PROCEEDINGS OF
THE 19TH INTERNATIONAL CONFERENCE
on Transformative Science and Engineering,
Business and Social Innovation

SDPS 2014 World Conference
Kuching, Sarawak, Malaysia
June 15-19, 2014
SMART INNOVATIVE SOCIETIES

EDITED BY: A.H. DOGRU, S. SUH
This year, our International Conference on Transformative Science & Engineering, Business and Innovation was held in Kuching, Sarawak, Malaysia: a beautiful and a most suitable setting for the conference theme “Smart Innovative Societies”. The enthusiastic participation by the supportive enterprises from government, university, and other affiliations added invaluable resources to the synergy. Keynote speeches conveying the interdisciplinary aspects of the scientific backgrounds for the investigation of such unstructured and complex problem sets were very interesting and leveraging for the ambitious goal. In a nutshell, it is possible to declare successful, the harmonious blending of various scientific contribution with the socially constructive views that has taken place.

The local organization was very effective. Such wide and high-quality participation was gathered in a very short time. Probably, continuity in focus inherited from the previous conference held in Brazil have supported the enthusiasm also. Anyway, the SDPS family is thankful for all the organizers and contributors. It was not difficult for anybody attending the gathering to feel the human touch due to the sincere aim to support the world society through science, luckily by the worldwide renowned names. The family keyword that is often used when SDPS is mentioned, is thus well deserved – most of our affiliates were already being asked about the date and the location for the next meeting!

The next expectation should be your participation for the continuation of the efforts towards founding of the transformative ideas that will lay the foundation for the future of knowledge integration and generation at a formal setting. Location for the next conference will be announced soon, to be followed in the web pages of the society. Also the journal of SDPS is another venue for your influencing the new science, innovation, and offering solutions to complex problems. It will not be too much to repeat our appreciation for the audience, authors, presenters, demonstrators, and speakers, besides an early appreciation for your future contribution. We are very thankful to the organizers for putting up such a successful establishment in a very short time, especially to Patrick Then. Also, we owe special thanks to Dr. Özgür Kaya for his far-fetched efforts in preparing this proceedings.

Ali H. Dogru
Sang Suh
# Table of Contents

1. REED SOLOMON CODES ASSISTED DISTRIBUTED VIDEO CODING IN NOISY ENVIRONMENT ................................................................. 5
2. SYNTHESIS OF SILICEOUS MESOPOROUS CAOTED SILVER MEMBRANE FOR OIL-WATER SEPARATION ...................... 11
3. DEVELOPING SUSTAINABLE COMPETITIVE ADVANTAGE AND RECONCILING STRATEGIC PARADOXES: A CASE STUDY OF THE COCA COLA COMPANY ................................................................. 16
4. SMART CARD TECHNOLOGY: CONTACTLESS PAYMENT SYSTEM AND NEAR FIELD COMMUNICATION ............ 21
5. BAT NEURAL NETWORK TO FORECAST IMPACT OF CLIMATE CHANGE ON LONG TERM RAINFALL TREND ...... 29
6. TELEMEDICINE PROGRAM FOR MANAGEMENT AND TREATMENT OF STRESS URINARY INCONTINENCE IN WOMEN: PROGRAM DESIGN .......................................................................................... 35
7. A PROPOSED SUPERVISORY WIDE AREA NETWORK FOR THE TRANS-BORNEO SMART GRID ................................. 39
8. CLASSIFICATION OF BLOG WRITERS ................................................................................................................................. 49
9. A COMPUTATIONAL APPROACH IN EPIDEMIOLOGICAL GAME THEORY ................................................................................ 53
10. STUDYING WITHOUT BORDERS ........................................................................................................................................... 59
11. INTERFERENCE EFFECTS ON SHORT RANGE TECHNOLOGY ZIGBEE PERFORMANCE ......................................................... 68
12. INTERFERENCE EFFECTS ON SHORT RANGE TECHNOLOGY BLUETOOTH PERFORMANCE ...................................................... 79
13. SOLUTION GENERATION IN CONCEPTUAL DESIGN USING ENVIRONMENT-BASED DESIGN: ALGORITHM AND CASE STUDY ......................................................................................................... 87
14. PERSUASIVE SUBTLETIES OF SOCIAL NETWORKING SITES: DESIGN IMPLICATIONS FOR BEHAVIOR CHANGE INTERVENTIONS ................................................................................................................. 94
15. TRACKING FUNCTIONAL DECLINE USING AMBIENT INTELLIGENCE FOR ALZHEIMER'S PATIENTS ............................ 106
16. REAL TIME HUMAN ACTIVITY RECOGNITION FRAMEWORK FOR KINECT SENSOR .......................................................... 115
17. NUMERICAL INVESTIGATION OF FINITE-AMPLITUDE PRESSURE WAVE PROPAGATION IN GAS SUSPENSIONS: APPLICATION TO AEROSOL CHARACTERIZATION ................................................................. 122
18. A FLEXIBLE DESIGN FOR SYSTEM CONTROL AND COMMUNICATIONS IN A ROBOTIC MOBILITY ASSISTANT FOR HANDICAPPED PERSONS .................................................................................. 129
19. INTELLIGENT SELF SUSTAINABLE GREEN ENERGY POWER SUPPLY MODEL FOR RURAL NETWORK ACCESS & HOME APPLIANCES .................................................................................................. 137
20. A SURVEY OF MAJOR INTER-USER INTERFERENCE MITIGATION TECHNIQUES IN WIRELESS BODY AREA NETWORKS ................................................................................................................................. 145
21. MAXIMIZING TRUE POSITIVES FOR SIGNATURE-BASED NETWORK APPLICATION IDENTIFICATION ............ 152
22. SOLUTION DECOMPOSITION FOR CURRICULUM DESIGN USING ENVIRONMENT-BASED DESIGN METHODOLOGY ........................................................................................................................................ 160
23. EXPERT SYSTEM FOR MAINTAINING A HEALTHY LIFE-STYLE ............................................................................................... 171
24. PSYCHOLOGICAL, MATHEMATICAL, AND PEDAGOGICAL ANALYSIS OF VIDEO STREAMS FOR MEASURES OF STUDENT ENGAGEMENT ........................................................................................................... 175
25. APPLYING JAPANESE EXPERIENCE TO RURAL COMMUNITY DEGRADATION CAUSED BY RURAL-URBAN MIGRATION IN SARAWAK .............................................................................................................. 182
26. DIGITAL MAMMOGRAM IMAGE ENHANCEMENT APPROACH BY CONTRAST STRECHING AND HISTOGRAM EQUALIZATION: AN IMPLEMENTATION AND RESULTS ......................................................... 184
27. EXPLORING THE POSSIBILITY OF COMPANION ROBOTS FOR INJURY PREVENTION FOR PEOPLE WITH
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.</td>
<td>MODELLING AND VALIDATION OF A DFIG WIND TURBINE</td>
</tr>
<tr>
<td>29.</td>
<td>A NOVEL AGGREGATION TECHNIQUE FOR DFIG WIND FARM MODEL</td>
</tr>
<tr>
<td>30.</td>
<td>IMPACT OF DFIG WIND FARM ON POWER SYSTEM STABILITY DURING TRANSIENT EVENTS</td>
</tr>
<tr>
<td>31.</td>
<td>EFFECT OF FLOCCULATED AND DISPERSED MICROFABRICS ON MECHANICAL PROPERTIES OF RECONSTITUTE KAOLIN SOILS</td>
</tr>
<tr>
<td>32.</td>
<td>SIGNAL INTERFERENCE EVALUATION OF EKO WIRELESS SENSOR NETWORK</td>
</tr>
<tr>
<td>33.</td>
<td>VARIABILITY MODELING IN COMPONENT ORIENTED SYSTEM ENGINEERING</td>
</tr>
<tr>
<td>34.</td>
<td>IDENTIFICATION OF LOCALLY FOUND BACTERIA FOR POTENTIAL USE IN GROUND IMPROVEMENT WORKS</td>
</tr>
<tr>
<td>35.</td>
<td>DEVELOPMENT OF PHOTOGRAMMETRIC MICRO UNMANNED AERIAL VEHICLE (UAV) SYSTEM FOR RESEARCH PURPOSE</td>
</tr>
<tr>
<td>36.</td>
<td>RIMBAILMU: AN INTERACTIVE SHARED SINGLE DISPLAY TECHNOLOGY FOR RESOURCE CONSTRAINED SCHOOLS IN SARAWAK</td>
</tr>
<tr>
<td>37.</td>
<td>MODERN INFORMATION OVERLOAD</td>
</tr>
<tr>
<td>38.</td>
<td>RAISING AWARENESS OF DENTAL PAIN THROUGH TWEETS</td>
</tr>
<tr>
<td>39.</td>
<td>THE REASONING PROCESSES FOR SELECTING TEACHING PRACTICES FOR STUDENTS WITH IMPAIRMENTS</td>
</tr>
<tr>
<td>40.</td>
<td>FALL DETECTION ALGORITHM BASED ON THE SLOPE OF THE VOLTAGE MEASUREMENTS FROM THE ACCELEROMETER</td>
</tr>
<tr>
<td>41.</td>
<td>AUTOMATIC PANNING SHOT GENERATOR USING TWO PHOTOS</td>
</tr>
<tr>
<td>42.</td>
<td>MALAYSIA’S HEALTH INFORMATION EXCHANGE (MYHIX): DEPLOYMENT BARRIERS AND SOLUTIONS</td>
</tr>
<tr>
<td>43.</td>
<td>UTILIZATION OF WOMEN RESOURCES WITH HIGHER EDUCATION IN KOREA</td>
</tr>
<tr>
<td>44.</td>
<td>A STUDY ON THE QUALITY FACTOR OF CULTURE MARBLE BATH-TUB &amp; WASHBOWL</td>
</tr>
<tr>
<td>45.</td>
<td>THE EFFECTS OF EXCHANGE RATE ON THE KOREAN REGIONAL TRADE BALANCE</td>
</tr>
<tr>
<td>46.</td>
<td>THE STOCK MARKET INTEGRATION AMONG US, JAPAN, AND KOREA</td>
</tr>
<tr>
<td>47.</td>
<td>COMOVEMENTS BETWEEN ASIAN EQUITY MARKETS</td>
</tr>
<tr>
<td>48.</td>
<td>AN EXPERIMENTAL STUDY OF WALL-PLATE U-BRACKET CONNECTION</td>
</tr>
<tr>
<td>49.</td>
<td>AN ASSET MANAGEMENT CASE STUDY IN A MALAYSIAN PUBLIC AGENCY – MOVING TOWARDS THE ACCRUAL BASIS OF ACCOUNTING</td>
</tr>
<tr>
<td>50.</td>
<td>EXAMINING THE ROLE OF INTELLECTUAL CAPITAL IN ELECTRONICS SMES IN KUCHING, MALAYSIA</td>
</tr>
<tr>
<td>51.</td>
<td>ORGANIZATIONAL INNOVATION AND BUSINESS PERFORMANCE: A MALAYSIAN SERVICE INDUSTRY PERSPECTIVE</td>
</tr>
<tr>
<td>52.</td>
<td>CORPORATE DISCLOSURE AND COST OF EQUITY: CASE OF MALAYSIAN LISTED COMPANIES</td>
</tr>
<tr>
<td>53.</td>
<td>REVISITING STOCK MARKET INTEGRATION PRE-POST SUBPRIME MORTGAGE CRISIS: INSIGHT FROM BRIC COUNTRIES</td>
</tr>
<tr>
<td>54.</td>
<td>ASSESSING THE QUALITY OF RETAIL WEBSITE AND AFTER SALES SERVICE USING IMPORTANCE-PERFORMANCE</td>
</tr>
<tr>
<td>55.</td>
<td>TOURISM-CO2 EMISSION NEXUS: THE CASE OF SOUTH KOREA AND MALAYSIA</td>
</tr>
<tr>
<td>56.</td>
<td>AN ECM ANALYSIS OF THAI TOURISM DEMAND IN MALAYSIA</td>
</tr>
<tr>
<td>57.</td>
<td>INTERNATIONAL FINANCIAL REPORTING STANDARD AND ITS VALUE RELEVANCE, MALAYSIA PERSPECTIVE</td>
</tr>
<tr>
<td>58.</td>
<td>DAY-OF-THE-WEEK EFFECT IN EQUITY MARKET: EVIDENCE FROM BURSA MALAYSIA</td>
</tr>
</tbody>
</table>
ABSTRACT

Distributed video coding (DVC) has been proposed as a new coding paradigm for video compression to achieve low complexity video encoding, where the bulk of computation is shifted to the decoder. To date, the proposed video codec based on DVC is utilizing the turbo codes as the forward error correction scheme. Moreover, most models are assuming that the transmitted data is subjected to perfect channel without any noise corruption. However, in emerging applications especially wireless networks, there exists fading channel that corrupts the transmitted data. In this paper, another forward error correction scheme, the Reed Solomon (RS) codes, are used to replace the turbo codes. The RS codes are chosen for their known capability to correct burst error, which is common in fading channel. Hence, the proposed model is also verified with the addition of Rayleigh fading channel as the noise. Simulation result shows that the proposed model can still give the acceptable performance in the present of fading channel.

NOMENCLATURE

- \( m \) = Number of bits per symbol
- \( n \) = Number of symbols per codeword
- \( k \) = Number of symbols per message
- \( t \) = Error correction capabilities of the RS codes
- \( d \) = Number of erroneous symbols in received codeword
- \( e \) = Number of erasures symbols in received codeword
- \( r \) = Code rate
- \( GF() \) = Galois field
- \( \beta \) = Primitive element in Galois field
- \( p(X) \) = Primitive polynomial over Galois field
- \( \Delta \) = Distortion between frames
- \( PSNR \) = Peak signal to noise ratio (dB)

INTRODUCTION

Pixel Domain Based Distributed Video Coding (DVC)

In recent years, applications where the memory and the computational power of the encoder are scarce are emerging. Hence, the conventional video coding approach with the statistical dependencies of frames are exploited at the encoder is no longer suitable although this design achieves high rate distortion (RD) performance. DVC is a new paradigm of video coding that shifts the complexity computation from the encoder to the decoder side [6]. The operation of the DVC is based on the Slepian-Wolf and Wyner-Ziv information theoretic results [5] [7].

One previous work in this field includes the pixel domain DVC architecture from Stanford University [1] [2] as proposed by the Aaron et al. (Aaron, et al., 2002).

Syndrome Based Transform Based DVC Architecture

In the same year, the syndrome based transform PRISM [3] [4] based DVC architecture is proposed by Puri et al. (Puri, et al., 2002) from the Berkeley University. Operating in the transform domain, the PRISM consists of a scalar quantizer followed by the syndrome coding of the least significant bitplanes of the discrete cosine transform (DCT) coefficients.

PRISM is shown to have good compression performance compared to the H.263+ technique [9]. However, the PRISM architecture also has some weaknesses. The true source statistics in the real world video sequences could not be predicted if there is no motion estimation at the encoder. Furthermore, the decoder is unable to control the bitrate and has the possibility of not finding any matching to the transmitted cyclic redundancy check (CRC) code bits.

These limitations are solved with the proposal of the DVC architecture with feedback channel [10]. It allows the decoder to control the bitrate by using the rate-compatible punctured (RCPT) codes.

More recently, the low density parity check accumulate (LDPCA) codes [11] have been proposed to replace the RCPT codes. The DVC kernel with the LDPCA codes has more superior error correction performance.
DISCOVER Architecture for DVC

Finally, in 2007, another transform-domain based DVC architecture, the DISCOVER [12], is proposed based on the Stanford architecture. It has improvement towards the transform-domain DVC scheme of [9].

Artigas et al. (Artigas, et al., 2007) adapted the LDPCA codes in DISCOVER over the turbo codes.

DVC Architecture with Adaptive Video Splitter

DVC has lower RD performance for video stream of high motion activity compared to video stream of low to medium motion activity. In the work [8], Kodavalla (Kodavalla, 2010) proposed the use of adaptive video splitter to adaptively split the incoming video stream into the key frames and Wyner-Ziv (WZ) frames based on the motion. Hence, the output of the video splitter is the group of picture (GOP) of various lengths. Longer GOP is preferred if there is more temporal redundancy between the frames and vice versa.

The adaptive video splitter also performs the decision on when and which frame shall only act as the key frame. This means that the GOP size is equal to 1. Thus, the frame shall only be encoded using the conventional intra encoder, bypassing the interframe encoder. As described in [8], for high motion sequences, the RD performance of H.264 intra coder is higher than the DVC. The video splitter could detect the high motion frames in the sequence, and bypass the WZ encoder. This makes the DVC to act as the H.264 intra coder.

Reed Solomon Codes

The Reed Solomon (RS) codes are a subset of BCH codes and class of linear block code. A RS encoder takes a block of digital data and adds extra redundant symbols, \(2t\). During transmission, there might be noise or interference which will cause errors. As a forward error correction scheme, the RS decoder processes each data block and attempts to correct the errors, recovering the original data. The number and type of errors that can be corrected depends on the characteristics of that particular RS codes and the number and distribution of errors.

1) If \(2d + e < 2t\) \((d\text{-errors, } e\text{-erasures})\), then the original transmitted codeword will always be recovered.
2) The decoder is not able to recover the original codeword and indicate this fact.
3) The decoder will mis-decode and recover and incorrect codeword without any indication.

Figure 1 shows a typical Reed Solomon codeword. In another publication [14], there are more details about the RS coding technique.

Objective

In this study, we utilized the RS codes as the forward error correction scheme to protect the source message prior to transmission. RS codes are chosen to replace the turbo codes in the DVC architecture proposed by the Stanford University. These codes are word oriented rather than bit oriented. This stems the codes the capability to correct burst errors, where a series of bits in the codeword is received in error. The burst error is relatively common in wireless communication due to fading. A bit oriented code would treat this situation as many independent single-bit errors. However, for RS codes, a single error means any or all incorrect bits within a single word. The proposed DVC model is verified in Rayleigh fading channel.

PROBLEM FORMULATION

Model

The model shown in Figure 2 uses the RS encoder and RS decoder to replace the turbo codes of the DVC architecture of the Stanford University.

The encoding process of DVC is very simple by design. The incoming video sequence is first into split into a GOP of eight frames. The first frame serves as the key frame while the remaining seven frames are the WZ frames. At the encoder, the key frame is encoded using the conventional H.264 as the intraframe encoding method. On the other hand, the WZ frames are first deciphered into a row of long message before being fed into the RS encoder.

In contrast, the DVC decoding process is relatively more complicated. The received key frame is first decoded using the H.264 intra decoder. Then, a frame interpolation method is used to generate the side information (SI) from the decoded key frame. The SI is an estimation of the next to be decoded WZ frame, which is only available at the encoder. To complete the decoding process, parity bits of the WZ frames are requested from the encoder using the feedback channel. If the correlation between the SI and the parity bits of the WZ frames is high, fewer parity bits need to be requested from the encoder to reach a certain decoding quality. However, if the decoding fails, request for parity bits is iterated until successful decoding is obtained.

To determine whether decoding is successful, the Hamming distance between the parity bits and the generated SI using the previously decoded frame is calculated. If the Hamming distance is non-zero, then the decoder proceeds to the next iteration to request more parity bits via the feedback channel. On the other hand, if the Hamming distance is zero, the decoding operation will then be verified with the 8-bits cyclic redundancy check (CRC) sum. Should the CRC sum computed on the decoded plane matches the value received from the encoder, decoding is considered successful and the decoded message is sent for reconstruction.
Reed Solomon Encoder

The function used to encode the data with RS codes is available in the MATLAB library. The RS function, RS \((n, k, m)\) has the following parameters: message length \((k)\), codeword length \((n)\), and number of bits per symbol \((m)\). Table 1 summarizes the meaning of each parameter and the allowable positive integers of the RS codes.

In this work, eight bits are chosen for each symbol, \(m = 8\). The code rate \((r)\) is 188/204, with \(n = 204\) and \(k = 188\). For the Galois field of the form of \(GF(2^m)\), the primitive element \((\beta)\) is 2. The primitive polynomial, \(p(X)\) over \(GF(2^8)\) that is primitive is defined as

\[
p(X) = 1 + X^2 + X^3 + X^4 + X^8
\]

To produce the narrow sense polynomial, the following function in the Matlab library is used.

\[
\text{genpoly} = \text{rsgenpoly}(n, k, \text{prim}_\text{poly})
\]

with the input argument \((\text{prim}_\text{poly})\) as the primitive polynomial. This function returns the narrow sense generator polynomial \((\text{genpoly})\) for the RS codes. The output represents the Galois row vector of the coefficients \((\text{genpoly})\) in the order of descending powers.

Another function in the Matlab library,

\[
\text{decode} = \text{rsdec}(\text{msg}, n, k, \text{genpoly})
\]

encodes the message \((\text{msg})\) with the RS \((n, k, m)\) codes with the narrow sense generator polynomial \([13]\). The source message is Galois array of symbol having \(m\)-bits each. Each \(k\)-element row of \(\text{msg}\) represents a message symbol, with the leftmost symbol as the most significant symbol. At most, the codeword length, \(n = 2^m - 1\). However, if \(n\) is not as exact as \(2^m - 1\), a shortened RS codes will be used. Redundant symbols are added at the end of each message symbol in the output Galois codes \([14]\).

Reed Solomon Decoder

During the transmission, there might be noise corruption due to the fading channel. In this study, the Rayleigh fading channel is simulated with a tool in the communication toolbox in SIMULINK. Hence, the received codeword \((\text{rec})\) shall be the summation of the transmitted codeword \((\text{encode})\) with the error \((\text{err})\).

\[
\text{rec} = \text{encode} + \text{err}
\]

The Matlab function,

\[
\text{decode} = \text{rsdec}(\text{rec}, n, k)
\]

will decode the received codeword \((\text{rec})\) with the narrow sense generator polynomial \([13]\). The codeword used in the decoder is a Galois array of symbols with \(m\)-bits each. Each \(n\)-element row of code represents a corrupted systematic codeword, with the leftmost symbol as the most significant symbol \([14]\).

If there are more than \(\frac{n - k}{2}\) errors detected in a row of the received codeword \((\text{rec})\), a decoding failure occurs. In this case, the decode function forms a corresponding row of decoded message by merely removing \((n - k)\) symbols from the end of the row of the codeword.

RESULTS AND DISCUSSIONS

The proposed DVC model as in Figure 2 is completely implemented with the C++ programming language. The RS encoding and decoding functions together with the Rayleigh fading channel are imported from the Matlab library and converted to the C++ language. The model is evaluated with the foreman QCIF video sequence. This video conferencing sequence has very high amount of motion activity.

Each reconstructed output frame in the GOP is compared with the corresponding original input frame. Based on the comparison, the peak signal to noise ratio \((PSNR)\) is calculated using the equation

\[
\text{PSNR} = 10 \log_{10} \left( \frac{255^2}{\Delta} \right)
\]

where \(\Delta\) is the distortion between the frames. The quality of the output frame is considered good if the value of the \(PSNR\) is above 30 dB.

Table 2 and Figure 3 depict the simulation result. From the data, it can be seen that the \(PSNR\) values for frames transmitted through perfect channel are above 36 dB. This also proves that there is very good reconstruction of the output frames at the decoder although there is no prior information available at the decoder.

The quality of the output frames without noise addition serves as a reference data. With the addition of Rayleigh fading channel and the source message is protected with RS codes, the \(PSNR\) values achieved are about 26 dB, close to the threshold \(PSNR\) values of 30 dB. However, the values are still very low compared with the \(PSNR\) values for frames transmitted through perfect channel, and this might be due to the multipath fading.

The results obtained are also compared with the findings in another publication \([6]\). The rate trace of the proposed model is compared with two other schemes: DCT-based intraframe coding and H.263+ interframe coding with predictive structure. The proposed model has almost similar performance with the conventional DCT-based intraframe coding. However, there is still a significant gap towards the H.263+ interframe coding as
the proposed model could not achieve the RD performance as the H.263+ interframe coding.

CONCLUSION AND FURTHER WORK

In this study, the Rayleigh fading channel is added in between the encoder and decoder of the basic DVC architecture. To better protect the source message against the noise, the RS codes are used as the forward error correction scheme. The implementation details and the results are presented.

It can be seen that DVC can be used in real wireless network world where there is multipath fading channel. RS codes serve as a better protection technique in this environment as the achievable PSNR is close to the threshold value for good frame quality (30 dB).

There are some key gaps and challenges in the practical usage of DVC. Firstly, the bitrate is inconsistent due to iterative decoding process and more parity bits might be requested for each iteration. This might cause decoding delay. Hence, there it is highly desired to use the rate estimation at the encoder.

To further improve the quality of the output frames smaller code rate \( r \) shall be used. As more redundant symbols are added, the better the RS codes in protecting the source message in the noise channel. Experiments should be carried out to determine the optimal code rate \( r \) that can best protect the source message in the noise channel and yet shall not increase the encoding complexity of the RS encoder.

Other type of noise like the Rician fading channel and the Nagakami fading channel shall also be added in the study to simulate the data transmission in the real world. At the same time, this study shall also determine that RS codes are performing the best in which fading channel.

Finally, there is no standardization yet for DVC compared to other video compression methods like MPEG2 and MPEG4. This is necessary to maintain consistency among various methods in the interest of interoperability among various encoders and decoders. The model presented in this study is adapted from the DVC architecture from Stanford University.

REFERENCES


Matlab Help, Matlab v7.4.0.287 (R2007a), The Math Works, Inc.


FIGURES AND TABLES

![Fig. 1 Typical structure of a RS codeword](image)

Fig. 1 Typical structure of a RS codeword [14]
Fig. 2 Proposed DVC model with RS encoder and RS decoder

Table 1 Parameters for RS codes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Value or Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>Number of bits per symbol</td>
<td>Integer between 3 to 16</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of symbols per codeword</td>
<td>Integer between 3 to $2^m - 1$</td>
</tr>
<tr>
<td>$k$</td>
<td>Number of symbols per message</td>
<td>Positive integer &lt; $n$, where $(n-k)$ is even</td>
</tr>
<tr>
<td>$t$</td>
<td>Error correction capability of the codes</td>
<td>$\frac{(n-k)}{2}$</td>
</tr>
</tbody>
</table>

Table 2 PSNR values for output frames

<table>
<thead>
<tr>
<th>Frame Number</th>
<th>Rayleigh Fading Channel</th>
<th>Perfect Channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>26.1747</td>
<td>36.1477</td>
</tr>
<tr>
<td>1</td>
<td>26.3132</td>
<td>36.3022</td>
</tr>
<tr>
<td>2</td>
<td>26.5241</td>
<td>36.3240</td>
</tr>
<tr>
<td>3</td>
<td>26.6725</td>
<td>36.2575</td>
</tr>
<tr>
<td>4</td>
<td>26.8941</td>
<td>36.2931</td>
</tr>
<tr>
<td>5</td>
<td>26.0945</td>
<td>36.2795</td>
</tr>
<tr>
<td>6</td>
<td>26.0813</td>
<td>36.2923</td>
</tr>
<tr>
<td>7</td>
<td>26.6729</td>
<td>36.2524</td>
</tr>
</tbody>
</table>
Fig. 3 PSNR values for output frames
SYNTHESIS OF SILICEOUS MESOPOROUS COATED SILVER MEMBRANE FOR OIL-WATER SEPARATION

Farouq Twaiq¹, Hussain Al Balushi²
¹Faculty of Engineering, Computing and Science
Swinburne University of Technology, Sarawak Campus
Jln Simpang Tiga, 93350 Kuching, Sarawak, Malaysia
²College of Engineering, Nizwa University, Brikat Al-Mouz, Nizwa, Oman

ABSTRACT
In this research, we have synthesized a siliceous mesoporous coated silver membrane for oil-water separation. The objective of the study is to testify the effect of mesoporous material deposited on the membrane surface on the removal of organic materials from aqueous streams. The mesoporous material is prepared using tetraethylorthosilicate (TEOS), hexadecyltrimethyl ammonium bromide (CTAB) template and nitric acid as a crystallizing agent, before it is allowed to deposit on the surface of the silver membrane. The membrane is characterized using scanning electron microscope (SEM) in order to observe the deposition and the morphology of the mesoporous material on the silver membrane. The coated membrane is tested for oil-water separation, and the diffusion factor was calculated from Fick’s law. The membrane coated with silica prepared from gel containing about 1 mL HNO₃ has higher value of diffusivity and therefore is preferable to be used as membrane.

NOMENCLATURE

\[ J \] = mass flux of the oil (A) across membrane section area (kg/m²·h)
\[ D \] = diffusion constant (m/h)
\[ dC / dx \] = concentration gradient of oil in the membrane (kg/m³)
\[ C_A \] = oil concentration in the feed tank (mg/L)
\[ C_B \] = oil concentration in the filtrate (mg/L)

INTRODUCTION
There is an increasing demand for clean water that is obtained from natural or recycled sources which contains an elevated levels of organic matter. The presence of organic matter causes a concerned environmental awareness. In the petroleum industry, the oil/water mixture occurs as waste streams in the operations of production, transportation and refining processes. This leads to the needs to treat these streams for either reuse this kind of water or for safe disposal. Membranes are a leading choice for water and wastewater purification. Membrane separation is widely used for water purification now a days due to their relatively low cost and easy to use with high efficiency. There is an increasing interest in the application of membrane filtration for removal of undesired contaminants. Examples of nanofiltration applications are the removal of pesticides from groundwater, the removal of heavy metals from wastewater, wastewater recycling in laundries, produced water treatment, etc. (Grassi et al, 2012).

Membranes are manufactured from materials such as synthetic polymers, ceramics, metals and metal oxides (Mulder, 1996). The material properties of the membrane may significantly impact the design and operation of the filtration system. The synthesis of inorganic membranes is an intensive research field, many contributions have been made concerning to select materials which are ordered microporous structures. Inorganic membranes are present as self supporting structures or multi-layer supporting structures. Compared to polymeric membranes, inorganic membranes can withstand more extreme conditions. Ceramic membranes are widely used as they are less reactive and less fouling membranes, especially when ozone treatment is used (Sagle and Freeman, 2014). Mesoporous membranes are used in water purification. To make them fully competitive, their manufacturing cost must be reduced by improvements in processing (Verweij, 2012).

Membrane fouling, caused by accumulation of rejected materials, reduces membrane performance and shortens the operation duration, thus increasing the operating costs, and shortening the membrane life. A limitation of more widespread application of filtration processes is the performance (i.e. Fouling) that occurs during potable water treatment (Zhao et al, 2000). As membrane technology is advancing, membrane fouling and membrane chemical stability still need the research attention for more economical, environmentally friendly, versatile, and easy to use.
The objective of the current research is to synthesize and test mesoporous silica coated silver membranes for separation of oil-water mixture and to evaluate the diffusion factor of these membranes.

**MATERIALS AND METHODS**

Synthesis of mesoporous silica on the silver membrane was carried out using the following procedure. A solution containing about 1.5 g of 99% tetraethylorthosilicate (TEOS) (Fluka) was dissolved under vigorous mixing in 45 mL deionized water and 70% nitric acid (Scharlau) (1, 2, 3, and 4.6 mL). A solution of the template of 1 mL 96% hexadecyltrimethylammonium bromide (CTAB) (Fluka) was dissolved in 40 mL absolute ethanol (J.T.Baker) at 40°C. The template solution is added to the silica solution slowly by using a dosing pump (1) as shown in Fig. 1, under vigorous stirring. The solution was circulated using a tubing pump (2) through the silver membrane (25 mm diameter, 5 micrometer porosity) obtained from Cole-Parmer, for 1 h. After that deionized water is circulated for 1 h and the silver membrane is removed from the membrane holder to dry at ambient temperature overnight, and then calcined in a muffled furnace at 550°C for 6 hours. The resulting coated membrane is characterized using JSM-6510LV scanning electron microscope (SEM).

The filtration testing of the coated membrane is performed in a rig setup as shown in Fig. 2. The rig consists of a five liter plastic container placed at 2.5 m height at the top of the rig containing oil-water mixture. The testing filter holder connected using 0.625 cm diameter rubber tube and is placed at 1 m below the bottom of the plastic container. The liquid flow is controlled by a ball valve. The filtrate is collected in a beaker below the filter. A sample of the collected liquid is analyzed using a Horiba oil analyzer in order to evaluate the oil content in the filtrate.

**RESULTS AND DISCUSSION**

**Synthesis and Characterization**

A purely silver membrane was characterized using SEM before mesoporous material coating is performed as shown in Fig 3. The SEM images show that the silver membrane has irregular porosity in the range of 4 – 5 μm (see Fig 3(a)). The silver membrane thickness is measured to be ~50 μm (see Fig 3(b)). The surface of the silver membrane is straight and clean without contaminants on the surface.

Self-assembly of mesoporous silica materials is of great interest for modern nanotechnology because of uniform pore size and high stability of the materials. The use of cationic surfactants in the presence of a silica precursor can result in synthesizing a variety of well-defined nanoporous silica shapes. Figure 4 shows the siliceous mesoporous material deposition at the silver membrane. The coating is prepared using mesoporous synthesis gel containing 1 mL of HNO₃, which is used as a crystallization agent that hydrolysis the TEOS as silica source. Particle depositions of less than 1 μm size are clearly observed at the silver membrane pores and surface. The deposition reduces the pore size of the original silver membrane and cause a roughness of the silver surface. It is difficult to control the shape of the mesoporous particles on the silver film.

Figures 5(a,b) and 6(a,b) are showing the mesoporous silica deposition using synthesis gel containing 2 mL and 3 mL HNO₃, respectively. The increase of HNO₃ contents accelerates the crystallization reaction and leads to the formation of larger particles, as can be seen in Figures 5(a, 6a). HNO₃ promotes the formation of rope like crystals as seen in Figures 5(b) and 6(b) whereas some particles are large and rope like shape of length exceeds 50 μm. One can see that the fiber formed at the surface is not uniform due to the decrease of the silica concentration during the filtration process. There are a number of round shapes, and the fibers are bent. This makes this material very attractive for various applications such are a filtration process (Kievsky and Sokolov, 2005).

It was observed that when the HNO₃ content increase, the brittleness of the membrane increases, which make them not suitable as membrane application as the case of sample prepared using 4.6 mL HNO₃.

**Filtration Testing**

Table 1 shows the results of testing for purely silver membrane and coated silver membrane with mesoporous silica coating for oil-water separation. Table 1 shows the data of solution flux through the membranes versus filtration time. It is observed that the solution flux through the membrane prepared using 1 mL HNO₃ is higher compared to all membranes prepared and higher than purely silver membrane only for the first 0.22 h running times. The reason is that the mesoporous materials prepared using 1 mL HNO₃ are small particles deposited directly on the silver surface and silver pores. The mesoporous material is a very good adsorbent for oil. The oil is removed from the water by adsorption allows high water flux. After 0.22 h of operation time, the mesopores are filled with oil and the pores are blocked disallowing the water stream to flow.

At the same time, the solution flux is very small in the samples prepared with high HNO₃ contents (3 and 4.6 mL HNO₃). That could be due to the fouling of the silver pores caused by the deposited particles which are larger in size and rope in shape.

Table 2 shows the oil content in the filtrate stream. The oil concentration in the stream is found to be decreasing with increasing the collected filtrate volume. The membrane prepared using 1 mL HNO₃ was the best in terms of oil reduction. The samples with high HNO₃...
content did not reduce the oil content in the filtrate even after long filtration time. That is because the flux is small and the particle sizes of the deposited mesoporous materials are larger than the porosity of the silver membrane.

In conclusion, the membrane fouling is affected by membrane surface morphology (morphology of the mesoporous deposition). And the oil adsorption on the membrane surface increases the fouling but enhances the oil removal for the oil water mixture.

**Modeling**

Many mathematical models used to describe diffusion assume steady state conditions. Steady state conditions assume that diffusant concentrations remain constant at all points on each side or surface of the membrane. The concentration changes linearly through the membrane and the rate of transfer for a diffusing substance is the same across all sections. Provided that the diffusion coefficient is constant, Fick’s law of diffusion for one dimensional and steady state process is being expressed as in Equation (1). Therefore, the rate of transfer per unit area of the section is calculated by Equation (2).

\[
I_A = -D_A \frac{dC_A}{dx}
\]  

(1)

\[
I_A = -D_A \left( C_A - C_B \right)
\]  

(2)

In order to fit the mass transfer model, the data obtained from experimental runs are mass flux versus concentration. The data from experiment are collected as volumes vary in time on stream. A plot of flux versus the concentration difference between the fed and filtrate is obtained from the experimental data in order to estimate the diffusion coefficient. The diffusion coefficients of different membranes are presented in Table 3. From the data membrane prepared using 1 mL HNO₃ has the highest value of oil diffusivity.

**CONCLUSION**

In this research, synthesis and testing of mesoporous coated silver membrane for oil-water separation is investigated. The coating is made of pure siliceous mesoporous material prepared using tetraethyl orthosilicate (TEOS). The synthesis coating is prepared using different amounts of nitric acid as crystallizing agent. The membrane is characterized using a scanning electron microscope (SEM) in order to measure the pores size and membrane thickness. The synthesized membrane is tested for oil-water separation. Data obtained from testing was used to fit a suitable mass transfer model and diffusion factor was estimated. Sample membrane prepared using 1 mL nitric acid performed the best in terms of oil separation, however pure silver membrane performed better in terms infiltrate flux.

**REFERENCES**


**TABLES AND FIGURES**

**Table 1. Filtration time (h) of oil-water mixture (112 mg/L oil in water) for silver and mesoporous coated silver membranes**

<table>
<thead>
<tr>
<th>Accumulative volume mL</th>
<th>Pure silver membrane</th>
<th>1 mL HNO₃</th>
<th>2 mL HNO₃</th>
<th>3 mL HNO₃</th>
<th>4.6 mL HNO₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0.26</td>
<td>0.22</td>
<td>0.33</td>
<td>2.61</td>
<td>5.13</td>
</tr>
<tr>
<td>90</td>
<td>0.96</td>
<td>1.66</td>
<td>2.72</td>
<td>5.66</td>
<td>23.83</td>
</tr>
<tr>
<td>135</td>
<td>2.02</td>
<td>4.49</td>
<td>9.45</td>
<td>18.92</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>2.88</td>
<td>9.59</td>
<td>22.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>6.03</td>
<td>19.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>48.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Concentration (mg/L) of oil in the filtrate stream by filtrating the oil-water mixture (112 mg/L) using silver and silica mesoporous coated silver membranes**

<table>
<thead>
<tr>
<th>Accumulative volume mL</th>
<th>Pure silver membrane</th>
<th>1 mL HNO₃</th>
<th>2 mL HNO₃</th>
<th>3 mL HNO₃</th>
<th>4.6 mL HNO₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>34.7</td>
<td>28.9</td>
<td>57.6</td>
<td>38.7</td>
<td>28.0</td>
</tr>
<tr>
<td>90</td>
<td>14.7</td>
<td>12.9</td>
<td>30.8</td>
<td>23.1</td>
<td>12.8</td>
</tr>
<tr>
<td>135</td>
<td>8.4</td>
<td>10.1</td>
<td>18.2</td>
<td>13.8</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>6.8</td>
<td>7.6</td>
<td>13.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>3.8</td>
<td>6.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Diffusivity coefficients of different membranes used to separate oil-water from mixture containing 112 mg/L oil in water

<table>
<thead>
<tr>
<th>Type of membrane</th>
<th>Diffusivity (m/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure silver membrane</td>
<td>0.386765</td>
</tr>
<tr>
<td>Mesoporous coated silver prepared using 1 mL HNO₃</td>
<td>0.554303</td>
</tr>
<tr>
<td>Mesoporous coated silver prepared using 2 mL HNO₃</td>
<td>0.374509</td>
</tr>
<tr>
<td>Mesoporous coated silver prepared using 3 mL HNO₃</td>
<td>0.045707</td>
</tr>
<tr>
<td>Mesoporous coated silver prepared using 4.6 mL HNO₃</td>
<td>0.056109</td>
</tr>
</tbody>
</table>

Fig. 1 diagram showing the setup used for the synthesis of mesoporous silica on silver film

Fig. 2 Membrane filter testing rig

Fig. 3 SEM image of porous silver membrane showing (a) Silver membrane surface and pore size distribution (b) silver membrane thickness

Fig. 4 SEM image of silica particle deposition on silver membrane prepared using 1 mL HNO₃
Fig. 5  SEM image of silver membrane coated with mesoporous silica prepared using 2 mL HNO₃ (a) siliceous mesoporous particle deposition (b) particle morphology

Fig. 6  SEM image of silver membrane coated with mesoporous silica prepared using 3 mL HNO₃ (a) siliceous mesoporous particle deposition (b) particle morphology
ABSTRACT

Coca Cola is probably one of the most recognized brands in the world. There are a lot of complexities involved in making strategic decisions at enterprises such as Coca Cola, a company with global operations in over 200 countries, with 250 bottling partners, 90 production facilities world-wide and over 500 brands according to their latest annual report. How do you manage complex strategic tensions in a firm which is so global and enormous? This paper attempts to unravel distinctive qualities which give Coca Cola a sustainable competitive advantage and enables us to understand better how the company reconciles some key strategic tensions that give it an edge over its competitors.

INTRODUCTION

It is intuitive that the performance of a firm depends on the quality of its products. Wernerfelt (1984) claims that firm resources and its products are like two sides of the same coin. So, inherently if the firm has good resources the products it churns out will be good as well. Grant (1991) lists six categories of firm resources: financial, physical, human, technological, reputation, and organizational.

These resources outlined by Grant (1991) are available to all firms, so there should be certain attributes about these resources which set a well performing firm apart from the average firm. Barney (1991) gives us an insight of the characteristics that the firm resources should have so as to be able to provide the firm with a competitive advantage. Barney’s approach is known in short as VRIO. According to Barney, the firm’s resources should be valuable, rare, in-imitable and organizational. The theory that firm resources are key drivers of performance and lead to a sustainable competitive advantage is known as the RBT (Resource Based Theory) and has gained prominence over the last two decades (Newbert 2007).

Dealing with strategic paradoxes is critical to the success of any modern day organization. Smith et. Al (2010) suggest that most managers traditionally think of the strategic paradox as choosing the best option from either strategy A or strategy B, the successful reconciliation of this leads to paradoxical strategies that accommodate both strategy A and strategy B to suit the interests of the organization. This accommodation and co-implementation of paradoxical strategies forms the cornerstone of the success that The Coca Cola Company has witnessed over the course of the century.

DEVELOPING SUSTAINABLE COMPETITIVE ADVANTAGE

If there is one thing that businesses have learnt from the GFC (Global Financial Crisis) it is that they should have enough liquidity or financial resources to survive through bad times. Coca Cola is a cash generator. The firm’s cash generation from operations has increased at an average rate of 12% from 2010 to 2012. This steady generation of cash enables re-investment, research and further brand development and makes Coca Cola an industry leader. Coming back to Barney’s VRIO approach, in today’s post GFC world, ample cash generation and liquidity is perhaps the most valuable resource a firm can have and it is the optimal deployment of such resources that indeed leads to a sustainable competitive advantage.

From Myanmar to Mayfair, from London to Lahore, from the bazaars of Damascus to the high end shopping malls of Dubai and from the Jewish part of Jerusalem to Muslim dominant Jeddah, Coca Cola crosses all boundaries, all tensions, and all parallels and is recognized as a premier beverage. The iconic contour shaped Coca Cola bottle stands for a brand name, for a product, for a taste, for an experience that millions around the globe have loved over the course of over a century. This global recognition is quite rare and perhaps even rarer is the ability to operate profitably across the globe and generate billions and billions of dollars in shareholder value year after year.

A thorough examination of their annual report and their investment relations overview provides valuable insights in to their operational and market strategies. In a nutshell, they aspire to be the number one beverage seller in all their operating markets. However, this goal oversimplifies the strategic variation that the company practices in different markets. Coca Cola points out that
one approach certainly doesn’t fit all markets. Their developed market approach for North America and Europe is to innovate constantly and provide better quality products. They are focused on providing low-calorie, less calorie and no calorie alternatives in these markets.

Coca-Cola’s approach for the developing markets is quite different than this innovation driven approach that they have for developed markets. For developing markets they focus more on market segmentation, primarily focusing on Coca-Cola drinks in different sizes, in different quality packaging and definitely more focused on providing low cost alternatives to rather expensive use and throw glass bottles etcetera. So, in a nutshell, their operational strategy in developed markets is to strive for greater quality whereas, providing products at the cheapest possible price is their strategy to compete in developing markets such as India and China.

At first glance, it may seem like Coca-Cola’s resources are fairly imitable, after their product is just another carbonated drink and there are so many alternatives available such as Pepsi, Dr. Pepper etc. However, their product is targeted towards conquering not the human taste buds, but in fact, the human mind. We all have a favorite desert, a favorite ice-cream flavor, most of us who smoke stick to a particular brand. Likewise, Coca-Cola has a loyal customer base that will pay for their preferred taste.

According to the online booklet “125 years of sharing happiness” available on Coca-Cola’s website, Coca-Cola has taken significant steps over the last century to make sure that its brand is inimitable. Coca-Cola’s unparalleled success encouraged other competitors to try to imitate Coke by offering products with slight variations on the trademarked name and logo. Koka-Kola, Koca-Nola, Celery-Cola and Koke were just a few of the products that tried to imitate Coca-Cola and capitalize on its success. In alliance with its bottlers, The Coca-Cola Company came up with a design that was so distinctive that it could be recognized by ‘feel in the dark’ or identified even if broke in to pieces. Hence, the iconic contour bottle was born and even today, it is still the most recognized bottle in the world.

**DISCUSSION ON RECONCILING STRATEGIC PARADOXES**

**1-Competition vs. Cooperation:**

Network Level Strategy of an organization dictates whether the organization adheres to inter-organizational competition or organizational cooperation (De Wit and Meyer). The corresponding strategy perspectives that an organization employs if it believes in competition or cooperation are the discrete organization perspective and the embedded organization perspective respectively.

Coca-Cola is an organization with well-defined values, organizational goals and structure. The philosophy of collaboration and cooperation is deep rooted in Coca-Cola’s organizational structure. In an article entitled “The science of alliance” from the Economist, we are informed that apparently there is no “piece of paper to fall back on” with regards to the alliance between McDonald’s and Coca-Cola.

According to “The science of alliance” article, in the words of the then Coke CEO Mr. Ivester the relationship between Coke and McDonald’s is driven just by “a common vision and a lot of trust”. In the same article, Mr. Ivester is asked about the proportion of Coca-Cola’s income depending on alliances, his answer without hesitation is “100%”. He further explains that every dollar that Coca-Cola makes is coming from some form of partner— a bottler, a distributor and so on. The article points out that the hallmark of the McDonald’s and Coke alliance is its informality and its open-endedness. It further stipulates that the extent of the alliance extends to the level of helping McDonald’s set up new stores in countries where Coca-Cola already has a presence. In doing so, Coca-Cola is clearly expanding its own outlets.

The article in the Economist discussed in the paragraph above was published in 1998. Over a decade later, according to Coca-Cola’s latest annual review, McDonald’s is Coca-Cola’s single largest customer, selling Coca-Cola in over 100 countries across 31,000 restaurants and still growing. Javier C. Goizueta who is Vice President of The Coca-Cola Company is the President of the global McDonald’s Division. He leads a worldwide organization that is responsible for building the strategic alliance with McDonald’s. It is with such collaboration that a win-win situation has been created for both the partners. There are no losers in this alliance.

Coca-Cola’s alliance with McDonald’s is strategically managed; the fact that they have an entire division to cater to the needs of their single largest customer is reflective of the significance of the alliance. The article ‘Creating a strategic centre to manage a web of partners’ by Lorenzoni and Fuller further supports this strategic alliance management approach practised by Coca-Cola. According to the article, the strategic centre should ensure that each partner gets to benefit from the core competencies of one another. This is evident in the fact that Coke actually helps McDonald’s in setting up restaurants in markets where McDonald’s has no prior experience, but Coke already has a presence.

Also, the article by Lorenzoni and Fuller, says that the central firms have to create an environment of trust and reciprocity. This is exactly what Coca-Cola has done with McDonald’s and with Disney as well. The only beverages you can buy at both are Coke products. Also, in the past Coke has featured Mickey Mouse on its bottles and likewise Disney has placed a large glass bottle of Coca
Cola on its theme park which dispenses the beverage to customers (The Coca-Cola Annual Review 2012-2013). The McDonald’s and Coca-Cola alliance is widely known. In this particular case, the alliance makes a lot of sense for both the partners. They are both market leaders and their products are complementary rather than substitutes for one another. In 1993, Coca-Cola re-entered Indian markets after a long absence. They marked the 20th anniversary of their return to India in 2013 and they have one entire report dedicated to their re-entry in the Indian markets on their website.

The entry plan in to the Indian market as suggested by the article, “20 years later: A Look Back at Coke’s Dramatic 1993 Return to India” on Coke’s website was a brave step and it began by forming an alliance which is quite different to that of the alliance with McDonald’s. McDonald’s and Coca Cola are complementary products; Pepsi and Coke are competitors and substitutes.

To tap the Indian markets Coca Cola joined hands with Parle which had control of 60% of India’s beverage markets. They acquired Parle beverages and became owners of India’s leading beverage brands such as Thums Up, Limca, and Gold Spot etcetera. It turned out to be a really wise strategy though; the acquisition gave them access to the existing bottling and distribution networks. These networks would now be used to bottle and distribute Coca Cola products as well.

The partnership between Coke and Thums Up was a big gamble as their leading executives describe in the article discussed above. However, the top executives also say that they didn’t promote Coca-Cola at the expense of Thums Up. They promoted Thums Up as the national icon and a symbol of nationalism for many young Indians. On the other side, they promoted Coca-Cola as aspiring new drink. The article “Collaborate with your Competitor- and Win” by Hamel, Doz and Prahalad actually explains how this strategy worked out in the favor of both Coca-Cola and Thums Up. Hanel et al. say that a collaboration can work and produce mutual gains as long as ‘The partners’ strategic goals converge while their competitive goals diverge’. Twenty years later, India, is Coca Cola’s seventh largest market. This shows the extent of collaborative and cooperative tactics that Coca Cola can employ.

2- Compliance vs. Choice:

Under Industry context, there are two paradoxical strategic perspectives namely- the industry dynamics perspective and the industry leadership perspective. The industry dynamics perspective requires firm strategy to be in compliance with industry whereas; the industry leadership perspective requires employment of strategy that enables a firm to exercise its own choice. Coca-Cola treats compliance vs. choice as a true paradox and manages to reconcile its strategic tensions in this context very efficiently.

Does Coca-Cola develop a business model distinctively different from its competitors? Or does industry logic largely determine its strategy?

As far a competitor for Coke is concerned, the only rival worthy of being called a global competitor of the Coca Cola Company is Pepsi Company. Coca Cola has stuck to its core competence in the beverage industry. Pepsi according to its latest annual report now earns close to half of its revenues from the snack food and Quaker breakfast oats. Also, interesting is the fact that the North Americas account for only 20% of Coke’s sales, whereas, for Pepsi it accounts again for half of their sales. So, in comparison to Pepsi, Coca-Cola is far more globalized and focusses primarily on the still and sparkling beverage industry.

As far as conducting the business is concerned, both of them work with external bottlers and distributors. The companies provide their bottlers with concentrate and marketing support. It is independently owned bottlers who actually manufacture the finished product and supply it to distributors who are again independent entities. So, technically in this arena industry logic prevails as running the company on a micro-level around the globe would be a tedious and inefficient exercise.

According to the annual reports of both companies, they both seem to be on course to delivering healthier alternatives in the form of fewer calories and zero calories alternatives. Also, they both seem to be tapping on the demand for fruit juice with Tropicana (Pepsi) and Minute-Maid (Coca-Cola) brands. This leads us to conclude that Porter’s article “Industry Evolution” from his book entitled Competitive Strategy holds true. According to Porter, firms that adapt to the operating environment profit and survive. Also, to quite an extent the firm’s strategy and actions are dependent on industry requirements. In this case, it is evident that the industry is demanding healthier choices and both Coke and Pepsi seem to be merely complying.

Coming back to the example of Coca Cola’s re-entry in to the Indian market, the strategy wasn’t limited to buying up Parle group and the Thums up brand alone. The manner in which Coca Cola was launched in the Indian markets itself re-defined the beverage industry in India. Pepsi was selling their cola at a price of INR 5.50 for a 250 ml bottle. Coca Cola introduced their bottle with a size of 300 ml and at a price of INR 5.00. They were offering higher value at lower cost. This is a classic example of a blue ocean strategy of breaking the value/cost trade off that authors Kim and Mauborgne discuss in their best seller “Blue Ocean Strategy”. The value/cost trade off implies that to increase product value you need to increase the price of the product. Companies which break this trade off and are able to deliver better value or quality at a lower cost fare better than their competitors.
According to annual reports of Pepsi and Coke, Diet Pepsi was first sold in 1964 and Diet Coke didn’t hit the shelves until 1982. However, today Diet Pepsi has a 5.3% share of the U.S. beverage market while Diet Coke accounts for 9.9% of market share. We see that Industry leadership didn’t prove profitable for Pepsi in this case, whereas, adherence to industry dynamics proved to be an extremely profitable move for Coca Cola. This brings us to an interesting point that the authors of the article, “The Firm matters not the industry”, Fuller and Stopford make in their article. According to them no industry is pre-disposed to be a successful business for all entrants. It is the strategic choices, the product quality and the differentiation attributes that the firm applies that really matter.

CONCLUSION

Coca Cola’s numbers speak a lot for its successful strategies and with a such a great presence around the globe and its ability to be a collaborator, a tough competitor, an industry leader and an industry follower all in appropriate proportions, it is bound to continue to grow and conquer greater heights.

There is a lot that can be learned from the manner in which Coca Cola handles its strategic tensions. This paper clearly demonstrates the complexity of strategic decision making. It also, tells us why it is important to reconcile strategic paradoxes without taking a bi-polar approach. It is imperative that firms realize both sides of the argument and accommodate paradoxical strategies wisely to serve their interests.

Every strategic tension is an opportunity in disguise and firms like Coca Cola who know how to exploit these tensions and reconcile them in an effective manner actually benefit from these tensions. The average firms aren’t able to compete with the better firms owing to their lack of proper management of strategic tensions.

This paper also demonstrates that traditional and conventional thinking of focusing on a single strategy and sticking to it isn’t very helpful. Strategy needs to be flexible and often the same firm will need to employ paradoxical strategies to fully attain its goals. This outcome is a dilemma for most firms as it is not always easy to employ counter-intuitive strategies within the same firm. However, it is these firms that are able to reconcile strategy paradoxes successfully that really reap great benefits in the long run and are successful in sustaining competitive advantage.

REFERENCES


GRI Report from company website


Annual Review from company website


Investor Relations Report from company website


Sustainability report from company website


http://www.pepsico.com/annual12/
Annual report from company website


Historical Report from company website

SMART CARD TECHNOLOGY: CONTACTLESS PAYMENT SYSTEM AND NEAR FIELD COMMUNICATION

Opeoluwa Tosin Eluwole, Mahboubeh Lohi and Nsima Sylvanus Udoh
Transaction Security Department, Underwriters Laboratories, Basingstoke, UK
Department of Electronic, Network and Computer Engineering, University of Westminster, London, UK

ABSTRACT

The metamorphic cycle in the payments and security aspects of smart card technology is characterized by transformation from paper cheques into magnetic stripe only cards, further into chip only cards, and then cryptographically secure chip-and-pin cards. The latter in recent years trendily developed into contactless card systems which consequently evolved into state-of-the-art near field communication (NFC); a technology based on radio frequency identification (RFID) principles. These developments have been remarkable over the last few decades to such a degree that nowadays, payments can be made on the move by simply waving or tapping a contactless smart card in the proximity of a point-of-sale (POS) terminal or reader, popularly dubbed the “wave and go” or “tap and pay” sensation. This is arguably the biggest buzz in the payments industry of the modern era; perhaps too huge of a buzz. This paper explicitly explains the contactless system, describes how it evolves into NFC as well as presents a survey with highlights on the numerous applications and benefits the technology promises to offer. Security threats posed to contactless systems are also underlined alongside potential remedies to such attacks. The paper seeks to enlighten people who may have recently heard the term “contactless or NFC” for the first time and need to know what it is and how it helps or affects them.

INTRODUCTION

Since the invention of smart cards over three decades ago, they have dominated most developments of the information age especially in the banking/payments industry. Figure 1 depicts the evolution of smart card technology as viewed from the banking/payments and security perspective. Historically, the first smart cards (prepaid telephone cards) were launched in Europe about three decades ago, precisely in 1984 (Finkenzeller, 2010). Today, smart cards make all services that relate to information or financial transactions simpler, more convenient and cost-effective. Numerically, between 1992 and 1995, over 600 million smartcards were already deployed globally. By the end of 1999, over 1.5 billion smart cards had been shipped around the world with an annual growth rate of 60% forecasted for United States’ smart card shipments between 1998 and 2003. If the growth rate from 1992 through 1999 to 2003 is sustained, then by 2020, at least 20 billion smart cards would have been deployed globally. Today, the smart card market can be said to represent one of, if not the fastest growing subsectors of the microelectronics industry (Finkenzeller, 2010; Gemalto, 2013).

Smart Card – Composition, Types, Applications and Deployment Statistics

By definition, a smart card is a microprocessor or memory chip-embedded card with processing power functionality to provide intelligence, data portability, security, and convenience. In terms of internal functionality or variety, a smart card can be memory-based (simply for data storage) or microprocessor-based (has in-built security features to add, delete or manipulate data). In terms of interface, a smart card can either be contact, contactless or dual (hybrid) (Gemalto, 2013). Figure 2 shows the various categories of smart cards and the standards governing them.

A contact smart card, as it implies, involves insertion into a reader (terminal), making physical contact with it in order to communicate (Finkenzeller, 2010). A contactless smart card, which is the crux of the matter in this paper, has embedded antenna that allows for proximity connection with the reader using an RF interface without making any physical contact (Inside Contactless, 2004). When both contact and contactless applications are integrated on a card using a single dual-interface chip, the resultant smart card is called a hybrid or dual or Combi card (Gemalto, 2013).

There are extensive applications of contactless transactions ranging from access control to financial payments; the latter is becoming progressively more popular particularly in Europe, Asia and the United States (Diakos, et al., 2013). Statistically, as of July 2013, at least about 60 million contactless cards, 1 million contactless readers, and 31 million contactless purchases were made in five European countries (France, Poland, Slovakia, Spain, Turkey and the United Kingdom);
purchases accruing to an overall average of approximately 15 euros per transaction. These values represent a six-fold growth from July 2012 (Visa Europe News, 2013). In fact, US research company Marketandmarkets forecasts that global NFC market revenues will reach 10 billion dollars by 2016 (Bodhani, 2013).

Modern-day smart phones and other mobile devices by design are equipped with NFC functionality to foster contactless payment transactions. Be that as it may, it is no surprise that even today, a lot of individuals seem unaware of the contactless technology as a survey conducted by Ceipidor, et al. (Ceipidor, et al., 2012a) in 2012 revealed that out of a total of 1001 respondents, about 91% indicated unfamiliarity with NFC and 60% were indisposed to the utilization of NFC for mobile proximity payments. Regardless of the numerous applications contactless and NFC technologies promise to offer, there exists rising concerns about the security risks involved, reflecting in the overall level of confidence in the technology from its consumers.

What do we mean by the term contactless or NFC and what standard(s) govern the technology? What are the applications and benefits derived? Are there any associated security risks? If yes, are there remedies? These are the core perspectives this paper seeks to elaborate.

CONTACTLESS AND NFC TECHNOLOGY

Contactless Smart Card – Definition, Composition and Characteristics

A contactless smart card is an Integrated Circuit Card (ICC) comprising a transponder (an inductive antenna plus microchip) that allows energy flow between the card and the interfacing device without direct physical contact but by electromagnetic (EM) induction or high-frequency transmission technique through an RF interface (Inside Contactless, 2004). It is a type of card which can communicate with other devices without any physical connection using RFID principles (Kfir, et al., 2005; Finkenzeller, 2010). The following features make a contactless smart card (Hanke, 2005; Inside Contactless, 2004):

- Without battery but powered by magnetic RF field.
- Operates at 13.56 MHz and interacts with a contactless reader through load modulation (EM coupling).
- Uses different types of load modulation/coding – Type A (847 KHz subcarrier at 106 Kbps On-Off Keying OOK + Manchester coding: card modulation, 100 % Amplitude Shift Keying ASK + modified Miller coding at 106 Kbps: reader modulation) and Type B (847 KHz subcarrier at 106 Kbps Binary Phase Shift Keying BPSK + Non-Return to Zero NRZ coding: card modulation, 10 % ASK + NRZ coding at 106 Kbps: reader modulation).
- Resonant frequency range of 14 – 19 MHz and operates typically within the range of 0 and 10 cm.
- Data transmission (to and fro reader) is by half-duplex (communication in both directions, but only one direction at a time) field modulation at 106 kbps.
- Compliant with ISO/IEC 14443 standard.

Throughout the paper, the terms card and reader (terminal) shall mean contactless- PICC and contactless-PCD respectively except stated otherwise.

The Contactless System – Operational Principle

There are 2 main components of a contactless system depicted in Figure 3 namely: the Proximity Coupling Device (PCD); also known as contactless reader or terminal and the Proximity ICC (PICC); also called a transponder or contactless card. The PCD basically is an antenna connected to an electronic circuit. The PICC equally consists of an inductive antenna and an IC connected to its ends. The PCD – PICC combination in terms of behaviour is analogous to the working principle of a transformer, whereby an alternating current (AC) passes through a primary coil (PCD antenna) and generates an EM field, which induces a current in the secondary coil (PICC antenna) whenever they are in proximity. In other words, inductive coupling takes place (EMV Book D, 2013). The air space between the transformer’s primary and secondary coils is known as air core (Inside Contactless, 2004). The PICC transduces the EM (or RF field) transmitted by the PCD into a DC voltage by means of a diode rectifier, and utilizes the latter to power the internal circuitry of the PICC. The configuration and tuning of both antennas determines the coupling efficiency from one device to the other; if in parallel might yield increase. Once the PICC receives the signal, it decodes it, and responds back to the PCD (EMV Book D, 2013).

Every time the PICC receives the current, it demodulates and recovers the data, at the same time uses the transmitted power to activate its circuitry; in other words, the RF energy transmitted by the PCD and received by the PICC does not only power up the PICC but is also used to transport the data through modulation of the carrier. Therefore, both data and power transfers occur in the PCD – PICC communication loop. The technique used by the PICC to achieve this is called load modulation (the point where the PICC changes the load or resistance which is sensed by the PCD) and how strongly or weakly this occurs is called load modulation sensitivity (EMV Book D, 2013). Load modulation is based on the EM coupling between PICC and PCD. If the PICC changes the current in its antenna by a small margin, then a corresponding small variation is sensed by the PCD antenna, which in turn translates to a small increase in voltage across a resistor in series with the PCD antenna (Inside Contactless, 2004).
Evolution into NFC

Contactless technology blends elements from extensive areas including EMC, RF, manufacturing and semiconductor technologies, data security and cryptography and telecommunications; the underlying technique for the technology is RFID which implies information is being conveyed by radio waves normally over short range (Finkenzeller, 2010). In the contactless payment system (Figure 3), popularly labelled the “wave-and-pay” system, the PCD will be a contactless reader, while the PICC will be a contactless card (Bodhani, 2013). This could be a standard card of ID-1 size (ISO/IEC 7810), or a mobile product including micro SD card, tablet or smart phone with NFC functionality. To complete a contactless transaction, the card is simply waved or tapped in the vicinity of a reader. The contactless symbol for reader and card appear at the bottom of Figure 3. The elements of the reader symbol consist of the contactless indicator, a hand holding a standard contactless form factor device and an egg-shaped key line that wraps the duo. Dimensionally, the symbol has an aspect ratio of approximately 1.71 (EMV Book D, 2013).

As contactless systems continued to advance, it became crucial to ensure interoperability between different wireless communication methods. This led to the concept of NFC (NFC online, 2013). NFC is an amendment to already existing contactless smart card technologies promoted and governed by the NFC forum; a non-profit industry association that supports the use of NFC short-range wireless communication in mobile devices and end user electronics. NFC devices are able to communicate with contactless cards or other NFC devices in a peer-to-peer (P2P) or half duplex mode similar to the PCD – PICC transformer-like operation (NFC Forum, 2014). One device acts as the master (initiator or the powered device) and the other as slave (target or tag). While the former primarily starts and dictates the flow of communication, the latter basically sends responses via load modulation procedure. By default, an NFC device stays in the target mode, awaiting orders from initiator (Madlmayr, et al., 2008a). While any mobile device having the capability of initiating and making a payment transaction via RFID means uses contactless technology, it does not necessarily mean that all mobile payments are contactless payments. For instance, a text message may be required to complete a mobile bank account transfer transaction. The amalgamation of mobile payments with contactless payments typifies how NFC provides the finest of both worlds (NFC Forum, 2014). NFC can be viewed as a subset of RFID having shorter communication range apparently for security motives (NFC online, 2013). EMV Contactless Mobile Payment, (EMV Contactless Mobile Payment, 2011) defined NFC as “a short range contactless proximity technology based on ISO/IEC 18092 (P2P), which provides for ISO/IEC 14443 compatible communications and enables devices to communicate with each other when brought into close range”.

NFC – Why it is so-called

Every NFC-enabled device has an antenna. By definition, an antenna is essentially a device for radiating and/or receiving EM waves designed to have maximum radiation with its active part being a multiple or fraction of the wavelength it is transmitting or receiving (Saunders and Aragón-Zavala, 2007). When an antenna is in operation, three field regions or zones exist namely: reactive, near field (Fresnel zone), and far field (Fraunhofer zone) regions. The reactive region, also called the reactive near field is the space immediately surrounding the antenna. Very close to an antenna, the field patterns change very rapidly with distance and include both radiating energy and reactive energy which oscillates towards and away from the antenna (Saunders and Aragón-Zavala, 2007). Beyond the reactive region, radiating fields begin to dominate within a particular distance (usually less than or equal to one wavelength from the transmitting antenna) since the reactive fields are now negligible; this is called the near field region, also known as radiating near field region. This region and the distance it occurs is where NFC technology is designed to operate, hence the reason for the name NFC (NFC online, 2013). Beyond the radiating near field region lies the far-field region where the wave front appears very closely as spherical waves. It begins approximately two wavelengths from the antenna extending outwardly. There also exists the transition zone in between the near and far field regions. In every region, EM field intensity is inversely proportional to the square of the distance the field is from the antenna (Fraunhofer zone) regions. The reactive region, also called the reactive, near field (Fresnel zone), and far field regions to it. Part 1, 2, 3, and 4 defines the physical characteristics, RF power and signal Interface, initialization and anticollision, and transmission protocol respectively (Smart Card Alliance, 2013). However, contactless card standards exist in three categories namely: ISO/IEC 14443 (Proximity ICCs), ISO/IEC 10536 (Close Coupling ICCs) and ISO/IEC 15693 (Vicinity ICCs). The trio only differs in operating distance between the card and the reader antenna; about 1cm, 10cm

Subsequently in the paper, the terms contactless and NFC shall be used interchangeably except otherwise specified.

Contactless/NFC Standards and Payment Schemes

The standard for contactless PICCs is ISO/IEC 14443 under the general title Identification Cards—Contactless Integrated Circuit Cards—Proximity cards. There are four parts to it. Part 1, 2, 3, and 4 defines the physical characteristics, RF power and signal Interface, initialization and anticollision, and transmission protocol respectively (Smart Card Alliance, 2013). However, contactless card standards exist in three categories namely: ISO/IEC 14443 (Proximity ICCs), ISO/IEC 10536 (Close Coupling ICCs) and ISO/IEC 15693 (Vicinity ICCs). The trio only differs in operating distance between the card and the reader antenna; about 1cm, 10cm
and 100cm for close coupling, proximity and vicinity ICCs respectively (Smart Card Alliance Report, 2002). Of all the three, when we refer to contactless technology standard, invariably means the ISO/IEC 14443 standard as this fully defines the contactless communication protocol (Hancke, 2005). Its main features include operating frequency of 13.56 MHz, and speed of 106 Kbps, 212 Kbps, 424 Kbps, or 848 Kbps. Up to 3 Mbps might be achieved as the technology advances. Other capabilities include up to 64 KB storage memory, crypto processors as well as wired logic mechanism for security purposes. All the three contactless technology standards utilize the application-level standards defined in ISO/IEC 7816 and ISO/IEC 7810 (contact card standards) because they define the structure of commands sent to smart card, data and file organizations, security mechanisms, application identification, inter-industry data features, card life phase management and smart card thickness or size (Smart Card Alliance Report, 2002).

Like the contactless card standard, a major specification underlying the NFC technology is ISO/IEC 14443. Additionally, NFC is based on ISO/IEC 18092 and ISO/IEC 18000-3 which specifies the RFID communication used by NFC devices by providing physical layer, collision management system and protocol values for RFID systems for item identification operating at 13.56 MHz (NFC online, 2013). More specifically, an NFC device can operate in three modes: card emulation mode (ISO/IEC 14443), P2P mode (ISO/IEC 18092) and Read/Write (R/W) mode (ISO/IEC 18000-3) (NFC Forum, 2014).

The smart card payment bodies governing smart card debit/credit transactions around the globe include Visa, American Express (AMEX), MasterCard, UnionPay, Discover, JCB and others. The named sextet make up members of the public corporation called EMVCo (EuroPay MasterCard Visa and Co). EMVCo work in tandem with accredited laboratories such as UL (Underwriters Laboratories), reader and card manufacturers to ensure smart cards and terminals deployed all over the world are compliant (environmentally safe, cryptographically secure, interoperable, and acceptable) to EMV specifications for payment transactions (EMVCo, 2014). Visa, MasterCard and AMEX are big players in the electronic payments industry of the modern era.

APPLICATIONS AND BENEFITS OF CONTACTLESS/NFC TECHNOLOGY

Applications

There are myriads of applications derivable from contactless/NFC technology (Finkenzeller, 2010; Inside Contactless, 2004; Walko, 2005; Smart Card Alliance Report, 2002; Özdenizci et al., 2010). For simplicity, the applications have been categorized according to Gemalto (Gemalto, 2013) into three areas expanded below:

- **Information Technology**: data encryption, secure log on, user authentication, storage of digital certificates etc.
- **Wireless**: inter-handover roaming, secure subscriber authentication, mobile value added services.
- **Business**:
  - Banking and electronic payment – payment transactions over intelligent readers.
  - Loyalty and promotions – smart posters and tickets for shopping, concerts and movies.
  - Access control – logical (accessing a network or system) and physical (automotive keyless entry systems for offices and residences).
  - Secure Identification – biometric ID cards, e-passports.
  - Transportation – electronic fare management in ticketing, parking and toll collection.

Benefits

- **Speed and Convenience**: Wave and pay, wave and go, touch and go, tap and go, tap and pay; these intuitively refer to speed and convenience that come with the technology. It feels good when you don’t have to queue up to pay for lunch meals especially during peak periods. If you can make fast payment transactions by simply waving your NFC device around an intelligent terminal, it saves you some waiting time. Better still, why wait for the bar man to bring your bill at the end of your meal if when upon making orders you could use your NFC-enabled mobile device and apps to initiate and complete the payment? Faster payment transactions are evidently achievable with the contactless technology. Oyster card used in London buses is a typical example. With a prepaid Oyster card, passengers can touch in upon boarding the bus without having to distract and pay the driver or bus attendants. A smart card for access control is another great benefit. Imagine unlocking the door to your office or home building by simply flapping or touching your portable card in the vicinity of the receptive NFC device installed on the door compared to inserting a key to unlock the door; it minimizes stress and time. Also, contactless payment offers customers a different method of paying for lower-priced goods, at 33 USD and below. Contactless cards have been issued by debit/credit card issuers such as PayPass (MasterCard) and PayWave (Visa) and adopted by some banks across the world. In recent years, various mobile phone manufacturers such as LG and Samsung have released NFC-enabled handsets to give their customers more seamless experience with mobile value added services (Bodhani, 2013; Gemalto, 2013).
- **Adaptability and Ease of Integration**: A contactless smart sticker (pay tag) attached to a mobile device can turn the latter into a contactless payment device and transactions can be made via NFC. Apps can also be
developed to work in unison with NFC-enabled handsets to complete payment transactions. An example of such is Visa’s PayWave NFC app installed on Samsung handsets; an overall successful trial was recorded during the London 2012 Olympics (Bodhani, 2013). Certainly, the much talked-of and gradually becoming widely utilized Google Wallets, m-Wallets, iWallets, digital wallets, or e-Wallets of very recent times undoubtedly prove that contactless/NFC technology is both adaptable and integrable. NFC is able to connect to other protocols, such as Bluetooth and Wi-Fi, by exchanging the configuration and session data via a P2P connection (Walko, 2005).

- **Intelligence and security:** Contactless cards do not only store information; they can decode, transmit and process data (Finkenzeller, 2010). Smart cards by design come with security features; information stored on the card can be read-write protected or PIN-coded as every smart card is uniquely identified by its own serial number (Gemalto, 2013). ISO/IEC 14443 cards have cryptographic features by design; nevertheless, the security of payment transactions is an ongoing argument in the contactless world. Perhaps this is the reason while contactless payment transactions have been limited to small amounts so as to prevent the potential of fraudulent usage (Bodhani, 2013).

**SECURITY CONCERNS AND POTENTIAL SOLUTIONS**

**Threats to Contactless/NFC Systems**

The primary objective of contactless smartcard technology is to deliver cost-effective no-touch communication, which can generate authenticated and optionally encrypted link of communication between the reader and the card using RFID techniques (Kfir, et al., 2005). The bulk of RFID-related risk is privacy (security) and this has been a growing concern for users of the technology worldwide. A number of attacks along with potential solutions on RFID/contactless/NFC systems are presented in Table 1. The attacks are elaborated below with some practical examples.

- **Relay attack:** Kfir, et al. (Kfir, et al., 2005) found that this type of attack occurs orthogonally and involves the activation of a victim’s card from a distance and sending explored information to an authorized reader in order to complete the transaction. This was also explained by Diakos, et al. (Diakos, et al., 2013). Akin to these, Hancke (Hancke, 2005) investigated that if a contactless card could be sensed and triggered while in a pocket, wallet or purse, then an intruder might be able to leverage that vulnerability by snitching while standing next to or walking past his victim. Since there is no user intervention, it is possible to execute a successful relay attack in order to impersonate, and take advantage. When users utilize NFC payments for movie tickets for instance so as to gain entrance, they are susceptible to this attack as there could be malicious relaying of tickets which inadvertently grants an attacker unauthorized entry (Dodson, et al., 2010).

- **Eavesdropping attack:** Jara, et al. (Jara, et al., 2010) defined eavesdropping in NFC as unapproved listening and communication interruption by collecting raw transmissions between the card and reader, so as to determine protocols and traffic pattern, ultimately collecting vital data for attack. Also, Diakos, et al. (Diakos, et al., 2013) presented a qualitative analysis on how NFC payments can be successfully eavesdropped in the card emulation mode using easily concealable antenna and inexpensive electronics. Performance assessments revealed that eavesdropping on contactless systems depend largely on the EM field generated by the “rogue” PCD; occurring at a distance of up to 90 cm range in a shielded surrounding.

- **Man in the middle (MITM) attack:** According to Dodson, et al (Dodson, et al., 2010), this type of attack is liable to occur whenever there is no authentication between communicating contactless and/or NFC devices. Stallings (Stallings, 2005) illustrated MITM attack as follows: Alice (PCD) thinks she is talking to Bob (PICC) and vice versa; meanwhile, Trudy (attacker i.e. third party with an active tag) intercepts the communications and modifies it unknown to Alice and Bob. In the MITM attack, the invader (Trudy) tries to intercept messages and either relays or substitutes with other messages after carrying out selective modification. If successful, one of the PCD or PICC is masqueraded during communication.

- **Skimming attack:** In this type of threat, an attacker is able to extract information from a vulnerable contactless device so as to exploit the user (Diakos, et al., 2013). Using the contactless system principle of operation, a powered up deceptive transmitter (PCD) can be used to quiz the target (PICC) and extract vital information that are useful in tracking the user of the target device. Skimming can be viewed as an intrusion on user privacy.

- **Malicious tracking or traceability:** In this type of attack, the communicating parties are firstly identified, followed by the time they communicate and how often they do so. This type of attack primarily hacks location privacy (Dodson, et al., 2010).

- **Spoofing and Cloning:** While a spoofing process collects, duplicates and transmits card data (tag) to a reader during operation in an R/W communication mode, cloning goes further to duplicate one tag data to another one thereby successfully making an exploit (Jara, et al., 2010).

- **Power analysis attack:** Messerges, et al. (Messerges, et al., 2002) investigated how the analytical control of power consumption signals by utilizing signal-to-noise ratio (SNR) characteristics can differentially pose attacking threat to smart card security systems.
Potential Solutions

Introduction of an activation button on an NFC device can protect against relay attacks. However, there are other possible attacks that could override relaying (Diakos, et al., 2013). Development of cryptographic mechanisms such as Data Encryption Standard (DES), Advanced Encryption Standard (AES) and Message Authentication Coding (MAC) might be necessary to counteract them.

Two potential solutions to eavesdropping in NFC over P2P mode were identified by Jara, et al. (Jara, et al., 2010) as application of symmetric ciphers and asymmetric ciphers. Both involve sharing keys during contactless interaction and allow intercepted data from being used to identify the user of the NFC device.

In other to eliminate MITM attack, it is imperative to introduce cryptographic safeguards in between the PCD and the PICC interactions based on authentication (Dodson, et al., 2010). In some cases, the use of digital signatures might be required (Stallings, 2005). A digital signature is an authentication mechanism that allows the initiator of a message to attach a cryptogram that serves as a signature which ascertains the source and integrity of the message. This can equally be a solution to phishing attack as well as malicious tracking and traceability to ensure privacy (Madlmayr, et al., 2008b).

A possible solution to information leakage called KerNeeS was proposed by Ceipidor, et al. (Ceipidor, et al., 2012b); this is a protocol for mutual authentication between NFC handsets and POS terminals for secure payment transactions. It sufficiently adds a secure level to the EMV protocol in such a manner that guarantees confidentiality and authentication between the card and reader.

Potential countermeasure of DES was highlighted and proposed by Messerges, et al. (Messerges, et al., 2002) as possible solution to power analysis attacks. In principle, NFC uses RFID tags in operation, thus authentication is critical during communication. Proper authentication can be carried out via intrusive (requires password) and non-intrusive (behavioural analysis) methods. An application level security can also be defined using zoning methods to demarcate low, medium-low, medium, medium-high and high security zones. Introducing authorization check points over the communication link providing contactless or NFC transactions is also very useful (Abu-Saymeh, et al., 2013).

To summarize, the implementation of cryptographic mechanisms such as MAC, AES, digital signatures, DES and others into RF chips, PICCs and NFC-enabled devices can help protect from majority of the security threats like relay attacks, MITM attacks, eavesdropping, skimming, phishing and cloning.

CONCLUSION

There is no doubt that contemporary times have witnessed a revolutionary wave of technological transformation in all fields of human endeavour, including smart card engineering; whose current trend of development primarily focuses on contactless payments systems. It has now become a global reality that smart card evolution is highly promising with the provision of great benefits to the society; resulting in a continuous expansion of its scope, application, deployment and utilization across the world. This paper has presented a concise but insightful background on smart card technology by emphasizing contactless and NFC systems. Useful information regarding the technology standards, applications, benefits, and security issues alongside their potential solutions have been proffered. Clearly, there are great benefits derived from the numerous applications of contactless smart card and NFC systems, however the future of the technology is somewhat uncertain; due to general lack of confidence in its security; and inadequate awareness cum low hi-tech savviness on the part of its users. Perhaps, a technology might not become classified as top-notch until it has satisfied the four essential requirements of modern cryptography namely confidentiality, authentication, integrity and non-repudiation. Surely, time will tell.

REFERENCES


Dodson, B., Bojinov, H. and Lam, M. S., 2010, “Touch and Run with Near Field Communication (NFC)”, Computer Science Department, Stanford University, USA.


**FIGURES AND TABLES**

![Image 1: Evolution of Smart Card in the Payment Industry](image1)

**Fig. 1 Evolution of Smart Card in the Payment Industry**

![Image 2: Smart Card Classification and Standards](image2)

**Fig. 2 Smart Card Classification and Standards**

![Image 3: The Contactless Technology System](image3)

**Fig. 3 The Contactless Technology System**

<table>
<thead>
<tr>
<th>Threat or attack</th>
<th>Potential solution or countermeasure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relaying</td>
<td>ON/OFF button to activate/deactivate NFC functionality; authentication</td>
</tr>
<tr>
<td>Eavesdropping</td>
<td>Application of ciphers or cryptograms (Encryption)</td>
</tr>
<tr>
<td>Man in the middle</td>
<td>MAC, asymmetric cryptography, use of digital signatures</td>
</tr>
<tr>
<td>Skimming</td>
<td>MAC, DES, AES</td>
</tr>
<tr>
<td>Malicious tracking</td>
<td>MAC, application of digital signatures</td>
</tr>
<tr>
<td>Spoofing</td>
<td>Application of cryptograms (encryption), use of digital signatures</td>
</tr>
<tr>
<td>Cloning</td>
<td>Application of cryptograms (encryption), use of digital signatures, authorization checkpoints</td>
</tr>
<tr>
<td>Power analysis</td>
<td>Authentication, DES, AES, application level security</td>
</tr>
<tr>
<td>Denial of Service</td>
<td>MAC, digital signatures, Encryption</td>
</tr>
<tr>
<td>Phishing</td>
<td>Mutual authentication, application of digital signatures</td>
</tr>
</tbody>
</table>

Table 1 Threats and remedies to contactless/NFC systems
ABSTRACT

Predicting long term rainfall trend is essential for agricultural, industrial and hydrological applications, particularly water resource planning and stormwater management system. Global circulation models (GCMs) are widely accepted climate models in assessing climate change. However, the resolution of GCMs are coarse, thus require downscaling technique to refine their resolution to that of local scale. This paper proposes the Bat neural network (BatNN) to downscale GCM outputs and to forecast long range rainfall scenarios. The performance of BatNN will be benchmarked with two other conventional training functions - Scaled Conjugate Gradient and Levenberg-Marquardt algorithms. From the validation test results, BatNN performed on par with SCGNN and significantly better than LMNN. BatNN predicted in increasing rainfall trend in the future, reaching up to 5% (relative to Year 1961-1990) during the late century.

INTRODUCTION

Climate change is a popular research area due to its widespread influence over natural ecosystems as well as human activities. It can be defined as a change or shift in global meteorological patterns caused primarily by global warming. Based on Intergovernmental Panel on Climate Change (IPCC) 5th assessment report (2013), it can be confirmed visually that the troposphere has warmed since the mid of 20th century. Carbon dioxide concentrations have increased by 40% since pre-industrial times. The main emitter of carbon dioxide has been from fossil fuel emission, followed by net land use change emissions. Human influence on the climate system is evident from increased greenhouse gases, positive radiative forcing, and global warming. IPCC (2013) has attributed the observed increase in global surface temperature from Year 1951-2010, which contributed to an increase of 0.5°C to 1.3°C as anthropogenic. As a result of higher surface temperature, there have been reported changes in global water cycle, reduction in snow and ice, rise of global mean sea level, as well as changes in several climate extremes.

In Malaysia, the influence of climate change has led to the change of rainfall and increase in surface temperature. This is evident through the increased flood occurrences and prolonged wet-spell duration during the Northeast Monsoon (November to March). Looking back at the history of Malaysia, the country has been constantly troubled by flash floods throughout the year, especially during the end and beginning of the year (December and January). Flash floods are mostly the after effects of heavy rainfall over a short period of time, which typically exceeds the maximum design of stormwater runoff of the drainage system. Most of the worst floods in Malaysia were flash floods due to the extreme heavy rainfall that not only caused severe damages to infrastructure but also claimed thousands of lives. Some of the worst floods recorded were: January 1971 - Kuala Lumpur; March 2006 - Shah Alam; December 2006 - Johor; January 2007 - Johor; June 2007 - Kuala Lumpur; December 2007 - East and South Coast of Peninsular Malaysia (Kelantan, Terengganu, Pahang and Johor); November 2007 - North of Peninsular Malaysia (Kedah and Perlis); and December 2012 - East and South Coast of Peninsular Malaysia (DID, 2013). As a nation that depends on water resource availability, changes to rainfall regimes are critical and should be addressed appropriately.

In order to estimate the extent of climate change impact on local weather, particularly rainfall pattern, it is essential to forecast future climatology. By far, the best way to achieve such deed is through Global Circulation Models (GCMs). GCMs are widely acknowledged tool for climate study as they model the responses of climate systems according to possible scenarios of different greenhouse gases concentration. GCMs produce climate variables (otherwise known as predictors), such as: air humidity, sea level pressure, temperature, and characterise them into numerical numbers. However, in order to incorporate the whole globe, these climate variables are integrated over coarse spatial scales, in the order of 100km, which are much larger than local atmospheric variables (usually 10km or less) that influence the rainfall processes (Wilks, 2012). Therefore it is necessary to process GCM outputs into finer resolution. One of such method is through statistical downscaling.
There are three types of statistical downscaling technique currently used by researchers: weather typing; regression method; and stochastic weather generators. Weather typing classifies predictand into weather categories by constructing indices of airflow. It derives the frequency distribution of local climate by weighting the local climate states with relative frequency of the weather categories (Conway et al., 1996). Regression methods quantify relationship between the predictors and predictand through linear or non-linear relationship. Stochastic weather generators utilise regression residuals and/or probabilistic state transition to emulate the statistical attributes of local variables. It generates random numbers which resembles observed local climate data (Wilks, 1999). For this study, neural network (NN), a regression type downscaling technique, will be used. NNs are commonly recognised as a ‘black box’ capable of distinguishing inherent patterns from a large amount of data (Hall et al., 1999). A basic architecture of NN model is illustrated in Fig. 1. It consists of an input layer, connected to a hidden layer where the training phase is conducted, followed by an output layer. NN is a viable alternate approach for non-linear statistical forecasting as it is less computational demanding and consume less time. Moreover, NNs are proven by Hornik et al. (1989), Musavi et al. (1994), Huang (1996) and Kak (1998) for their generalisation and learning ability, even with incomplete data input.

NNs normally require training function to calibrate weights and biases of data during the training/learning phase. Conventional training functions are back-propagation (BP), Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) optimisation method. Recently, metaheuristic algorithms have gained recognition in solving optimisation related problems. Osman and Laporte (1996) define metaheuristics as an iterative process which guides a subordinate heuristic through search space exploration and exploitation, by using learning strategies to process information for the purpose of finding a near optimal solution. Hence the motivation of this paper is to implement a relatively new metaheuristic algorithm – Bat algorithm, developed by Yang and Gandomi (2012), into NN models to downscale GCM predictors for rainfall forecasting purposes.

Objective

The objective of this paper is to implement Bat algorithm as a training function for neural networks, henceforth named as BatNN, and use it to forecast long range rainfall scenario, from Year 2011 to 2100. As benchmark, the performance of BatNN will be compared to NN models trained using SCG and LM methods. The former will be known as SCGNN, the latter as LMNN throughout this paper.

Study Location and Research Data

Kuching City is the capital city of Sarawak, Malaysia with an estimated land area of 431 km². The city is located at latitude 1°33'N and longitude 110°20'E. Kuching receives an average rainfall of 4096mm annually (from annual rainfall data for Year 1958 – Year 2010).

There are two distinct seasons in Malaysia, namely Southwest Monsoon (May to September); and Northeast Monsoon (November to March). The former normally signifies drier weather; while the latter usually carries heavy rainfall. Kuching typically experiences maximum rainfall during January; while minimum rainfall events occur during June and/or July (MOSTI 2013). The maximum daily rainfall can reach up to 485.4mm; while the minimum is recorded to be 77.8mm, with 186.22mm being the average daily rainfall. As a result of high daily rainfall, Kuching is susceptible to floods, especially during the Northeast Monsoon season.

In general, there are two types of data input for downscaling technique, namely the predictand (to be predicted) and predictor (used to predict), for the training phase of NN. The predictand data are historical monthly rainfall data of Kuching, which was obtained from Kuching Airport Rainfall Station. Predictor data are GCM outputs in numerical form, and are available via IPCC website as an open source for researchers. GCM predictors were produced based on several hypothesized scenarios, such as population growth and technological advancement. The scenarios are the guidelines that define the concentration of greenhouse gases in the atmosphere which affect the intensity of atmospheric variables.

For this study, predictors from scenario 20CM3 and A2 were used. 20CM3 is constructed based on the climatology of the 20th century. A2 is based on the storyline of a future with higher emission of carbon dioxide (CO2) due to increased population, intensified land use and slower technological advancement. The climate-carbon cycle feedback in Scenario A2 was projected to cause 2°C to 5.4°C increase in global temperature with 0.23m to 0.51m increase in sea water level (IPCC, 2007). There are research centres that provide GCM data for public use. GCM data from Max-Planck-Institut for Meteorology (MPI-M), Germany were selected for this study as the resolution of MOI-M is the highest – latitude 1.9° and longitude 1.9°, or 140 x 140km. NCEP (National Centers for Environmental Prediction) reanalysis data was not used as all variables for the grid box are zero. Hence 20CM3 dataset is used.

METODOLOGY

Bat Optimisation Algorithm

Bats are mammals of the order Chiroptera with flight ability. As most of them are nocturnal – active during the night, they employ sound (echolocation) as a device for communication, navigation and during prey hunting (Richardson, 2008). Bat algorithm is a nature inspired
algorithm by Yang and Gandomi (2012) that bio-mimic the echolocation ability of bats. When encountering small insects (preys), bats will adjust the frequency of sonar to a higher rate and reduce its loudness to a constant minimum. (Møhl, 1988) (Surlykke et al., 2009). Yang and Gandomi (2012) proposed several rules in developing the algorithm:

- Bats use echolocation to sense distance and have the ability to differentiate between prey and obstacles.
- Bats fly with random velocity, \( v_i \) at position, \( x_i \). Their sonar have a fixed minimum frequency, \( f_{\text{min}} \), with varying wavelength, \( \lambda \), and loudness, \( A \), when searching for prey. Bats have the ability to adjust its sonar frequency and the rate of pulse emission, \( r \), in relation to the proximity of their target.
- The loudness of sonar is assumed to vary from a large \( A_0 \), to a minimum but constant \( A_{\text{min}} \).

The following pseudo code by Fister et al. (2013) describes how Bat algorithm works:

```
Objective function \( f(x) \), \( x=(x_1, \ldots., x_d)^T \)
Initialize bat population \( x_i \) and \( v_i \) for \( i=1\ldots n \)
Define pulse frequency \( Q_i \) \( \in [Q_{\text{min}}, Q_{\text{max}}] \)
Initialize pulse rates, \( r_i \) and the loudness, \( A_i \)
While (\( t < T_{\text{max}} \)) // number of iterations
  Generate new solutions by adjusting frequency
  Update velocities and locations/solutions
  If (\( \text{rand}(0,1) > R_i \))
    Select a solution among the best solutions
    Generate a local solution around the best solution
  End if
  Generate a new solution by flying randomly
  If (\( \text{rand}(0,1) < A_i \) and \( f(x_i) < F(x) \))
    Accept the new solutions
    Increase \( r_i \) and reduce \( A_i \)
  End if
  Rank the bats and find the current best
End while
Postprocess results and visualisation
```

For this study, \( Q_{\text{min}} \) and \( Q_{\text{max}} \) are 0 and 1.0. Loudness, \( A \), and pulse rate, \( r \), are set to 0.5 respectively. It should be noted that implementing a self-adapting mechanism where \( A \) would be high initially and decreases rapidly once a prey have been found (vice versa for \( r \)), might improve the performance of the model. This is a subject for future research.

**Performance Indicators**

The performance of the models is evaluated through several accuracy indicators, which are: root mean squared error (RMSE), mean absolute error (MAE), mean bias (MB), and correlation coefficient (R). RMSE is the square root of the sum of each squared errors over \( N \) - total number of outputs. MAE is the summing of absolute error values then divided by \( N \) (Willmott and Matsuura, 2005).

- MB shows the mean forecast error of forecast against observed values (Fekete et al., 2004) (Ramirez, 2005). A positive value indicates overestimation and vice versa.

The formula for each performance indicator and the ideal/optimal point are as follows (Elshorbagy et al., 2010):

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}}
\]

\[
\text{MAE} = \frac{\sum_{i=1}^{N}|O_i - S_i|}{N}
\]

\[
\text{MB} = \frac{1}{N} \sum_{i=1}^{N}(O_i - S_i)
\]

\[
R = \frac{\sum_{i=1}^{N}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \bar{O})^2} \sqrt{\sum_{i=1}^{N}(P_i - \bar{P})^2}}
\]

where \( O_i \) represents the observed values; \( P_i \) represents predicted values; \( \bar{O} \) is the mean of observed values; and \( \bar{P} \) is the mean of predicted values.

**Research Procedures**

The input for NN models were GCM predictors and historical rainfall data over Year 1961-1990. A schematic of research procedure for this study is shown in Fig. 2. The experiments consist of simple feedforward NN (FNN) trained according to Bat, SCG and LM optimisation method. These three NN models will be validated with observed monthly rainfall data from Year 1991-2000 (test 1), and Year 2001-2010 (test 2). The NN models were evaluated through performance indicators stated in previous section. Additionally, the monthly percentage error of the simulated rainfall was compared with observed historical rainfall for test 1 and 2.

All models were trained with lagged data to incorporate temporal recognition capability into NN by feeding it with antecedent data (Saharia and Bhattacharjiya, 2012). This is because simply training NNs using data that are temporally constrained will induce failure in the network to distinguish time related pattern (Dibike and Coulibaly, 2006).

**RESULTS AND DISCUSSIONS**

The predictors used for this study were: ts (surface skin air temperature), tas (2m surface air temperature),
ta200 (air temperature at 200hPa), ta500 (air temperature at 500hPa), and i850 (air temperature at 850hPa). They were chosen based on their high correlation with local observed rainfall data. Through trial and error method, the optimal hidden nodes and training iterations (epochs) for each NN were found. In this case, BatNN and SCGNN shared similar parameters; they were trained with 250 hidden nodes for 1000 epochs; while LMNN was trained with 8 hidden nodes for 1000 epochs.

Validation test models were conducted by comparing the observed rainfall with simulated rainfall of the three NN models. It can be observed that BatNN displayed great potential during validation test 1 (Fig. 3) and test 2 (Fig. 4). Based on the score of performance indicators, BatNN was on par with SCGNN, while it is significantly better than LMNN.

In order to determine implicitly the accuracy of each model, the monthly percentage error was inspected. Fig. 5 and Fig. 6 show the comparison of BatNN, SCGNN and LMNN according to their monthly percentage error. As a whole, the monthly percentage error (margin error) for BatNN ranges between +29% to -23%; SCGNN’s was between +31% to -24%; while LMNN’s was between +29% to -25%. Again here, BatNN shows that its margin error is lower than the other two models. The margin errors were mostly concentrated on the mid-year (May, June, July, August and September). Upon closer investigation, these singularities can be attributed to the nature GCM’s predictor data. In this case, the predictors are assigned a much higher value, which induce NN models to identify it as an important factor hence assigned a higher weight and bias value.

Simulated future rainfall were subdivided into three decadal periods, which are 2030s (Year 2011 – 2040); 2060s (Year 2041 – 2070); and lastly 2080s (Year 2071 – 2100). In order to provide a better perspective of how the future rainfall changes in relative to current rainfall pattern, the observed rainfall from Year 1961-1990 (denoted as OBS1970s) was be used as base period for comparison.

Table 1 records the forecast by BatNN and its percentage increase/decrease in comparison to the base period. Fig. 7 shows the equivalent graphical representation. According to the forecast, annual mean rainfall during the 2030s and 2060s will increase by 4% (relative to the base period); while by the late century (2080s), an increase of 5% can be observed. Interestingly the forecast show major increase of rainfall volume during June and July, especially during the 2080s (24% and 19% increase for both months respectively). This can be attributed to the margin error of NN models previously discussed.

CONCLUSIONS

This paper explored the possibility of implementing metaheuristic algorithm, in this case, Bat algorithm, as a training function into NNs. Two other widely used training functions – SCG and LM were used as benchmark. Based on the results, the performance of BatNN was on par with SCGNN; and proved to be better than LMNN. BatNN showed high confidence level during the validation stage, with correlation coefficient r value ranging between 0.94–0.96 for both tests.

Percentage errors of each NN models were also observed on a monthly scale during the validation stage. It was found that the margin error of BatNN was less than that of SCGNN and LMNN.

From the future rainfall forecast by BatNN, an increase of annual mean rainfall of 4% during the 2030s and 2060s; and 5% during 2080s was shown. The forecast revealed that June and July will receive major rainfall increase.

REFERENCES


DIDS, 2012, Recorded Flood Events, Department of Irrigation and Drainage (Sarawak).


Fister, I., Fister, D. and Yang, X., 2013, “A hybrid bat algorithm”.


MOSTI 2013, *Seasonal Rainfall Variation in Sabah and Sarawak*.


---

**FIGURES AND TABLES**

**Fig. 1 Basic architecture of NN**

![Basic architecture of NN](image1)

**Fig. 2 Procedure of training a NN model**

![Procedure of training a NN model](image2)

**Fig. 3 Performance of BatNN, SCGNN and LMNN for test 1 (Year 1991-2000)**

![Performance of BatNN, SCGNN and LMNN for test 1 (Year 1991-2000)](image3)
Fig. 4 Performance of BatNN, SCGNN and LMNN for test 2 (Year 2001-2010)

Fig. 5 Monthly percentage of BatNN, SCGNN and LMNN for test 1 (Year 1991-2000)

Fig. 6 Monthly percentage of BatNN, SCGNN and LMNN for test 2 (Year 2001-2010)

Fig. 7 Simulated future rainfall by BatNN

Table 1 Future rainfall forecast by BatNN and percentage difference (in comparison with OBS 1970s)

<table>
<thead>
<tr>
<th>Month</th>
<th>OBS 1970s</th>
<th>2030s</th>
<th>% difference</th>
<th>2060s</th>
<th>% difference</th>
<th>2080s</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>21.24</td>
<td>21.23</td>
<td>0%</td>
<td>20.94</td>
<td>-1%</td>
<td>20.63</td>
<td>-5%</td>
</tr>
<tr>
<td>Feb</td>
<td>17.33</td>
<td>18.50</td>
<td>7%</td>
<td>18.82</td>
<td>0%</td>
<td>18.73</td>
<td>0%</td>
</tr>
<tr>
<td>Mar</td>
<td>11.78</td>
<td>11.54</td>
<td>-2%</td>
<td>11.52</td>
<td>2%</td>
<td>11.94</td>
<td>2%</td>
</tr>
<tr>
<td>Apr</td>
<td>8.99</td>
<td>9.22</td>
<td>3%</td>
<td>9.94</td>
<td>2%</td>
<td>9.92</td>
<td>0%</td>
</tr>
<tr>
<td>May</td>
<td>7.52</td>
<td>7.69</td>
<td>2%</td>
<td>7.32</td>
<td>-3%</td>
<td>7.54</td>
<td>0%</td>
</tr>
<tr>
<td>Jun</td>
<td>6.56</td>
<td>7.52</td>
<td>15%</td>
<td>7.74</td>
<td>18%</td>
<td>8.15</td>
<td>24%</td>
</tr>
<tr>
<td>Jul</td>
<td>5.92</td>
<td>6.78</td>
<td>14%</td>
<td>7.11</td>
<td>20%</td>
<td>7.06</td>
<td>19%</td>
</tr>
<tr>
<td>Aug</td>
<td>6.64</td>
<td>6.89</td>
<td>4%</td>
<td>7.01</td>
<td>6%</td>
<td>7.40</td>
<td>12%</td>
</tr>
<tr>
<td>Sep</td>
<td>8.96</td>
<td>9.13</td>
<td>3%</td>
<td>9.33</td>
<td>5%</td>
<td>9.50</td>
<td>7%</td>
</tr>
<tr>
<td>Oct</td>
<td>10.40</td>
<td>10.82</td>
<td>4%</td>
<td>10.86</td>
<td>4%</td>
<td>10.54</td>
<td>1%</td>
</tr>
<tr>
<td>Nov</td>
<td>12.10</td>
<td>12.34</td>
<td>2%</td>
<td>13.03</td>
<td>7%</td>
<td>12.91</td>
<td>0%</td>
</tr>
<tr>
<td>Dec</td>
<td>15.05</td>
<td>15.47</td>
<td>3%</td>
<td>15.41</td>
<td>3%</td>
<td>15.39</td>
<td>2%</td>
</tr>
</tbody>
</table>

mean: 11.06 11.46 4% 11.51 4% 11.36 5%
TELEMEDICINE PROGRAM FOR MANAGEMENT AND TREATMENT OF STRESS URINARY INCONTINENCE IN WOMEN: PROGRAM DESIGN

Anna Abelló Pla PT¹,², Anna Andreu Povar PT², Jordi Esquirol Caussa PhD¹, Vanessa Bayo Tallón PT¹, Dolores Rexachs PhD,³ Emilio Luque PhD.³

¹. Servei Universitari de Recerca en Fisioteràpia (SURF). Escoles Universitàries Gimbernat. Sant Cugat del Vallès, Barcelona (Spain).
². NuboCare. Barcelona (Spain).

ABSTRACT

Stress Urinary Incontinence is defined as involuntary urine leakage caused by physical activity and/or efforts, and is a highly prevalent pathology among women; this pathology significantly affects their quality of life. Stress Urinary Incontinence treatments are often less effective than expected because they require a conscious effort by the patient to correctly follow the treatment programs, that usually have drawbacks as high costs, time and/or schedule requirements.

ICT mediated Physical Therapy treatment programs can be useful to improve Stress Urinary Incontinence symptoms and pelvic floor function in women while maintaining total confidentiality, with an at home comfy treatment and accomplishing a higher treatment adherence, while keeping a low budget for the patients and saving health systems’ economic resources.

INTRODUCTION

The use of Information and Communication Technologies (ICT) and Telemedicine can help improve the adherence to the Physical Therapy treatment and, additionally, it can improve the possibilities of healthcare providers to monitor if patients follow the treatment correctly (both technically and in the recommended schedule), to control the evolution of the patient and to interact with every patient individually.

Urinary Incontinence (UI) in women is a very important and prevalent health problem that affects quality of life in patients and can be psychologically threatening for them. Its treatment usually consists in Physical Therapy interventions and exercises in groups of affected women (added or not to a medical/pharmacologic management) and to do some individual exercises regularly at home.

This is a quite non affordable program in many cases, due to several reasons and adherence is often quite low.

An ICT mediated domiciliary intervention program for women with UI and need for Physical Therapy intervention is presented.

With this program, Physical Therapists can see and interact with every patient confidentially, receive all medical data from the patient to control the quality of the performed exercise and monitor and evaluate how every patient’s UI is improving day by day.

The aim of this poster is to show how ICT can be applied to conservative Stress Urinary Incontinence (SUI) treatments (Physical Therapy).

BACKGROUND

Telemedicine is defined as the use of telecommunication and information technologies in order to provide clinical health care at distance. Telemedicine and Telehealth are terms that may refer to quite different meanings or definitions but usually are considered to be interchangeable.

Telemedicine has been reported to have huge benefits, as improved access, quality and cost-effectiveness.

Additional benefits are found in the use of Telemedicine: it usually is an inexpensive user-friendly service that allows patients to have a flexible schedule and the possibility of offering health assistance from the comfort of patient’s home (Figure 1).

Patients and costumers want Telemedicine: it reduces travel time and related stress. It shows greater satisfaction and they support its use.

The most common use of Telemedicine is still consultation: patients use this service to consult different healthcare providers about health-related issues.
In Physical Therapy, Telemedicine application began about fifteen years ago.

In 2009, the International Continence Society (ICS), along with the International Urogynaecological Association (IUGA), wrote a joint report where they defined UI as the “complain of involuntary loss of urine” (Haylen et al., 2010). UI causes social and family troubles to the patient too, being nowadays an important health problem that has a high influence in medical, psychological, social, and economic bearings affecting their quality of life. There are three types of UI, based in symptomatic signs (Espuña Pons, Castro Díaz, Carbonell, & Dilla, 2007): Stress Urinary Incontinence (SUI), referred as involuntary urine leakage caused by physical activities, efforts or after simple acts (as sneezing or coughing); Urge Urinary Incontinence (UII), defined as involuntary urine leakage accompanied or immediately preceded by urgency symptoms; and Mixed Urinary Incontinence (MUI), which includes the simultaneous symptoms of both previously defined types.

The large amount of bibliography about UI shows the huge importance that the scientific community gives to this pathology in women. Nevertheless, one of the most important difficulties is establishing its prevalence (Minaire, Sengler, & Jacquetin, 1995; V.A. Minassian, Drutz, & Al-Badr, 2003) that can be established from 17% to 47% (Abelló, Esquirol, Salas, & Bayo, 2009) or more than 50% (Milsom et al., 2009).

One of the most recent published studies (analyzing all types and severities in middle-aged and elderly women in the general population) estimates UI prevalence between 30% and 60% and shows that prevalence increases with age (Milsom et al., 2009).

It has been suggested that the female population tends to underestimate UI as a health problem if it’s slight: it is often a taboo subject and women have the tendency of not talking about it if the issue is not presented in a way that it is very discreet, confidential and direct (Ricatte, 2004).

In order to diagnose and evaluate UI severity, there are two commonly used questionnaires that help achieve the task: ICIQ-SF and ISI test.

Physical Therapy usually is the first therapeutic choice for SUI, because it has no secondary effects and is minimally invasive (Berghmans, 2006) (Agence Nationale d’Acréditation et d’Évaluation en Santé., 200)

Pelvic Floor Muscle Training (PFMT) and Hypopresive Abdominal Technique (HAT) are the two main physiotherapy techniques to treat this disease.

Nowadays, SUI’s evaluation, management and treatment, use to follow an on-site model with Biofeedback techniques and devices.

A Telemedicine-mediated intervention program would aid patients and health providers to control and improve patient’s symptoms of UI, avoiding most of the problems of classical Physical Therapy treatment problems (lack of adherence due to lack of privacy, higher cost and patient trips to the clinic); in addition, physiotherapists can offer the correct treatment to patients located hundreds or thousands of miles away.

**MAIN FOCUS**

A new treatment model has been designed and has been proposed to evaluate, manage and treat SUI, involving the use of ICT to replace on-site appointments, while always trying to maintain and -if possible- improve treatment’s efficacy and quality, while guaranteeing patient’s privacy and confidentiality. This program uses QR codes in the ads, individual live on-line videoconference, using Skype® (http://www.skype.com/; Internet connection, PC and a webcam is required) and a Birdi® device (http://www.birdisolutions.com, Figure 2), a user-friendly biofeedback device with Bluetooth® connection and a specific app (named Birdi Kegel Trainer, Birdi Solutions S.L.; available for smartphones, tablets and computers).

The Physical Therapist will register every patient on the system and program the first measurement, performed with the device and will propose the individual treatment plan considered to be the most adequate for the patient and will explain all the proposed personalized exercises and actions via videoconference (Figure 3).

The patient has individually daily programmed pelvic floor exercises (PFMT) on her Birdi® Kegel Trainer App (Figure 4). All results are transmitted via Bluetooth to the patient’s smartphone, tablet or computer and then to the Physical Therapist’s control program; the Physical Therapist can go over the results of any patient at any time, checking if the patient really does the routines or not. It is also possible to revise patient’s progress and her adherence to the treatment.

Confidentiality can be maintained absolutely in any case if necessary, by not showing one or more (or any) of the group participants to the others, while Physical Therapist can see all of them on the screen.

At any moment, any patient can use different ways of communication with the Physical Therapist to solve her doubts and/or problems that may develop along the treatment (phone, message apps, videoconference, e-mail, etc.). All the obtained results are digitally stored in a confidential individual clinical record.

The progress of patient’s intravaginal pressure is easily seen thanks to the Birdi® device, allowing the Physical Therapist to see and compare the results of all the exercises, day by day (Figure 5). To assess the improvement (or lack thereof) of the symptomatology, patients are asked to monthly answer the questionnaires ICIQ-SF and ISI. Once a month, the patient and the Physical Therapist will have a short individual videoconference appointment to go over the results and to analyze the progress of the treatment program.
Limitations of this model are merely those derived of data line failures, lack of understanding of ICT in some patients or patients that do not have the necessary ICT devices and connections to the Internet at home. When treatment program ends, Birdi® device can remain with the patient to do follow up exercises by her own, or can be returned to the Physiotherapist to be used for another patient by simply changing some expendable parts.

A pilot study has already been started with a first small group of women with SUI. First preliminary results are better than expected while, as expected, adherence to the treatment seems to be really higher than in the classic model of treatment.

Patients feel this program as comfy and friendly; the device is really easy to use. Patients are comfortable with the program and really proud of their progress, showing that this ICT treatment model is really affordable.

CONCLUSION
SUI is a highly prevalent disease that affects quality of life in women. SUI treatment is often less effective than expected due to different circumstances as high costs, time and/or schedule requirements.

ICT mediated Physical Therapy programs can be useful to improve SUI in women while maintaining total confidentiality, with an at home treatment (anywhere in the world), achieving a higher adherence to the treatment, keeping a low budget for the patients and saving health systems’ economic resources, both public or private.

REFERENCES


FIGURES AND TABLES

Fig. 1 ICT contribution advantages in Telemedicine.

Fig. 2 Birdi Kegel Trainer. Remote biofeedback device.

Fig. 3 Remote device & treatment plan.

Fig. 4 Birdi Kegel Trainer APP screenshots.

Fig. 5 Global treatment plan.
A PROPOSED SUPERVISORY WIDE AREA NETWORK FOR THE TRANS-BORNEO SMART GRID

Duaa Fatima S. Khan, Mujahid Tabassum, Manas K.Haldar and Salim S. Maajie
Faculty of Engineering, Computing and Science
Swinburne University of Technology
Sarawak Campus, Malaysia

ABSTRACT

This paper studies a proposed supervisory computer network for a Borneo-wide power grid. This grid was described in detail earlier. The grid included proposed renewable power generation. For the power grid to be smart, an Internet based communication network is considered. The centralized Supervisory Control And Data Acquisition system (SCADA) is replaced by Wide Area Monitoring System (WAMS) comprising of Phasor Measurement Unit (PMU). The communication network runs on Multiprotocol Label Switching (MPLS) topology including core and edge routers which are connected to Local Area Networks (LAN) at points of generation and distribution is connected by Fiber Optic cables to form a Wide Area Network (WAN). The proposed Wide Area Network (WAN) is simulated using OPNET Modeler and compared with the existing WAN used by Sarawak Energy. Renewable Energy monitoring systems at proposed locations are also simulated with the proposed WAN.

INDEX TERMS: MPLS, OPNET, Power grid, PMU, Solar and Wind, SCADA, SDH

INTRODUCTION

Electric utilities worldwide are increasingly interested in grid interconnection. In Sarawak, Malaysia the Energy, Green Technology and Water Minister of Malaysia, Datuk Seri Peter Chin also expressed interest in the idea (Then, 2012) of a Borneo – wide power grid involving Sarawak and Sabah (East Malaysia), Brunei and Kalimantan (Indonesia). Connecting power grids of countries can be beneficial for two reasons. Firstly, power can be traded. Secondly, and most importantly, in the event of a blackout in one country, power can be obtained from other countries. For example, the 2012 blackout of the North Indian power grid was the worst in history (Daniel, 2012). Roads were gridlocked; coal miners were trapped underground, hospitals were plunged into darkness and millions of train travellers were stranded, Bhutan a tiny country with its grid connected to India offered help. However, as evident by the Indian experience, mere connection of power grids was not sufficient. Rapid response to such blackout requires a smart grid. A smart grid enables automatic Fault Location Isolation and Service Restoration (FLISR), which is achieved by using next generation data acquisition and control systems with high sampling rates.

One such technology is a Phasor Measurement Unit, which has gained considerable attention from countries worldwide, including the US, Europe, India, China and Brazil. It provides synchronized measurements at rates of about 30 to 60 samples per second. This increased granularity of data not only gives insight into grid dynamics, but can also open opportunities for deploying new Smart Grid applications to monitor and enhance the operation of the grid. Most of these smart grid applications have high bandwidth and stringent latency requirements (NASPI, 2009). These issues of bandwidth and latency become more and more critical when Phasor Measurement Units are installed throughout a power grid. In order to take advantage of the measurements provided by the Phasor Measurement Unit a fast communication infrastructure capable of handling huge amounts of data movement and providing real-time data delivery is required (Kansal, et al., 2012)

This next generation communication system that will layer on top of the power grids to form a Smart Grid is currently under intense discussion by organizations like NASPI, EPRI and IEEE. In this paper we have made an attempt to design, simulate, and test the efficiency of such a next generation data acquisition and communication system for the Trans-Borneo Smart Grid. Preliminary studies on the power grid are reported in an earlier paper (Maaji, et al., 2013). Incorporation of renewables is also proposed in this paper and is described in greater detail elsewhere (Maaji, et al., 2013).

AN OVERVIEW OF TRANS-BORNEO GRID WITH INCORPORATION OF RENEWABLES

Figure 1 show the single line diagram for the proposed Trans-Borneo Grid, which connects the power grids of Sarawak, West Kalimantan, Brunei and Sabah. Electrical buses are numbered from 1 to 13 for easy reference and alphabets are used to represent substations. Dotted lines are used to
represent non-existing solar and wind power plants. High Voltage DC (HVDC) links are used at the borders of Sarawak, to transport electricity to Sabah, Brunei, and Kalimantan i.e., bus 1, 6 and 13 respectively. This connection enables bi-directional power transmission between the countries (Wang, et al., 2010). It also allows asynchronous interconnections and protects the system from cascaded ac-systems outages, due to the problems that might be initiated across the border. Bus 3 and bus 7 connect wind farm and photovoltaic arrays to the power grid. HVDC transmission lines are also used to carry the wind farm energy collected from the coast of Bintulu to Simalajau substation (E). The remaining part of the grid consists of 500kV three phase transmission lines and 275kV also HVAC from A to 1, 5 to D and J to K. Grid distances are calculated using Google map.

Wind and solar plants are also proposed to be incorporated into the Trans Borneo Grid. Based on the solar data (Jakhrani, et al., 2013), Limbang, Sarawak with average solar radiation of 17MJ/m² per day is selected to be the choice of sites for PV plant. Based on the same source, Bintulu with the wind power density of about 3.9W/m² was identified to have the highest wind power density in the entire region. We have proposed an offshore plant near the coast of Bintulu as wind speeds are greater in open water than on land and exhibit less shear and turbulence.

### DATA ACQUISITION AND CONTROL SYSTEM

#### Why Has SCADA Failed?

Even though SCADA has managed to provide utility operators with the necessary data to monitor the power grids since the late 1960s, the growing number of outages occurring in the grids worldwide suggests that SCADA systems have now reached their limits. One prime example will be the 2003 USA - Canada blackout which resulted in 50 million people being plunged into darkness for up to two days, 11 deaths and an estimated cost of $6 billion (Minkel, 2008).

As a consequence, major power utilities have shifted their focus on deploying Phasor Measurement Units (PMUs) for Wide-Area Monitoring Systems (WAMS). Unlike PMUs which provide a higher resolution, SCADA systems observe grid conditions after every 4 - 6 seconds, fail to monitor key indicators such as phase angles and hence cannot provide grid operators real-time wide area visibility.

Due to their low sampling rate, SCADA systems are also unable to reveal new information about transient stability events on the grid. This is evident from the Figure 2, which compares SCADA data to phasor data collected for a same event on February 7, 2010 in Florida (NERC, 2010). It can be noted that phasor data not only revealed the system dynamics, but also provided accurate information on frequency excursions as compared to SCADA data.

Furthermore, an investigation report on the 2003 USA-Canada blackout hypothesized that if a phasor system had been in operation at that time, the blackout preconditions, in particular the growing voltage problems in Ohio, could have been identified and understood earlier in the day (NERC, 2010).

Like the majority of the power grids, Sarawak Grid also relies on SCADA systems for monitoring and data acquisition purposes. Approximately 100 Remote Terminal Units are currently installed at critical nodes in the field to monitor the stability of the power system (SESCO, 2002). However, despite the presence of SCADA, a total of 6,338 supply interruptions were experienced in Sarawak during 2010. The breakdown of the causes of these supply interruptions is shown in Figure 3.

On considering the above facts, it is evident that in order to realize a Trans-Borneo Smart Grid the Data acquisition system of Sarawak Grid needs to be upgraded. Thus we have proposed a Wide Area Monitoring System (WAMS) that makes use of PMU for grid monitoring and control (Karam, 2009).

#### Phasor Measurement Unit

A synchrophasor is a phasor that is time stamped to an extremely precise time reference such as a GPS clock as shown in Figure 4. A unit operating on the principle of synchrophasor is known as a Phasor Measurement Unit (PMU) (Sadamiak, et al., 2012). PMUs are placed at the transmission substations and measure parameters such as positive-sequence bus, line voltage phasors, line current phasors, bus frequency and line MW/MVAR flows.

#### PMU Data Format (C37.118)

Currently IEEE C37.118 is the globally used standard for synchrophasors interoperability. It allows simple processing of the synchrophasors from different measuring systems and proposes to standardize several synchrophasor reporting rates as shown in Table 1 (Hauer, et al., 2004).

In IEEE C37.118, PMU packet consists of four types of frames: Configuration frame, Data frame, Header frame and a Command frame. Data frame is sent from substation during normal system operation and is most frequently transmitted (Narendra, et al., 2008).

#### Phasor Data Concentrator

After measuring the electric signals, PMU transfer the phasor data over a communication network to the Phasor Data Concentrator (PDC) (Armenia, et al., 2010). The PDC performs concentration, collection, correlation and
synchronization of PMU data. The PDC data can also serve as an input to SCADA systems (Wang, et al., 2007).

Smart Grid Applications Provided by Phasor Data

Phasor data can also be utilized in numerous Smart Grid applications; some of the most important are as mentioned as follows (Agarwal, et al., 2011):

1- State Estimation: State estimation is an essential tool utilized in the power industry to eliminate effect of inaccurate measurements on the final calculation of the state. Most of the Energy Management System (EMS) applications are fed from state estimated data. Research has revealed that if PMUs are installed at 1/3rd of the power system locations, the state of the entire power system can be measured instantaneously.

2- Transient Stability: A typical power system can get transiently instable in approximately 10 cycles. This can be prevented by forming islands within the system or by shedding load/generation using Special Protection Scheme (SPS), also utilized in Sarawak’s grid. PMUs with their high sampling rate can provide real time information regarding the transient stability of a system.

3- Voltage Stability: Voltage instability spreads over time starting from reactive power (VAR) deficient area and can ultimately cascade and lead to a blackout. The problem can be solved if the voltage in an area can be measured and corrected in time by balancing VAR in the particular area or by islanding the area.

4- Post-Mortem Analysis: This is a key application to correct power system models and to update engineering settings for the system. The engineering settings are bound to change as the system changes.

WAMS Architecture

The architecture of a Wide Area Monitoring System (WAMS) in a power grid is shown in Figure 5 (Jitlikhit, 2010). The architecture consists of three layers:

1- Measurement Layer consists of PMUs connected to the current or voltage transformer.

2- Data Collection Layer consists of PDCs to collect and synchronize data streamed in real time from PMUs grid wide.

3- Application Layer consists of tools, in the utility LAN, for the conversion of PMU data into real time information for grid operators.

PMU Latency Requirement

Latency requirement depends on the particular application and the kind of system response it deals with. The latency requirement of PMUs for WAMS was estimated to be in the range of 30 – 100ms (Selinc, 2012). The latency requirement for the above mentioned smart grid applications along with their data type are shown in the Table 2 (Kansal, et al., 2012).

COMMUNICATION NETWORK

Challenges for SDH Networks

Electric utilities are responsible for maintaining power delivery and controlling grid equipment during all circumstances, even when public service provider networks are congested or an extended power outage occurs. To achieve this goal, most utilities have used Synchronous Digital Hierarchy (SDH) for their Wide Area Network (WAN). Sarawak Energy has also deployed a SDH communication network for its grid which operates at STM4 (622 Mbps) and STM1 (155 Mbps) level to link the four Dispatch Centres (Kuching, Sibu, Bintulu and Miri), major Power Stations, EHV Substations and Regional Offices (SESCO, 2002).

In the past, SDH was considered as the most suitable solution as it delivers carrier-class performance, supports the deterministic traffic critical for grid operations, is straightforward to deploy initially and provides more security than the shared IP networks (Cisco Inc., 2013). However, with reference to smart grid communication requirements the disadvantages of SDH network have become obvious.

In SDH, bandwidth provisioning is fixed as the circuits are established in a static configuration. This approach is inefficient as it reserves bandwidth for a particular application, whether it is active or inactive. As a particular application can only use its predefined bandwidth, when new IP-based services are being integrated over a common infrastructure with SDH, the network is unable to give priority to mission critical applications by allocating them the required additional bandwidth. Deployment of smart grid applications will also increase communications traffic which will in turn increase the bandwidth consumption. As in SDH networks bandwidth increases in increments of 1.5 Mbps or 2 Mbps hence smart grid applications such as WAMS, Advanced Metering Infrastructure (AMI), distributed generation monitoring systems, distribution automation etc. will quickly exhaust the network’s available bandwidth. Eventually the utility will be required to increment its bandwidth which is not cost effective. Equipment for SDH networks has also become obsolete. It is also complex to integrate SDH networks with new systems and operational processes as they are built for specific applications. This inflexibility promotes the deployment of more specialized overlay networks, which further increases the complexity, requires more manual administration and increases maintenance costs.
Considering the above challenges, packet-based systems appear to be a strategic option, which can not only enable utilities to deploy new Smart Grid technologies but also enable their coexistence with traditional operations. One of the most prominent packet-based WAN topologies is Multiprotocol Label Switching (MPLS) which has now been widely deployed for efficiently operating and managing IP networks. MPLS enables Traffic Engineering (TE), Virtual Private Network (VPN) and controllable quality of service (QoS) features by utilizing classification, queue and scheduling (CQS) traffic. In case of link failure, MPLS provides an alternate path and fast reroute in less than 50ms whereas SDH provides an alternate route with 50ms switch over time. MPLS communication network is capable of providing other benefits to utilities such as reliability, flexibility, manageability and maintainability. Furthermore, MPLS is a mature technology and is backed by communication giants such as Cisco and Alcatel Lucent. Cisco provides MPLS solution in its Cisco Grid Blocks™ reference model for Smart Grid (Cisco Inc., 2012). Cisco has also successfully deployed MPLS network for many electric utilities which involves SDH/TDM to MPLS conversions, WAMS, system protection etc.

On considering the above factors, we have proposed MPLS/IP as the WAN topology for the Trans-Borneo Smart Grid (Cisco Inc., 2013).

**Multi-Protocol Label Switching Network**

MPLS enables high performance packet control and forwarding mechanism for routing the packets in the data networks. The main functionality of MPLS is to attach a short fixed-label between Layer2 (Data Link Layer) and Layer3 (Network Layer) of the packet to form Layer 2.5 as shown in Figure 6. This label enables the packet to enter the MPLS domain via the Label Switched Path (LSP) (Jannu, et al., 2012).

MPLS network consists of MPLS core and MPLS Edge routers. The core routers make up Label Switch Routers (LSRs) and the edge routers make up the Label Edge Routers (LERs). The LSRs forward packets based on label swapping mechanism. The LERs are of two types, Ingress routers through which a packet enters into the LSP and Egress router through which the packet leaves LSP. A brief description of the MPLS network is shown in Figure 7.

**SIMULATION OF DATA ACQUISITION AND CONTROL SYSTEM**

OPNET Modeler was used to simulate the proposed Trans-Borneo Smart Grid WAN. The WAN consisted of PMUs installed in seven substations along the 500 kV Trans-Borneo Transmission grid. The PDC was installed as a server component, Wide Area Monitoring (WAM) Server, in the Head Quarters of Sarawak Energy known as State Dispatch Centre (SDC), located in Kuching. The location of the PDC and PMUs is shown in Figure 8. The distance between each substation and between the substations and the SDC was obtained from Google maps.

The main objective of the simulation was to observe if the proposed WAN topology can transfer phasor data to the PDC within the latency requirements of the PMU. In order to compare the proposed and existing Tran-Borneo WAN the simulation was divided into three scenarios.

**Scenario 1: Smart Grid with MPLS/IP Core Network**

In this scenario the WAN consisted of the proposed MPLS/IP core network that connected all the substations to the SDC through fiber optic cables. The installed PMUs were connected to the LERs of the chosen substations, which then linked to the SDC LER via the core network. The core network was made up of LSRs. In order to save cost of using additional routers for the core network, the substation routers that formed the communication structure of the current 275kV Sarawak Grid were utilized as core LSRs. LSPs were created between the substation LER and SDC LER to complete the MPLS network. The simulated model is shown in Figure 9.

As per IEEE C37.118 requirements, all the PMUs were configured to generate 30 packets per second with a packet size of 76 Bytes. TCP/IP protocol was chosen for the network layer and was configured with the default settings. OSPF was used as the routing protocol with single area. Fig. 10 shows the node modal of the Mambong PMU. Background traffic was introduced in the network to represent the existing communications traffic. The background traffic flow was configured to generate ~1,550,000 bits/hr to represent 50% usage of the network.

**Scenario 2: Smart Grid with IGP Routing:**

In this scenario the WAN consisted of the proposed MPLS/IP core network that connected all the substations to the SDC through fiber optic cables. The installed PMUs were connected to the LERs of the chosen substations, which then linked to the SDC LER via the core network. The core network was made up of LSRs. In order to save cost of using additional routers for the core network, the substation routers that formed the communication structure of the current 275kV Sarawak Grid were utilized as core LSRs. LSPs were created between the substation LER and SDC LER to complete the MPLS network. The simulated model is shown in Figure 9.

As per IEEE C37.118 requirements, all the PMUs were configured to generate 30 packets per second with a packet size of 76 Bytes. TCP/IP protocol was chosen for the network layer and was configured with the default settings. OSPF was used as the routing protocol with single area. Fig. 10 shows the node modal of the Mambong PMU. Background traffic was introduced in the network to represent the existing communications traffic. The background traffic flow was configured to generate ~1,550,000 bits/hr to represent 50% usage of the network.

**Scenario 3: Smart Grid with OSPF Single Area Routing:**

The current communication network of Sarawak Grid makes uses OSPF single area as the routing protocol. Hence in order to compare the performance of the proposed and existing WAN, the network in this scenario was configured to route traffic purely based on OSPF routing protocol as shown in Figure 11.
Scenario 3: Addition of Renewable Energy Monitoring System

In this scenario, offshore wind farm and PV monitoring system were added in the MPLS/IP WAN network. The latency requirement of a wind farm monitoring system is 12 cycles or 200ms and that of PV monitoring system is 6 cycles or 100ms. The objective of this simulation was to observe if the proposed network can deliver the monitoring data within the specified latency requirements and also the effect of the communications traffic generated by the renewable energy monitoring systems on the phasor traffic.

Figure 12 shows the simulated Ethernet LAN for Wind Farm monitoring and control placed at the coast of Bintulu. The LAN consisted of eight Wind Turbine (WT) control centres connected to an Ethernet switch, which was connected to the Wind Farm LER shown in following fig. The eight WT workstations were arranged in a tree topology.

The Solar subnet, placed in Limbang, consisted of a PMU and four sensors to detect fluctuations in wind speed, temperature, humidity and irradiance. The four sensors simulated as computer components were connected to an Ethernet switch, which was connected to the PV Plant LER, as shown in Figure 13. The four sensor nodes were arranged in a tree topology. The PMU was directly connected to the PV Plant LER.

Both Wind Farm LER and PV Plant LER were configured with MPLS & static LSP was established between the two and SDC LER to provide connectivity to the WAN. For both the subnets, the Physical Layer used Ethernet 1000BaseX series cable as the transmission link with a data rate of 1Gbps. The Link Layer’s protocol was Ethernet. The applications chosen for the Application Layer for the wind turbine controllers were medium load: database access, load file transfer, remote login and printing. The Applications for the four PV sensor nodes were of (medium load): database access and load file transfer.

RESULTS AND ANALYSIS

The simulation metrics focused in this paper include: LSP Delay, End-To-End (ETE) Delay and TCP Delay. By observing these metrics the latency of the phasor traffic in the WAN Network can be calculated.

Results of Scenario 1:

To accurately analyse the response of PMU traffic, it was essential that each PMU unit would generate 30 packets/sec. Hence the seven installed PMUs would have to generate 210 packets/sec. This can be observed in Figure 14 which shows the traffic received (packets/sec) by WAM server.

Figure 15 shows the LSP delay from all the seven LERs to the SDC LER. It can be observed that all the LSP delays are less than 50ms, which shows the efficiency of deploying MPLS/IP on the core network. It can also be observed that the highest LSP delay is for (Lawas LER – SDC LER). This is because the distance between the Lawas substation and SDC is the largest (~705km). The lowest LSP delay is between Mambong LER and SDC LER as the distance between Mambong substation and the SDC is the smallest (~23,803km).

The ETE delay between the seven PMUs and the WAM server can be observed in Table 3. It can be observed that the highest ETE delay is from Lawas PMU to WAM server, which is 27.97ms and the lowest is from Mambong to WAM server which is 17.85ms. Again, it can be observed that because of deploying MPLS/IP core network the ETE delay is below 50ms.

The TCP delay for WAM server in SDC was observed to be 23.0204ms as shown in Figure 16.

Results of Scenario 2:

Figure 17 shows the TCP delay of WAM server, which has tremendously increased to 115.41ms as compared to the TCP delay of the MPLS/IP core network.

Similarly the ETE Delay between each PMU and the WAM server was also observed to have increased to more than 100ms as shown in Table 4.

Results of Scenario 3:

The addition of renewable energy monitoring system did not affect the LSP and ETE Delay of the PMU units. The LSP Delay of the wind farm subnet was observed to be 15.446ms and that of PV Plant subnet was observed to be 20.66ms as shown in Figure 18.

Table 5 shows the ETE Delay of the Wind Turbines. It can be observed that the ETE Delays are between 23ms – 24ms.

Table 6 shows the ETE Delays of the sensors and PMU placed in the PV Plant subnet. It can be noted that the ETE Delays are ~27ms.

After the addition of the renewable energy monitoring system subnets in the Tran-Borneo Transmission communication network, the TCP delay of the WAM server was observed to be 24.910ms as shown in Figure 19.

From the above results it is evident that MPLS/IP communication network is the ideal choice for Trans-Borneo Smart Grid as it transfers phasor data in less than 50ms, which is way below the latency requirements of the PMUs. On the contrary a network operating purely on routing protocols without the presence of the fast switching mechanism offered by MPLS is not able to meet the latency requirements of PMUs. Furthermore, it can be noted that our proposed architecture is also...
able to support other smart grid applications such as renewable energy monitoring systems without affecting the delay of phasor data.

CONCLUSIONS AND FUTURE WORK

In this paper, a Wide Area Monitoring System consisting of PMUs operating on MPLS/IP communication network for the Trans – Borneo Smart Grid has been simulated and sample results are given. We have also simulated wind farm and PV plant monitoring systems with the proposed WAN. The simulation results indicate that by using MPLS/IP Wide Area Network delays for phasor data as well as renewable energy monitoring systems can be contained within 50ms. Hence it can be concluded that such a network will be able to satisfy all smart grid applications and can significantly aid in realizing the Trans – Borneo Smart Grid.

Future work will study on evaluating the appropriate communication technologies for the incorporation of Advanced Metering Infrastructure (AMI) in the Trans – Borneo Smart Grid.

REFERENCES


42


**Figures**

![Fig. 1 Single line diagram of the integrated power grid (Maajie, et al., 2013)](image1)

![Fig. 2 Comparison of SCADA (top) and Phasor Data (bottom) for February 7, 2010 event (NERC, 2010)](image2)

![Fig. 3 Causes of supply interruptions in Sarawak 2010 estimate](image3)

![Fig. 4: Phasor Technology Overview](image4)
Fig. 5: Wide Area Monitoring System Architecture (Jitlikhit, 2010)

Fig. 6 MPLS Label (Jannu, et al., 2012)

Fig. 7 MPLS Network (Jannu, et al., 2012)

Fig. 8 Location of the PMU and PDC for the Proposed WAMS for Trans-Borneo Grid

Fig. 9 OPNET Model of Proposed MPLS/IP Wide Area Network for the Trans-Borneo Smart Grid

Fig. 10 Node Model of the Mambong PMU

Fig. 11 OPNET Model of Trans-Borneo Smart Grid with OSPF Routing
Fig. 12 Bintulu Wind Farm Subnet

Fig. 13 Limbang PV plant subnet

Fig. 14 Traffic Received (Packets/sec) by WAM server

Fig. 15 LSP Delay (sec) between the selected substation LER and SDC

Fig. 16 TCP Delay (sec) of WAM server at SDC

Fig. 17 Comparison between TCP Delays (sec) of Smart Grid with solely OSPF Routing (top) and MPLS/IP (bottom)
TABLES

<table>
<thead>
<tr>
<th>System Frequency</th>
<th>50 Hz</th>
<th>60 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting Rates (samples/sec)</td>
<td>10</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 1: Reporting Rates of PMU (Hauer, et al., 2004)

<table>
<thead>
<tr>
<th>Application</th>
<th>Phasor Data Type</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>State estimation</td>
<td>P, Q, V, O, I</td>
<td>1 s</td>
</tr>
<tr>
<td>Transient stability</td>
<td>Generator internal angle, d/dt, f</td>
<td>100 ms</td>
</tr>
<tr>
<td>Voltage stability</td>
<td>V phasor</td>
<td>1 – 5 s</td>
</tr>
<tr>
<td>Post-Mortem Analysis</td>
<td>All measurements</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 2: Smart Grid Applications Data Type and Latency Requirements (Kansal, et al., 2012)

<table>
<thead>
<tr>
<th>Source (IP address) – Destination (IP address) Pair</th>
<th>ETE Delays (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tada PMU (192.0.15.2/24) – WAM (192.0.27.2/24)</td>
<td>22.933</td>
</tr>
<tr>
<td>Mambong PMU (192.0.17.2/24) – WAM (192.0.27.2/24)</td>
<td>17.854</td>
</tr>
<tr>
<td>Similajau_PMU (192.0.44.1/24) – WAM (192.0.27.2/24)</td>
<td>22.099</td>
</tr>
<tr>
<td>Bunut_PMU (192.0.49.1/24) – WAM (192.0.27.2/24)</td>
<td>24.686</td>
</tr>
<tr>
<td>Tundan_PMU (192.0.50.1/24) – WAM (192.0.27.2/24)</td>
<td>24.686</td>
</tr>
<tr>
<td>Limbang PMU (192.0.51.1/24) – WAM (192.0.27.2/24)</td>
<td>27.563</td>
</tr>
<tr>
<td>Lawas PMU (192.0.52.2/24) – WAM (192.0.27.2/24)</td>
<td>27.974</td>
</tr>
</tbody>
</table>

Table 3: ETE delay (ms) of the seven PMUs

<table>
<thead>
<tr>
<th>Source (IP address) – Destination (IP address) Pair</th>
<th>ETE Delays (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT_1 (192.0.54.2/24) – WAM</td>
<td>23.500</td>
</tr>
<tr>
<td>WT_2 (192.0.54.1/24) – WAM</td>
<td>23.200</td>
</tr>
<tr>
<td>WT_3 (192.0.54.6/24) – WAM</td>
<td>24.108</td>
</tr>
<tr>
<td>WT_4 (192.0.54.8/24) – WAM</td>
<td>24.156</td>
</tr>
<tr>
<td>WT_5 (192.0.54.7/24) – WAM</td>
<td>24.131</td>
</tr>
<tr>
<td>WT_6 (192.0.54.5/24) – WAM</td>
<td>24.215</td>
</tr>
<tr>
<td>WT_7 (192.0.54.4/24) – WAM</td>
<td>23.300</td>
</tr>
<tr>
<td>WT_8 (192.0.54.3/24) – WAM</td>
<td>23.250</td>
</tr>
</tbody>
</table>

Table 4: ETE Delay (ms) of the eight wind turbines

<table>
<thead>
<tr>
<th>Source – Destination</th>
<th>ETE Delays (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMU – WAM</td>
<td>27.609</td>
</tr>
<tr>
<td>Humidity – WAM</td>
<td>27.001</td>
</tr>
<tr>
<td>Temperature – WAM</td>
<td>27.41</td>
</tr>
<tr>
<td>Wind Speed – WAM</td>
<td>27.432</td>
</tr>
<tr>
<td>Global Irradiance – WAM</td>
<td>27.000</td>
</tr>
</tbody>
</table>

Table 6: ETE Delay (ms) of the PMU and Sensors simulated in the PV Plant Subnet
CLASSIFICATION OF BLOG WRITERS

Yasuaki Nishimoto, Tad Gonsalves
Department of Information and Communication Sciences,
Faculty of Science & Technology,
Sophia University, Tokyo, Japan

ABSTRACT
Classification is a supervised learning method, which groups data into classes with labeled samples. We present a Particle Swarm Optimization (PSO) algorithm and a Neural Networks algorithm that adopt supervised learning. In this paper, we do not use the traditional Back Propagation algorithm, but PSO for the evolution of the Neural Networks learning weights. The advantage of this method is the rapid convergence of PSO. The PSO-NN algorithm is tested on a publicly available online data set about blog writers.

INTRODUCTION
Nowadays, most firms hold a lot of data in their databases about the products and the users buying the products. The problem with the data preserved into the databases is that they are hardly utilized to improve the business prospects. To improve the business prospects, the firms have to make or sell the products with greater efficiency which means they have to know which products are preferred by the users and which are not. But it is almost impossible to ask all users to understand their needs. To understand the users’ needs, most of firms have to analyze the database which holds the users’ past purchase information.

In order to analyze these data sets and get valid, novel and potentially useful information, the data mining community has developed a various analysis and data mining tools. The data mining techniques have been used to operate on a large amount of data to discover useful information, patterns and relationships to help make decision. There are many kinds of data mining techniques depending on the data set or the purpose of the users. In this paper, we will implement classification by supervised learning.

Classification is one of the most useful data mining techniques (Bhardwaj and Pal, 2011). It is a predictive data mining technique, making prediction about unknown data using known results found from different data (Bhardwaj and Pal, 2011). Classification uses a variety of algorithms, for example decision tree, statistical machine learning and neural networks. In this paper, we will use neural networks to implement classification.

NEURAL NETWORKS
Biological neural networks are composed of many interconnected neurons. Each neuron possesses axons and dendrites. These elements enable the neuron to communicate with the neighboring neurons by transmitting and receiving electrical signals. The dendrites provide input signals to the cell and the axon sends output signals to another neuron through the axon terminals. These axon terminals merge with the dendrites of another neuron. Signals can be transmitted unchanged or they can be altered by synapse which is able to increase or decrease the strength of the connection from neuron to neuron. Synapse cause excitation or inhibition of the subsequent neuron.

In artificial neural networks, the artificial neurons receive the sum of information which is from other neurons (input) and then pass on the transformed information to other neurons (outputs). If the summation of the product of the input and the weights is larger than threshold value, the neuron outputs the signal, which is represented as 1 in this paper and if not, the neuron outputs nothing, which is represent as 0 in this paper.

Neural networks can have one or more layers of neurons (Online Chapter). These neurons can be highly interconnected. Connections between neurons have an associated weight. In essence, the learning possessed by the networks is encapsulated in these interconnection weights. Each neuron calculates a weighted sum on the incoming neuron value, transform this input, and passes on its neural value as the input to subsequent neurons.

Each neural network is composed of a collection of neurons which is grouped in layers. Figure 1 shows a layered neural network which has three layers: input, hidden and output layers. Hidden layer is a layer of neurons that takes input from the previous layer and converts those inputs into outputs for further processing. Several hidden layers can be placed between the input and output layers, although it is quite common to use only one hidden layer. In this case, the hidden layer simply converts inputs into a nonlinear combination and passes the transformed inputs to the output layer.
In the input process, each input corresponds to a single attribute. Several types of data can be input data such as text, pictures, and voice. Preprocessing may be needed to convert the data to meaningful inputs from symbolic data.

In the output process, the outputs of a network contain the solution to a problem. In the neural network, most of the time, the numeric values to the outputs will be 0 and 1 which means “no” and “yes”. The purpose of the network is to compute the values of the output.

Weights are the key elements of the neural network. They express the relative strength of the input data or connections of each layer. Weights express the relative importance of each input to a processing element and the output. They are crucial in that they store learned patterns of data. The patterns are learned through repeated adjustments of weights.

The goal of the neural network is to mapping an input to the network to a specific category, as identified by an output of the network. The neural network learning can occur in supervised learning or unsupervised learning. In this paper, we implement the neural networks by supervised learning.

**PSO AND NN EVOLUTION**

**PSO**

Particle Swarm Optimization (PSO) is an evolutionary computation method (Xiao, et al, 2003). This algorithm is an optimization technique based on the intelligent movement of the swarm. In order to find an optimal solution to the problem, PSO uses a number of particles that constitute a swarm moving around in the search space looking for the best solution. Each particle is treated as a point in an N-dimensional point which adjusts its flying according to its own flying experience as well as flying experience of other particles.

Each particle has two associated properties, a current position and a velocity. Each particle also has a memory of the best location from their flying experience in the search space that it has found so far which is called personal-best (p-best). Every particle knows the best location found to date by all the particles in the population which is called global-best (g-best).

The velocity size or direction is influenced by the velocity in the previous iteration of the algorithm and the location of a particle relative to p-best and g-best. Therefore, each step of the algorithm, the size and direction of each particle’s move is a function of its own flying experience.

The update formula for the particle i’s velocity vector \( \mathbf{v}_i \) and position vector \( \mathbf{x}_i \) are follows:

\[
\mathbf{v}_i(t + 1) = \mathbf{w} \cdot \mathbf{v}_i(t) + c_1 \cdot (p_{\text{best}} - \mathbf{x}_i(t)) + c_2 \cdot R_2 \cdot (g_{\text{best}} - \mathbf{x}_i(t)) \tag{1}
\]

\[
\mathbf{x}_i(t + 1) = \mathbf{x}_i(t) + \mathbf{v}_i(t + 1) \tag{2}
\]

where, \( c_1 \) and \( c_2 \) are the weights associated with the p-best and g-best terms in the velocity update equation which is represents the relative impact of the p-best and g-best locations on the velocity of a particle. \( R_1 \) and \( R_2 \) are randomly drawn from uniform random number distribution. The parameter \( \mathbf{w} \) in equation (3) represents a momentum coefficient. This parameter is reduced as the algorithm iterates. \( \text{Iter}_{\text{max}} \) and \( \text{Iter} \) are the total number of iterations the algorithm will run for and the current iteration value respectively. \( \mathbf{w}_{\text{max}} \) and \( \mathbf{w}_{\text{min}} \) set the upper and lower boundaries on the value of the momentum coefficient.

\[
\mathbf{w} = \mathbf{w}_{\text{max}} - \left( \frac{\mathbf{w}_{\text{max}} - \mathbf{w}_{\text{min}}}{\text{Iter}_{\text{max}}} \right) \cdot \text{Iter} \tag{3}
\]

**Neural Networks evolution using PSO**

In this paper, we will adapt the weights of neural network into the particles of the PSO. We will call this learning method as PSO-NN learning. When we modify the values of weight to get the expected output, we will use PSO update equations (equations 1 & 2). The reason we use PSO-NN learning instead of back propagation learning algorithms is because back propagation algorithm takes a long time to determine the optimal weights value when there are many data points to classify. The advantage of PSO is rapid convergence to optimal value and easy implementation.
RESULTS AND DISCUSSIONS

In order to assess the utility of the PSO-NN learning, we will use the data set from (Gharehchopogh and Khaza, 2012) which is the data of the blog writers of Iran. The attributes of the data include: education, political caprice, topics, local media turnover (LMT), local, political and social space (LPSS). By using these attributes, we will classify bloggers into two groups: professional bloggers and seasonal bloggers. Professional bloggers are those who adopt blog as an effective digital media and interested in digital writing in continuous time intervals (Gharehchopogh and Khaza, 2012).

We have to transform the text into numbers so that we can calculate the summation of the input and weights. The transform of each attribute is shown below:

In education:

- Low → 0
- Medium → 1
- High → 2

In political caprice:

- Left → 0
- Middle → 1
- Right → 2

In topics:

- Impression → 0
- Political → 1
- Tourism → 2
- News → 3
- Scientific → 4

In LMT, LPSS, PB:

- No → 0
- Yes → 1

After transforming the attributes, we will normalize each attribute value. Then, we will divide the data set into two groups: training data set and testing data set. In this paper, 75% of data will be training data set and the remaining 25% will be testing data set. Each of them is picked up randomly in each learning epoch. The parameters of PSO and neural network are shown in Table 1 and the results are shown in Table 2.

**Table 1: Parameters of PSO-NN Learning**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of neurons</td>
<td>1000</td>
</tr>
<tr>
<td>c₁</td>
<td>3</td>
</tr>
<tr>
<td>c₂</td>
<td>3</td>
</tr>
<tr>
<td>Iterations T</td>
<td>1000</td>
</tr>
<tr>
<td>Max velocity bound V_max</td>
<td>1</td>
</tr>
<tr>
<td>Min velocity bound V_min</td>
<td>-1</td>
</tr>
<tr>
<td>Number of neurons in input layer</td>
<td>5</td>
</tr>
<tr>
<td>Number of neurons in hidden layer</td>
<td>10</td>
</tr>
<tr>
<td>Number of neurons in output layer</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 2: The Accuracy of PSO-NN Learning**

<table>
<thead>
<tr>
<th>No.</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96</td>
</tr>
<tr>
<td>2</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
</tr>
<tr>
<td>5</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
</tr>
<tr>
<td>9</td>
<td>84</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>Average</td>
<td>86.8</td>
</tr>
</tbody>
</table>

We conducted the PSO-NN learning 10 times. Each learning got different training data set and testing data set randomly from bloggers’ data set. The results show the accuracy of classification of testing data by using the model learned with optimal weights by training data.

CONCLUSION

This paper presented a PSO-NN learning algorithm which updates the weights of neural network by PSO update equations instead of the traditional back propagation algorithm. The applicability of PSO for NN training is illustrated by using the bloggers’ dataset.
Classification accuracies reveal that PSO-NN learning produces more than 80% in all of the test results.

The optimal number of neurons in the hidden layer should be determined so that we can get better results in less iterations. Further, we plan to use the PSO-NN learning method on different data sets to demonstrate its competitiveness with the other learning methods.

REFERENCES


Neural Networks for Data Mining (Online Chapter) http://www70.hompage.villanova.edu/mattew.liberatore/Mg2206/turban_online_ch06.pdf (accessed on 16/5/2013).

A COMPUTATIONAL APPROACH IN EPIDEMIOLOGICAL GAME THEORY

Teoh Shian Li, Phang Piau, Jane Labadin
Department of Computational Science and Mathematics
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Sarawak, Malaysia.

Abstract

Health phenomena and disease causality have been discussed widely and there are many cases of mortality caused by epidemic. Individual behavior plays an important role in the epidemiology (study of epidemic) and it can be presented through game theory. However, most of the conventional models for disease study do not incorporate human behavior and game theory together. In this paper, we would like to incorporate human behavior in classical epidemic compartmental model by parameterizing fear feeling and its response to vaccination program. The basic SIR model is extended by introducing “Fear” and “Vaccinated” compartments. Game theory is used to assist the individuals in the Fear compartment in making decision to whether or not to vaccinate. With this model, the disease dynamics will be affected by human behavior and the results are compared to a conventional epidemiological model. The spread of disease is slower in our proposed model with the introduction of fear and vaccination behavior compared to the basic SIR model. By creating awareness, the disease outbreak is postponed. Hence, it is essential to capture human behavior in the study of epidemic to control the prevalence.

INTRODUCTION

It is well-recognized that the spread of infectious diseases is the consequence of several inter-related components such as viral, socio-demographic, geographical and environment factors (Altizer, et al., 2013). Public health organizations had taken much effort to control the disease from spreading, which includes vaccination program, closure of school, and intensified focus on hygienic care. However, human behavior is a central key which will significantly determines the success or failure of the public health efforts made to control the spread of infectious diseases.

In the context of modelling disease spreading, Kermack-McKendrick model which was already developed back in 1927 marked the beginning of the classical mathematical study of epidemic (Li and Zou, 2009). There are three governing equations in the basic model of epidemiology, indicating three main groups of individuals with different disease status, namely Susceptible (S), Infected (I), and Recovery (R) occupying in their respective compartment. Although the parameters taken into account in these three compartments are mainly rate of disease transmission and rate of recovery (Gerstman, 2013), it was not until the last decade that human behavioral changes in response to the epidemic outbreak began to be quantified and included as it has great impact on disease dynamics. For instance, the dramatic reduction in travel and social contact were observed in Hong Kong and Singapore during the 2003 Severe Acute Respiratory Syndrome (SARS) epidemic (Ferguson, 2007). The discipline of focusing on individual behaviors as a key determinant to the dynamics of infectious diseases can be specifically known as behavioral epidemiology (Bauch, et al., 2013) and the models constructed are called behavior-disease model.

In the context of health beliefs, according to Nisbet and Gick (2008), in order for behavior to change, people must feel personally vulnerable to a health threat, view the possible consequences as severe, and see that taking action is likely to either prevent or reduce the risk at an acceptable cost. Attitudes, belief system, opinion or awareness of a disease may all affect one’s behavior in the course of epidemic and these factors change over time, both in individual and population level (Funk, et al., 2010; Barakova and Ruyter, 2011). World Health Organization (WHO) reported that there are estimated 2 to 3 million deaths every year due to infectious diseases such as whooping cough, measles, and tetanus (WHO, http://www.who.int/mediacentre/factsheets/fs378/en/). In order to reduce the mortality caused by infectious diseases, global immunization through either mandatory or voluntary vaccination has been introduced to most of the countries. Public are highly encouraged to engage in the voluntary vaccination programs as this self-initiated, voluntary behavior is believed to be able to increase the population vaccine coverage and community immunization level and therefore reduce the infection’s risk (Perra, et al., 2011). When the vaccine uptake level reaches a high enough level, herd immunity can be achieved which means that the vaccine coverage in the population is sufficient to eradicate the disease (Bauch, 2005).

On the other hand, a successful immunization with a high proportion of people opt for vaccination will strongly reduce the prevalence of disease and therefore also reduce the perceived risk of a disease, thereby killing the vaccine demands (Bauch, et al., 2013). This situation is even complicated by the existence of non-vaccinators who benefit from the herd immunity and being protected from contracting the disease. This created free-riding effect to the remaining unvaccinated individuals (Fu, et al., 2010). Therefore, individuals are losing their incentive to vaccinate and thus leading to the low vaccine coverage. Hence, the voluntary vaccination program is actually a social dilemma (Wu, et al., 2011). Individuals have to make decision whether or not to vaccinate and their decision is important in determining the vaccine coverage and hence the disease prevalence in the population. This individual-level of decision making is well-suited the language of game theory (see, e.g. Bauch (2005); Fu, et al., (2010)). Specifically, imitation dynamics is often used in the study of vaccination behavior in epidemiology. The individuals make their vaccinating decision based on information that they received...
such as vaccination cost, perceived or actual risk of the diseases and neighbor’s decision.

Despite significant medical advances in the 21st century and today we know more about the infectious disease, emotional responses such as fear of being infected are still induced in different intensities whenever an individual is exposed to extensive media coverage about the infectious disease related deaths or even overhearing an informal conversation about someone experiencing adverse health effects due to contraction an infection. Epstein, et al. (2008) attempted to incorporate human behavior by introducing two processes of interacting contamination in which individuals can be infected by disease pathogen or fear or both. The contact patterns that adapted by individuals have affected the dynamics of fear transmission and thus affected the disease transmission. In the other words, an individual interacted with a sick person in the population, he/she may be infected by pathogen or fear or both. If he/she is infected by fear, whether or not infected by pathogen, he/she will adapt self-isolation. Self-isolation is an example of protective behavior that will reduce the risk to be infected by the pathogen (Epstein, et al., 2008). However, the disease status of the self-isolated individuals does not change unless they take up any vaccination.

In this paper, we proposed the integration of fear and vaccination behavior as we believed these two processed could occur almost at the same time. Our objective is to incorporate human behavior in classical epidemic compartmental model by parameterizing fear feeling and its response to vaccination program. Based on literatures, we found that the behavior of voluntary vaccination and fear as a social contagion are considered separately. In fact, during the course of disease outbreak, individuals will acquire certain intensity of fear and hence have the awareness to take some preventive actions such as to vaccinate. Therefore, it is important to capture individuals’ feeling of fear in behavior-epidemic model. The feeling is more or less affecting the decision on vaccination.

The rest of this paper is organized as follows: in Section of Model Formulation, the mathematical model is formulated. In the following section, we presented the results and analysis. Discussion and conclusion is given in the last section in this paper.

**MODEL FORMULATION**

Susceptible-Infected-Recovered (SIR) compartmental model (see Figure 1) concept has been deployed in our behavior-epidemic model (see Figure 2). Both models assumed all individuals are well mixed in a homogenous population. In SIR model, the population is divided into three major groups which are Susceptible (S), Infected (I), and Recovered (R). Initially, all individuals in the population are considered as susceptible with some randomly chosen infected individuals. As time progress, individuals in Susceptible compartment may be infected by the infected individuals with certain probability and they will be transferred to the compartment of Infected. The interactions among these individuals are modeled by non-linear term, $\beta SI$. At the same time, some of the patients will be removed from the Infected compartment to the Recovered compartment at the given rate of recovery.

In this model, we extended SIR model by introducing two new compartments, namely “Fear” and “Vaccinated” (see Figure 2). At the beginning, all individuals are considered as susceptible with some infected, feared, and vaccinated. During an outbreak of disease, some of them will acquire the feeling of fear and they will be moved to the group of Fear. In Fear compartment, individuals will make a decision whether or not to vaccinate. The individual level of vaccinating decision making is modeled in the language of game theory.

Game theory is a study of decision making where the players taking part in the game have to make selection that potentially affects the others’ interest (Turocy and Stengel, 2001). The problem of free riding caused vaccination game to be described as a social dilemma game. Bauch (2005) proposed that vaccination game should be modeled through imitation dynamics, which is similar to replicator dynamics. The difference between replicator dynamics and imitation dynamics is the strategies are inherited from the ancestor to the next generation in replicator dynamics whereas in imitation dynamics, the players are learning the strategy from the others or based on their past experiences.

The players (individuals in population) in the vaccination game are sampled randomly. In this game, the payoffs are formed by taking into account of cost of vaccination, perceived risk of vaccination, and perceived risk of infectious disease. The imitation dynamics equation is adopted from Bauch (2005):

$$\frac{dx}{dt} = \kappa x (1-x)(-1 + \omega l) ,$$

where $\omega$ is the product of sensitivity of vaccinating behavior to prevalence changes and ratio of perceived probability of significant morbidity from infection to perceived probability of significant morbidity from vaccine, $\kappa$ denotes the product of
imitation rate and perceived probability of significant morbidity from vaccine. \( x \) and \((1-x)\) indicates the frequency of vaccinators and non-vaccinators respectively. The prevalence \( I(t) \) has been considered in the formulation of imitation dynamics model as it is one of the factors affecting individuals’ decision. \( I(t) \) can be defined from the compartmental model as in Figure 2.

In our proposed model (see Figure 2), we had made the following assumptions:

i. The population is homogenous.
ii. The interaction among individuals is random.
iii. Total population is constant, i.e. growth rate is omitted.
iv. Recovered individuals are considered as protected from the disease.
v. Each individual in Fear compartment has to make decision whether or not to vaccinate.

There are five compartments in our proposed model, which are Susceptible (\( S \)), Infected (\( I \)), Fear (\( F \)), Vaccinated (\( V \)), and Recovered (\( R \)). During the disease outbreak, information is spread through some mass media such as newspaper, televisions or internet. The information has created a sense of fear feeling in population. The individuals in the Susceptible (\( S \)) compartment will either interact with the Infected (\( I \)) or Fear (\( F \)) individuals. The interaction can be interpreted as \( SF \) or \( SF \) respectively. The interaction between Susceptible and Infected will cause the Susceptible individual to move to the Infected compartment with the probability of \( \beta \). The interaction between Susceptible and Fear will cause the Susceptible individual to move to the Fear compartment with the probability of \( \alpha \). We model the Fear compartment whereby the individual in this compartment will make a decision whether or not to vaccinate. With the probability of \( x \), he would make a decision to vaccinate. This is modeled as an imitation process (Bauch, 2005) where he will make the decision based on his neighbors’ vaccination status, cost of vaccination, and the perceived risk of infection. Hence, this is denoted as \( xFV \) which represent removal of individual from the Fear compartment to the Vaccinated compartment. On the other hand, an individual may decide not to vaccinate. Hence, the probability for this is \((1-x)\) based on the imitation process (Bauch, 2005). The probability of unvaccinated individual to get infected is denoted by \( \beta_f = \tau_{\beta} \beta \) which is a reduced probability to get infected (controlled by a parameter of \( \tau_{\beta} \) where \( 0 < \tau_{\beta} < 1 \) and \( \tau_{\beta} \beta < \beta \) (Perra, et al., 2011)) due to the fact that the feared individual has the awareness towards the disease. Hence, the interaction between unvaccinated individual and Infected will cause the unvaccinated individual to move to the Infected compartment from the Fear compartment and this interaction is represented by \((1-x)\beta_f FI\). The vaccine “protection” is not for lifelong. After a period of time, the vaccinated individuals will become susceptible again with the probability of \( h \). The infected individuals will experience the recovery after \( D \) unit of time which is represented by \( 1/\lambda \). To simplify this condition, a system of governing equations is formed:

\[
\frac{dS}{dt} = -\alpha SF - \beta SI + hV, \tag{2}
\]

\[
\frac{dF}{dt} = -xFV - (1-x)\beta_f FI + \alpha SF, \tag{3}
\]

\[
\frac{dI}{dt} = \beta SI + (1-x)\beta_f FI - \lambda I, \tag{4}
\]

\[
\frac{dV}{dt} = xFV - hV, \tag{5}
\]

\[
\frac{dR}{dt} = \lambda I, \tag{6}
\]

\[
\frac{dx}{dt} = kx(1-x)(-1+\alpha I). \tag{7}
\]

We denoted the total population size as:

\[
N(t) = S(t) + F(t) + I(t) + V(t) + R(t). \tag{8}
\]

Henceforth, with these Ordinary Differential Equations (ODEs), we found the basic reproductive number \( R_0 \) for the proposed model. \( R_0 \) is an expected number of secondary cases that caused by an infected individual in the susceptible population (Heffernan, et al., 2005). The basic reproductive number for our proposed model is a function of rates of disease transmission and recovery and it relies on the number of feared individuals who decided not to vaccinate as follows:

\[
R_0 = \frac{\beta + (1-x)\beta_f F}{\lambda}. \tag{9}
\]

Based on the governing equations, a disease-free, pure vaccinator equilibrium is found to be:

\[
\xi_1 = (S^*, F^*, I^*, V^*) = (0, \frac{h}{x}, 0, 0). \tag{10}
\]

The disease-free, pure vaccinator equilibrium is a steady state where the disease is not exists in the population. The feared individuals will make a decision to vaccinate and this will contributes to the global immunization, thus the disease will be eventually eradicated. There is also an endemic, mixed state of vaccinators and non-vaccinators equilibrium determined in the model:
The endemic equilibrium is a steady state where there is prevalence in the population. It is a mixed state of vaccinators and non-vaccinators as the feared individuals will have the choices to whether or not to vaccinate. If the individuals decided not to vaccinate, the immunity in the population is not sufficient to eradicate the disease, therefore the disease is still spreading in the population.

**RESULTS AND ANALYSIS**

We have tested the efficiency of our proposed model by comparing with the basic SIR model. The basic SIR model retrieved from Weiss (2013) is defined as:

\[ \frac{dS}{dt} = -\beta SI, \]

\[ \frac{dI}{dt} = \beta SI - \nu I, \]

\[ \frac{dR}{dt} = \nu I, \]

where \( \beta \) is the disease transmission rate and \( \nu \) is denoted as recovery rate. A case of flu epidemic in a town of total population of 50,000 people is simulated by the basic SIR model (Weiss, 2013) where the parameters \( \beta = 0.7/50000 \) and \( \nu = 1/5 \). The initial population for each compartment is \( S(0) = 49995, I(0) = 5, \) and \( R(0) = 0 \). The result of reproducing this basic SIR model is shown in Figure 3.

To test the efficiency of the proposed model, we conducted an experiment by simulating the same case as in Weiss (2013) by the same value of parameters and the additional parameters which are \( \alpha = 0.001 \) (fear transmission rate that established in Epstein, et al. (2008)), \( \beta_{f} = (0.7)(\frac{0.7}{50000}) \) (\( \beta_{f} \) is adopted from Perra, et al. (2011)), \( h = 0.1, \kappa = 0.001 \), and \( \lambda_{f} = 5000 \) (both \( \kappa \) and \( \lambda_{f} \) are adopted from Bauch (2005)). The initial population for each compartment is \( S(0) = 49985, F(0) = 5, I(0) = 5, V(0) = 5, R(0) = 0 \). Figure 4 shows the simulation result produced by our proposed model. The infectious disease is transmitted in a slower mode where Fear and Vaccinated compartments are introduced into the model. With the behavior-epidemic model, although the peak of infected individuals is the same as basic the SIR model, but the spreading of the pathogen is slower.

Since the population of susceptible, fear, vaccinated, and recovered will never go beyond the total population, \( 0 \leq S(t) \leq S(0) \leq N, \) \( 0 \leq F(t) \leq N, \) \( 0 \leq V(t) \leq N, \) and \( 0 \leq R(0) \leq R(t) \leq N, \) therefore the limits of \( S(\infty) = \lim_{t \to \infty} S(t), \) \( F(\infty) = \lim_{t \to \infty} F(t), \) \( V(\infty) = \lim_{t \to \infty} V(t), \) and \( R(\infty) = \lim_{t \to \infty} R(t) \) exist. At the same time, these limits implied that the limit of \( I(\infty) = \lim_{t \to \infty} I(t) = N - S(\infty) - F(\infty) - V(\infty) - R(\infty) \) exists. In this case, with all of the limits presented, it is proven that the infectious disease will always dies out for all population, i.e. \( I(\infty) = 0 \).

Using the basic reproductive number obtained in equation (9), the stability of both disease-free and endemic equilibrium has been analyzed. When \( R_{0} > 1, \) the disease-free, pure vaccinator equilibrium, \( \xi_{1} \), is stable if the rate of transmission is equal to the rate of recovery. It is destabilized when the rate of disease transmission is large. The stability of this equilibrium is also depends on the size of fear population. If the size of feared population is large enough, \( \xi_{1} \) is stable.
When $R_0 < 1$, the disease is eventually dies out, hence individuals have no incentive to vaccinate. Therefore, $\xi_1$ is unstable. The endemic, mixed state of vaccinator and non-vaccinator equilibrium, $\xi_2$, is a locally stable equilibrium. The stability is based on the population of Fear and the number of feared individuals making a decision to vaccinate.

The number of vaccinator is affected by imitation rate. With a higher imitation rate, the feared individuals are more probable to decide for vaccination. The immunity in population can be achieved efficiently and thus disease will be prevented. If the disease prevalence is high, the feared individuals making a decision to vaccinate. The immunity in population can be achieved efficiently and thus disease will be prevented. If the disease prevalence is high, the feared individuals have more incentive to take the vaccine. Hence, immunity to the disease is created.

**DISCUSSION AND CONCLUSION**

In our proposed model, we found that with the presence of the behavioral changes in epidemiology (Fear and Vaccinated in our case), the spreading of an infectious disease is slows down. It took longer time to transmit the viruses, compared to the basic SIR model. It is most likely caused by the vaccination program introduced by the authorities. During the disease prevalence, the interaction of individuals will create fear and they would like to make the decision to vaccinate. Game theory is suggested to model this as it is a tool for decision making. The perceived risk of both infections and vaccination, cost of vaccination, morbidity by pathogen are taken into consideration to calculate the payoffs for the individuals. There is also a little probability for vaccinator to become susceptible again. In this simulation, some individuals may become susceptible again after a period of time even though they had been vaccinated. This means that the vaccine is not for lifelong protection and immunization from the pathogen.

Fu, et al. (2010) mentioned that voluntary vaccination is a possible policy to eradicate disease. Based on the simulation, to achieve this goal, it is important to publicize vaccination program. WHO has put on large effort to introduce vaccination program to most of the countries. According to a report by WHO (http://www.who.int/mediacentre/factsheets/fs378/en/), the Haemophilus influenza type B (Hib) vaccine has only reached 45% of global coverage by the end of 2012. To increase the public awareness on vaccination, the information about vaccination program should be spread widely through different ways, such as news on television, newspaper, and campaigns. If the individuals are not provided with sufficient and correct information, they will only make decision based on their neighbors’ opinion. Subsidized vaccination program which can reduce the cost of vaccination might also increase the public motivation to vaccinate.

A preliminary study on the effect of human behavior in the dynamics of epidemic is presented in this paper. During the prevalence of disease, individuals are making decision based on the payoff of each strategy where game theory is applied. The decision making process is modeled as an imitation process which is a social influence in the population. The decision of each individual is made through learning from their neighbors to gain higher payoff. Besides the lure of payoffs, there are other factors that affecting the decision making process such as cognitive factor. The individual’s age is affecting the structure of the brain and how it handles the risks and therefore affecting how the individual makes decision (Samanez-Larkin and Knutson, 2013). In our study, cognitive factor is neglected.

In short, the results in this paper suggested that it is essential to capture human behavior in the study of epidemics. The spread of disease can be slow down according to the individuals’ behavioral changes. The model in this paper serves as a proof of concept in the effort to incorporate human behavior in epidemiology. Our next research stage is to consider the validation of our proposed model.

**ACKNOWLEDGEMENT**

This work was supported by Universiti Malaysia Sarawak and Ministry of Education through Fundamental Research Grant Scheme number FRGS/2/2013/ICT07/UNIMAS/02/6 to J. Labadin.

**REFERENCES**


STUDYING WITHOUT BORDERS

Djamshid Tavangarian, Ingolf Waßmann
Research Group Computer Architecture
University of Rostock
Joachim-Jungius-Str. 9, 18059 Rostock, Germany

ABSTRACT

After an explanation of different e-learning systems that help overcome the regional, national and international boundaries of learning provision, their structures are critically evaluated in order to draw conclusions for a reorganization of such systems. The discussion results are reflected in the development of a novel concept called “learning hub”, which helps the user to find the most suitable learning content by gathering, reprocessing and evaluating materials from different learning repositories. As a first prototype, the “Wiki-Learnia” platform is introduced, which combines the character of a social network with conventional learning management systems (LMS). By means of Web 2.0 and Web 3.0 technologies it provides new services for cross-border acquisition of knowledge and thus presents a useful way to offer studying without borders.

INTRODUCTION

Due to shorter cycles of innovation, the structure and organization of basic and further education becomes more and more important. Media based learning has taken a dominant part in both individual and regular education processes. Learning can be integrated into the everyday life, for example as a single online course (like MOOCs [LAW13]) or its combination with presence sessions (blended learning), mobile on a portable device like the smart phone or stationary at workspace or private computer – meanwhile there are almost no borders, neither local, temporal or technological. Over the course of globalization the borders of state education and private education mingle between regions and states. Time differences between countries are being bypassed with advanced recording and playback devices. Adapted repositories and communication technologies like social networks connect communities in different regions and supply possibilities for extensive cooperation and collaboration [BS00] [DTS09].

The best way to support the individual process of learning and creating knowledge is to support self-organized and self-controlled activities [GM95]. Therefore, the ability to judge and reflect the material in a critical way is of primary importance. This, however, is also trained during the learning process itself where you learn how to learn. For authors and teachers, this requires the increasing of their professional expertise to a methodical expertise. It will also change the role of the teacher from an expert to a moderator, tutor and advisor who is able to incorporate the complex relations between communicative and cognitive processes. At the same time, we should not ignore the aspects of context-creation and context-usage [LZL10], eventually enriched with ontologies [LT06]. Especially these instruments help supplying individually tailored material to the recipients which optimizes the learning process.

In this article, examples from both own and foreign developments will be shown that help to partially overcome the national and international limitations on knowledge transfer. They are also the foundation to critically rate the established structures, review them and obtain consequences for a restructuration. These discussions have led to the development of a new system called “Wiki-Learnia” which implements the novel “learning hub” concept. Wiki-Learnia combines the functionality of social networks and conventional learning management systems for learning and teaching. By means of Web 2.0 and Web 3.0 technologies it creates new services for the trans-border organization of knowledge transfer with an interesting potential for future development in the area of computer supported learning and teaching. This new concept allows the access to learning materials from wherever they are in the web by taking into account the participant’s context.

PRACTICAL EXAMPLES

Organizational structure

In fact, most e-learning systems are geared to a remote education for recipients who - due to various reasons - cannot attend seminars held at the university or in a certain class room. Our system, however, claims to provide e-learning seminars for both presence and remote attendees as a cost efficient and well organized approach. In this way, remote attendees will get accustomed to the procedures and actual situations during the seminar. A three way organizational structure [NT04] will be the foundation for this (fig. 1).
The first phase intends to homogenize the varying knowledge levels of the participants who come from different geographical areas with different background knowledge. The literature required for preparation will therefore be provided. In a first presence seminar substantial procedures and additional aspects will be discussed.

Remote attendees, just like presence attendees, are put in planned seminars synchronously, get the backing materials, contribute to discussions, ask their questions as soon as they occur during the seminar and much more. For this, the second phase will favor an internet live streaming of all seminars (for example with Adobe Connect), which is also the basis for all following practical examples, as well on a national and international level, even though it may not be possible in every single case. Since the live streaming is bounded to a given date, scheduling problems or local time differences will keep people away from attending live. That is why seminars will be recorded and stored in repositories for their future use. Of course it is necessary for the users to work with the material as soon as possible to still be able to keep up synchronously and to participate in the discussions. To make that easier, tutors can be contacted by e-mail, (video) chat or similar methods and will help the recipients with their questions. Beside these recordings, other print- and digital media will be provided, like journals, e-books, well edited online manuscripts and a literature list with content that can be acquired in the form of digital documents (for example from the university library) and many more.

The third phase is designed to intensify the knowledge that was acquired in phase two. A limited timeframe of about two weeks will be available in which the recipients can ask questions and put them up for discussion with the tutors. To further enhance this, features of social networks will be supplied, enabling recipients to connect to groups, work collaboratively and discuss with each other by means of chat, forum or e-mail. Assessment tools will enable them to qualify and reassess their knowledge in order to determine deficits in their learning process.

Media supported junior studies

The transition from school education to university education leads to a lot of problems for students often resulting in canceling the whole course of studies. Most pupils have no idea how their selected field of studies is structured with regards to content or organization. A steady estimation is often not gained before one or two terms. Canceling university or switching subjects after some period not only leads to a huge loss of time but also to financial losses for the students and also governmental supporters.

The media supported junior studies, however, enable school students in their last years in secondary graduation to participate remotely at university lectures synchronously with university students. They can also watch and work on their material after school or on the weekend, in case they are unable to participate live. To compensate for their lack of time and knowledge, an extensive support by tutors as well as special discussion sessions in person will enable them to keep up and to participate in tests and the final exams. Every test passed will be certified and accredited for future university studies. This synthesis of adding live streaming and recordings of seminars opens up a new dimension in the field of e-learning. As a method it is therefore also called live blended learning.

A successful participation not only allows attendees to shorten the overall time needed but also gives a valuable orientation for choosing their future field of studies. It is the practical implementation of “studying before you study”.

This program is successfully offered since 2008 at the University of Rostock. Nearly 600 school students have already participated with great response.

M. Sc. Programs

In this program, a number of master degree courses at the University of Rostock were held since winter term 2010/11 until summer term 2013 in cooperation with the Yerevan State University YSU in Armenia. The students were located in Armenia but were matriculated at the University of Rostock. The lectures were held in English. As a test course, the master degree course “Computer Science/Visual computing” was offered. It was held in Rostock as a conventional seminar. For Armenian students, the seminar as a whole was exported in form of a live blended learning course as mentioned above. Due to time zone differences, lecture courses not only were streamed but also were recorded. The reprocessed seminar scripts were provided in online repositories that are time- and place-independently accessible through a management platform for learning and teaching. Advanced search mechanisms allowed the users to search
for individual passages that are the most important for the current learning process. There were also compact courses held at the YSU by Rostock University lecturers in the term breaks that supplemented the program. Additionally, exercises and seminars were created by lecturers from Rostock which were then held by Armenian colleagues in close cooperation. There were a video conference system and online forums to assist students in the solution of problems and for discussion. Further presence courses were available in the form of block seminars and summer school programs.

The whole course of studies was determined for 4 terms of which 3 were attended in the form of live blended learning in Armenia. The fourth term was dedicated for participants to write their master thesis in Rostock.

Internships were also an integral part of the course of studies. They were conducted in laboratories and media centers to fortify the knowledge gained in lecture courses and in exercises on practical examples. Additionally, an industrial internship in a German company could be performed.

In 2010, the M. Sc. Program started with a number of 10 students. All students were able to finish their studies including their master thesis successfully after four terms. This proofs the possibility to effectively execute programs like this in spite of differences in countries, language, systems and culture.

Programs for further education

Education and learning are about to change fundamentally in our information society. These changes have a huge impact on knowledge acquisition after studying or after job preparation. Due to constant progress, further education to keep the level of knowledge up to date has become essential. Special knowledge that has been important in the past may no longer be needed anymore. The half-life of knowledge is decreasing permanently.

In the light of these changes a union of northern universities started a prototype project [UNT05], developing an online lecture course in the field of computer engineering for further education [MT11][WT12].

The offer that was planned with remote participants in mind delivers programs for a B. Sc. degree as well as a M. Sc. degree. Since the former requires six terms, the latter is designed to only take three terms. The study course is extended by one term if the students have no experience by an earlier job or other relevant activities. Since the program is intended for further education, study times can be modeled in a flexible way and may be extended up to two times. Even individual topics are selectable, enabling partial studies. For this, a certificate can be acquired as soon as the required exams have been passed.

Lectures will be supplied by the involved partners who are also responsible for support and for taking the exams. Administrative work will be done by a central institution.

Due to the autonomy of German federal states in questions of education (like tuition fees, approval of different lecture courses or embedding running lectures), the program is still in development. It is planned to start the M. Sc. program Computer Engineering in the winter term of 2014.

National and international developments

The amount of online material that is available for personal education, office education, school education and university education is similar to the general information overload on the internet. A lot of state funded programs as well as initiatives of certain educational institutions have led to a lot of different repositories that contain different formats for files and data. Generally, there are a lot of possibilities available to use learning material or access them for learning processes.

Table 1, columns 2-5 show exemplary and as a general overview a variety of e-learning systems that are currently available. Here, Massive Open Online Courses (MOOCs) are representative of portals like Udacity, edX or Coursera. There are a lot more systems and platforms that could be added to this table: proprietary material on local websites, learning (content) management systems, Open University, University of the People, iTunesU, e-lecture portals and many more. It is therefore restricted to selected examples in order to illustrate the broad spectrum that is available. One thing, however, they all have in common: All of them are separate entities of teaching units that require a lot of manual work to form a consistent education potential. Furthermore, a cross platform search function that reaches beyond a single product is difficult, even if semantic information for single learning contents do exist. So there are few opportunities for a continuous use of these systems.

<table>
<thead>
<tr>
<th>Course Management</th>
<th>MOOCs</th>
<th>Copendia</th>
<th>Wiki</th>
<th>EduTubePlus</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Management</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Learners Can Create Content Intuitively</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Earn A Degree</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tutor Support</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Quality Assurance</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Collaborative Tools</td>
<td>+/-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

This situation demands the extension of expertise to methodological competences and to change the role of the learning subject away from a recipients view to an expert’s view who is organizing the learning process by himself. The former learning subject has to be able to take into account the complex relations between content
processes and cognitive processes, to understand them and to use them to structure his learning progress.

This process shows a lot of analogies to a new culture of learning in its many variations and is often influenced by constructivist and epistemic concepts. Therefore, the amount of theoretic preliminary studies is huge. Apart from essays and lectures that are utilizing the term “New Learning Culture” as a catchphrase without any further definition, a few approaches that sharpen existing ideas do exist. One of those is for example [Me05] which gives an overview on the state of research on “New Learning Culture” or [HG08] and [Wi08] which formulate the requirements for “good education”. One needs to add, however, that despite of the new name, the idea of “learner centering” of [Me05] is not new at all. There already exist a lot of variations of it like for example as an integral part of the German progressive educational movement in the 1920s or included in partly politically motivated educational reforms in the 1970s [JM03]. Internationally, similar conceptual approaches are found in “situated cognition” [BCD89] or “discovering learning” [BC94].

A concluding overview results in the following process-oriented basic features of learning processes, which are partially supplied by many of those systems:

- active and passive learning;
- formal and informal learning;
- individual and social (collaborative) learning;
- assisted and self-motivated learning.

While films, videos or audio books create a passive access, interactivity by means of simulators or self-tests enables an active learning process. These solutions are also partly community-supported. Formal learning relies on pre-defined learning contents, whereas on the contrary informal learning enables the user to define the content and approach all by himself [DST11]. Individual learning in contrast to collaborative learning is a single handed attempt.

E-learning systems can be divided into supported but self-contained variants and unsupported but open variants (see fig. 2). If the system is structured and allows certification, the users have no influence on contents of the courses. If it has an active and open community which can add and edit contents, a certification is generally impossible. Fig. 2 summarizes selected providers and systems with an easy access to learning material and a broad user base. Those features can also be applied to a number of similar providers.

To unite the implications of the above mentioned e-learning systems in practice the next chapter will introduce a framework called “Wiki-Learnia” to sum up the different paradigms. Their implementation and inclusion will integrate currently available material. This way, mechanisms can be developed that are suited and can be applied in school education, office education and university education as well as for different languages and cultures.

Fig. 2 Comparison of open and closed systems

SUGGESTING A NEW CONCEPT

Both the practical examples shown in the previous chapters as well as the e-learning systems developed up to this date imply the growing importance of e-learning systems in a conventional organized educational system. This process fortifies the classical approaches with structures that increase the learner’s flexibility, take their individual needs into account and enable them to participate in a variety of education processes.

It is therefore essential to reflect: How should and can both e-learning and the learning process itself be optimized and restructured in a border-crossing way for different groups of various ethnic origins and from different communities?

Educational borders

Passing from one part of your life into another (school, university, work) requires managing those passages. This includes chapters like school education over education for a job, university education and further education up until pension age. Besides local and temporal borders [LT06], further problems can be identified in the organization of learning processes [LNT05]. Due to globalization tendencies, there are also linguistic and cultural differences [MDT0]. The next chapters will therefore point out concepts to overcome the most important boundaries.
Social networks

A high educational and political significance is held by social networks [JH10] due to their communicational possibilities. With an adequate conception, they can be an important fountain of knowledge for different learning groups which unite in the form of virtual communities. Structured educational processes that support the recipient enable target-orientated learning. Complementing presence lecture courses that rely on a method mix of blended learning or live blended learning can further enhance the learning process as needed.

Learning Hubs

A hub can be understood as a center point that unites different single entities. In technical terms, like for example in local networks (LANs), hubs are a point that concentrates on directing data communication between different segment nodes and terminals.

The term “hub” is being used in many different areas. Besides redirecting streams of communication in computer networks [HA10], the term is also used in relation to economic structures or as an education hub [KN11] in creating branches of traditional educational institutions in foreign countries. We are going to use the term as a “Learning-Hub” in another definition: Hubs deliver mechanisms that interpret information from a website based on a user defined search object. It is then gathered and reprocessed in a structured way as a result of a user’s specific query. A hub is therefore a valuable instrument in researching specific topics. Combined with contextual and semantic procedures for evaluation and sorting criteria, a hub supported system delivers results in a high quality suited for the intended purpose.

In social communities with connected structures, hubs are the central stimulus. They are characterized by their high recommendation activity which leads to specifying the requested information.

For the interested recipient, an e-learning hub chooses the most suitable material from all the available repositories based upon which material is most adapted to the requested topic. The hub is searching context-orientated, gathers the results, evaluates them, adapts the material from any media of the internet and reprocesses them for the individual learning process.

This introduces a new instrument in the area of e-learning which will deliver an important aspect to enhance the quality of learning over the next years. A learning hub extends the term “ubiquitous learning systems”: It not only includes local and temporal independence for learning processes but also includes and overcomes aspects of age, culture and linguistics.

The next chapter is going to introduce the possibilities of such a learning hub at the example of the “Wiki-Learnia” system and to describe it’s characteristics as a limitless social oriented learning system.

Wiki-Learnia

Wiki-Learnia is going to combine innovative web 2.0 and web 3.0 technologies, the wiki platform and multimedia techniques for teaching and learning. Its aim is to certify and qualify learning in all personal development stages with the help of communities.

Fig. 3 shows the schematic architecture of the Wiki-Learnia portal. It is based on the Wiki idea to widen your own knowledge level by the collaborative help of other users. It can be seen as a virtual marketplace where agents like universities, vocational schools, private education institutions, companies, groups or individual people can put their content. The user community can then reprocess this content by optimizing or accrediting it. Wiki-Learnia gathers resources, ports them for a certain task, structures it and forwards it to the user. If a user is registered and his portfolio is known, another level of content optimization becomes available. Based on the knowledge about the user, the best fitting content can be preselected.

In this way, updated content is always available. Multimedia material is not only being consumed but also being generated by means of editors in order to increase the quality of the learning material significantly. By creating learning modules, different parts can be combined in form of personal integrated curricula with subject combinations that are unique and demand-orientated but can still be certified. Gradual user levels allow them to take the role of a recipient, teacher or lecturer. The former comes from various social classes. But university students are also addressed as candidates for new professions or individuals looking for further job improvement. Lecturers are university professors, freelancers, experts from companies, or any other author who has acquired expertise in a certain subject. Users can therefore assume several roles at once. They are members of various communities in Wiki-Learnia at once.
Besides user created material by means of Wiki-Learnia editors, common internet information can be used as an input for the learning process of recipients and authors too. This task is accomplished by the learning hub system. It crawls web portals, sorts the information content- and context-wise and reprocesses them in form of modules which will then be published for the users.

The learning material involves a number of facets: One single unit includes a video and/or an audio recording of a lecturer, supplemented by his presentation and several mechanisms for an advanced search. Illustrations of special topics, exercises and reading material for the selected topic as well as explanations of subtopics conclude the recordings. The learning material is rounded up by linked external literature (digital libraries, online encyclopedias, lexicons, conferences, journals and web contents). Chats, forums and interest communities can play a primary part in initiating and hosting discussions as well as intensifying the material. Additionally, it is possible to get internal support by organized and institutionally secured organizations (public academies or private education centers) for a fee. This tutorial support is enacted by subject matter experts, which is the basis for future certification of what was learned. If needed, the support can be extended by presence training or practical work with specialized institutions. Using the community material for self-oriented learning is, however, free of charge.

Achievements explicitly made by the students can be registered by an institution and will be certified according to an exam proofing the success. These certifications will be charged and can therefore only be conducted by an officially approved institution. They can be filed for a single course, a combination of courses and can also be accumulated to reach vocational or academic degree certification.

The Wiki-Learnia concept is an innovative basis for basic and further education that realizes knowledge acquisition in a variety of subjects. The intended fees on individual support and the creation of certificates as well as the input from sponsors focus on the sustainability of our community.

The development and prototyping of the Wiki-Learnia concept is done by a team at the University of Rostock. Other contributors (like it is common in social networks) are very welcome. We especially focus on vocational schools, independent organizations and companies. The basic concept and server capacity is currently hosted at the University of Rostock, Research Group Computer Architecture.

**Didactic concept**

The principles of teaching in Wiki-Learnia are based on a variety of learning units from different subjects that can be combined with each other. The following chapter will explain this concept in detail.

A learning unit in Wiki-Learnia is an object based on a didactic concept with a clear pre-defined learning target (see Fig. 4). Learning units are characterized by embedding helpful elements like simulations or practical examples and schematics that can be integrated in the learning process. It includes learning material of about two academic hours and is enriched by a variety of additional content as there is:

- video recordings or podcasts with presentation slides;
- supplements like manuscripts;
• exercise material like different practice tasks;
• additional literature in form of bibliographies and links;
• practical tasks for extended theme complexes (e.g. with simulation tools);
• self-controlling mechanism to monitor the learning progress.

These components can be commented and extended by the recipient for personal use, to make recommendations and to leave private notes. Content can be reviewed by an assessment or can be marked as accredited by an organization that is allowed to issue a certificate or final degree. This marked material, which is characterized as high quality, can no longer be edited by the community but can still be commented in order to take these comments into account for updating the article in a timeframe of about two years.

Sequence of a learning process

An important cornerstone of the concept is the structuring of contents which allows the user to attend singular modules (that can individually be selected) as well as thematic combinations of modules (combined module) up to combining learning content in the form of university courses. According to personal interest, the user can either choose small units (knowledge bits) depending on the competence he currently needs. Regarding the completion of a whole university degree, the gained time flexibility is a huge advantage.

As mentioned above, unsupported learning has been separated from supported and certificated learning. Courses without support that are structured for self-learning supply the material free of charge. For charged courses with support an explicit registration at a central server has to be done.

Technical realization

For technical realization the widely known MediaWiki system is in use. Through a variety of own modifications and extensions the systems has overcome its original Wiki character and has now adapted features of social networks and LMS. These new additions make it especially useful for learning and teaching tasks. It is populated with scripts and diverse learning material that can be viewed or downloaded. Authors have the possibility to record video lectures, upload them, link them, supply them with text based material, and rework it cooperatively and directly out of the browser they are using. These additional materials can include features like Google Books, Wikipedia articles, YouTube videos or digital libraries.

Various tools allow communication and collaborative work between the participants and for tutorial help. The most important ones are a chat for synchronous communication and forums for asynchronous communication. Both can be realized peer-to-peer or group sessions with several users. The Wiki system already provides a rudimentary technical basis for community collaboration.

For support and tutorial assistance a web conference solution is integrated, including a video chat as well as a virtual whiteboard. Interactive seminars for groups can easily be conducted this way.

Creating a user specific profile enables the social integration of the participant in the community and to comply with the professional needs and interests. Users with similar interests can easily be found for a professional exchange of ideas. A personal profile also increases the connection to Wiki-Learnia. A better virtual connection between tutors and learners is also realized. Anonymous profiles are enabled but are contrary to the learning process.

In addition to textual learning material, recordings from lecture courses and discussions are provided. This is realized through a system that synchronizes audio and video material with presentation slides. A user friendly interface similar to a video recorder enables a repeated viewing, pausing, advanced search mechanisms and categorizing of the material into smaller learning units. All software just requires an actual web browser and is independent from operating system.

Another feature of the learning platform is the functionality of self-assessments to control the own learning status. This can also be used as a grading test to make the selection of appropriate learning modules easier.

Advantages

The Wiki-Learnia concept has a lot of advantages over other concepts, especially on didactics but also on technological areas. Modularizing learning contents lead to a high flexibility and enable an individual structure of the learning process. The learner only needs the necessary material that expands his current knowledge level. Material that is already known can be opted out. This flexible use in all different live stages helps to overcome those boundaries.

The highly user friendly interface and the well know basis of the MediaWiki software make it easier for people with limited technical knowledge to enter Wiki-Learnia. A completely re-written user layout further lowers the inhibition level. The huge number of media and communication tools enables a user-specialized learning arrangement.

Similar to social networks, users can fast and easily contribute to knowledge creation and knowledge transfer due to a balanced integration of web 2.0 features.

The Wiki system allows switching languages according to the user and thereby helps to overcome linguistic and cultural barriers.
CONCLUSION

Basic and further education will have an increasing importance in society. The Wiki-Learnia portal stands for a flexible knowledge acquisition and is a contribution for lifelong learning. Everyone in the community can participate actively. Wiki-Learnia is an easily accessible internet platform, on which a variety of learning content on different levels (academic, school, job-oriented) is reprocessed in a modular way. In addition to that, a certification of modules similar to traditional courses is possible, enabling even degrees.

In a unique platform concept, Wiki-Learnia unites the openness of a modern social network of web 2.0 with a strong community character, as well as a certified quality of a professional educational institution with accredited learning offers.

The introduced learning hub shows an innovative dimension in the area of media supported learning. The unstructured variety of online learning material currently available is inventoried, reprocessed and accessible by means of Web 3.0 mechanisms in a clear form for recipients.

Summarized, the Wiki-Learnia platform delivers a number of supportive features that are usually only available singular and incomprehensive at a few different agencies as you can see in fig. 3.

INVITATION

Not every single feature has been implemented at the current date. Starting in summer 2014, a first version including the basic characteristics will be provided. We hope to get a broad response from the community. Therefore, we have different partners in mind.

One group consists of those who – due to their expertise – are able to execute internal tasks to enhance the system. These include for example software engineers. The following table 2 gives some examples.

Since a lot of new and interesting questions regarding legal rights, copyrights and financial foundation will come up support from acquainted experts is very welcome. Another group could engage themselves in terms of quality assurance in university and self-improvement areas. A number of subjects will require surveyors to rate the contents and certificate them after use. These also include structures for quality assurance. For the didactic frame the support of educationalists would be helpful.

Questions regarding the way of communication inside the communities and between the people involved in the development of the portal as well as aspects of usability and user experience are other interesting topics for a closer engagement. Regarding the optimization in creating organizational structures and management structures we welcome help from economic or social sciences.

And at last we are looking towards the cooperation with institutions and organizations that can deliver appropriate content, produce it or offer programs for getting certificates for interested recipients. This includes universities, vocational schools, chambers of trade, specialized education institutions, common schools or schools in foreign countries as well as companies that are willing to transfer their education programs as a whole inside Wiki-Learnia.

We therefore welcome any support that can assist us in developing the ambitious project of Wiki-Learnia (contact: info@wiki-learnia.org).

Table 2  Learning objects in Wiki-Learnia

<table>
<thead>
<tr>
<th>Examples of Wiki-Learnia tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chat and video conference possibilities for synchronous communication and feedback acquisition</td>
</tr>
<tr>
<td>Content ranking and commentary function to enable feedback functionality for the community and thereby set signs of quality</td>
</tr>
<tr>
<td>Optional integration with other social services (like Facebook, Twitter, Wikipedia or YouTube), content creation</td>
</tr>
<tr>
<td>Simple forms of editing user documents</td>
</tr>
<tr>
<td>Interfaces to different document systems</td>
</tr>
<tr>
<td>Collaborative work at shared documents</td>
</tr>
<tr>
<td>Upload and download of learning material</td>
</tr>
<tr>
<td>Content administration to sort and assign new material to existing areas</td>
</tr>
<tr>
<td>Structuring of courses and combined modules to define learning paths</td>
</tr>
<tr>
<td>Version management to track and control content changes</td>
</tr>
<tr>
<td>Tagging of learning content in order to categorize similar or identical material</td>
</tr>
<tr>
<td>Annotations with sharing and collaborative editing</td>
</tr>
<tr>
<td>Bibliographies and embedding of online libraries</td>
</tr>
<tr>
<td>Integrated plagiarism control to increase the quality of written papers</td>
</tr>
<tr>
<td>Possibilities to automatically analyze and evaluate exams and exercises</td>
</tr>
<tr>
<td>Creating e-votes and automatically process them</td>
</tr>
<tr>
<td>Integration of audio and video contents</td>
</tr>
<tr>
<td>Web based production of videos assisted by slideshows, webcams and headsets, modular plugin structure for an easy adaption and extension of the system</td>
</tr>
<tr>
<td>Optional use of special 3D environments for adequate special disciplines</td>
</tr>
<tr>
<td>Customer friendly pay model and integration of a micropayment system</td>
</tr>
</tbody>
</table>

REFERENCES


Scenarios", 14th International Conference of Interactive Collaborative Learning, Slovakia.


INTERFERENCE EFFECTS ON SHORT RANGE TECHNOLOGY ZIGBEE PERFORMANCE

Natalie Lawrence Muda, Kuru Mathew, Mujahid Tabassum
Faculty of Engineering, Computing and Science
Swinburne University of Technology, Sarawak Campus,
Kuching, Sarawak, 93350, Malaysia

ABSTRACT
Zigbee is a wireless technology that is based on the IEEE 802.15.4 standards. It is a very reliable, cost effective, low power and much secure wireless technology. The physical and MAC layers of this protocol are based on the IEEE 802.15.4 standards, while the application, network and security layers are by the Zigbee alliance. The numbers of devices equipped with Zigbee are growing drastically in numbers. ZigBee’s sensing capabilities have been identified and found niche in smart home technology. The result is the number of co-located devices, say within 10 meters, grown to a limit, which may cause interference issues in the 2.4 GHz radio frequency spectrum. In this paper, we have performed a survey on the interference issues of 2.4 GHz frequency band and their subsequent effects on throughput of Zigbee.

INTRODUCTION
Zigbee is a wireless technology that is based on the IEEE 802.15.4 standards. The IEEE 802.15.4 standard is a type of standard that is used particularly in low rate-wireless personal area network (LR-WPAN). It is designed to be a low cost and lower speed communication device. The physical and MAC layers of the Zigbee model is based on the mentioned standard. As for the other layers in the architecture, it is defined by the Zigbee Alliance and users. Zigbee devices can communicate in a range of 10meters to 30 meters and supports up to 250kbps of data rate in the 2.4GHz frequency band. It consumes much lesser power compared to other wireless technologies as the devices can switch to standby mode when not in used. With characteristics of being low rate, low cost and low power, Zigbee have become a suitable choice for sensors and actuators.

PROBLEM STATEMENT
In today era technologies are getting more and more advance in order to help user in every field to enhance their productivity. Engineers and companies are coming out with more advancement in technologies that we have at hand. For example few year back user was using the Bluetooth version 2.0 but now due to advancement, this technology evolves and the Bluetooth 4.0 is available in the market that provide more and significantly higher data rate functions. Similarly, WiFi technology has also had an increase in its throughput from 802.11g to 802.11n. In this paper the main focus will be on Zigbee. We will be making use of the latest version of this technology and do an analysis of their working. Our main focus is how Zigbee works in different kinds of interference and how the environmental condition does affect the performance.

LITERATURE REVIEW
Introduction of 802.15.4
802.15.4 (Zigbee) is a standard that is created by the Institute of Electrical and Electronics Engineers (IEEE) organization. It is used to specify the two layers, the media access control (MAC) and the physical (PHY), of a wireless personal area network (WPAN). Networks that are running on this standard are normally low in cost, lower range, low power consumption and lower data rate. The IEEE 802.15.4 standard also known as Low Rate – Wireless Personal Area Network (LR-WPAN). Unlike other technologies such as the WLAN and Bluetooth, which runs on the IEEE 802.11 and 801.15.1 standards respectively, LR-WPAN is widely used for actuator and sensor devices (Gislason 2008).

Fig 1 illustrates the various technologies and its respective applications while Table 1 shows the summary of specification of wireless technologies.

The WLAN (IEEE 802.11) is able to support the most data rate when compared to Bluetooth and LR-WPAN. Since it can cover up to approximately 100m in distance, it has a larger network size. Thus, the power consumption is also higher. With this, the WLAN it is very suitable to be used in large area such as an office size environment. Besides having higher data rate and power consumption, it is also more costly among the other technologies mentioned. Therefore, it is not suitable to be used as embedded sensors (Gislason 2008).

As for Bluetooth (IEEE 802.15.1), it serves as a good cable replacements for end devices and can operate in medium sized areas of approximately 10m to 100m such
as a small personal space as it has data rate and power consumption that slightly lower than WLAN. Due to its nature, it is not suitable to be used for transferring media with high data rate such as videos. The cost of implementing such technology is lesser than implementation of WLAN (Gislason 2008).

The LR-WPAN (IEEE 802.15.4) is the cheapest technology. It is suitable for industrial or commercial use with shorter range of 10m to 30m. This means that it is only capable of covering a small network area. This technology serves as wireless links to devices such as actuators and sensors. Since it has very low data rate, it is not able to support real time media like voice and video.

The IEEE 802.15.4 is defined by the following (Gislason 2008):

- Have a system for network discovery
- Able to form and join networks
- Capable of changing the channels in the network
- Ability to detect interference or noise in a specific channel
- Use Carrier Sense Multiple Access - Collision Avoid (CSMA-CA) to prevent collision from occurring
- Use the method of data-broadcast that is unacknowledged and single hop

Briefly said, the main characteristic of LR-WPAN is that it is able to support data rates of 20kbps, 40kbps and 250kbps. The data rate that is supported is dependent on the frequency bands used. For the 868MHz band, it only has 1 channel and thus, is only able to support a small portion of data with the maximum rate of 20kbps. The 915MHz band is able to support up to 40kbps of data. The data rate is slightly higher than the 868MHz band as it has more channels (10 channels) in it. The frequency band that can support the most data rate will have to be the 2.4GHz band with support for data up to 250kbps. This band has a total of 16 channels which helps explain the ability of it to support higher data rates. However, there is a limitation for the use of different frequency bands. Due to the licensing of bands, the 868MHz band can only be used licensed-free in Europe while the 915MHz band is licensed-free in the United States. Globally, the 2.4GHz band can be used licensed-free. Besides, the LR-WPAN also have characteristics of being reliable in transferring data as it is able to support devices of low latency and uses fully handshake or acknowledged protocol (IEEE 2003). Not to mention, it can operate in either a star topology or a peer-to-peer topology.

In Table 2, it is seen that frequency band of 868MHz and 915MHz uses the same type of modulation technique for transmission of data. This technique is BPSK modulation, which is the simplest Direct Sequence Spread Spectrum (DSSS) approach. The same type of technique is used for the spreading of data. The 2.4GHz band on the other hand uses the 16-ary orthogonal modulation technique for data transmission. This is also a technique that is based on the DSSS approach. Unlike the 868MHz and 915MHz bands, the modulation technique used for spreading of data is not the same as the one used for transmission of data. In this case, the Offset Quadrature Phase Shift Keying (OQPSK) technique is used. This technique is also similar to Minimum Shift Keying (MSK) modulation technique (Callaway et al. 2002). Some examples of network protocols that is built on the IEEE 802.15.4 standard:

- Zigbee
- MicorChip MiWi™
- Synap SNAP
- 6LoWPAN

**Introduction to Zigbee**

Zigbee is a wireless technology that is built based on the IEEE 802.15.4 standards. Low data rates, low power and low cost are the main key features of Zigbee. Since other technologies have higher specifications, Zigbee became the ideal choice in the sensor and actuator industry. This industry does not require high data rates for data transfer and only uses a short range for transmission. Since sensors and actuators do not require high power consumptions, Zigbee can provide long lifespan, by the decades, for the battery used. In short, Zigbee is known as ‘Wireless Control’, it comes from the term ‘Wireless Sensor Networking and Control’ (Gislason 2008). The Zigbee Alliance gave Zigbee the slogan ‘Wireless Control that Simply Works’. Fig 2 shows the various applications of Zigbee in the world today.

**Detailed Characteristics of Zigbee**

Generally, Zigbee can be characterized into 6 main categories, as seen below:

**High Reliability**

Based on the IEEE 802.15.4 Standard Zigbee is a protocol that runs on the IEEE 802.15.4 Standard. The physical and media access control layer are based on this standard. With this, Zigbee is a very reliable wireless technology that runs in low range environment.

**Modulation Schemes**

In wireless communication, Direct Sequence Spread Spectrum (DSSS) is used to spread the data bits to larger bit stream in order for the data to have bigger bandwidth than the original data before DSSS modulation. When the original bandwidth is bigger than the original bandwidth, the data will have very little Power Spectral Density (PSD). Thus, there will be lesser interference to other wireless services that are operating in the same band (Panchal 2012).

Offset Quadrature Phase Shift Keying (OQPSK) modulation on the other hand is used by the transmitter to map bits of 1's and 0's to symbols between 0 to 15.
These symbols will then be mapped to 16 different 32 bit long chip sequence based on the Zigbee standards. With the combination of DSSS and OQPSK being used together, these technologies contribute to ensure the performance of Zigbee in environments with low Signal-to-Noise ratio (SNR).

**Uses CSMA-CA**

Before any transmission is done, the Zigbee transceiver will first listen to the channel to make sure that the channel is clear. The transceiver will only transmit when the channel is clear. This is to ensure that the data transmitted will free from collision from other data or signals in the channel. Thus, keeping the data uncorrupted.

**Uses 16-bit CRCs**

The Cyclic Redundancy Check (CRC) is used on each packet that is transmitted in the network. It is also called a Frame Checksum (FCS) that is used for error detection in the data bits. This ensures that the data bits that are received at the nodes are correct, free from error.

**Acknowledgement of Packets at Each Hop**

During transmission of packets from transceiver to receiver, the 3-way-handshake protocol is applied. Each time a packet is received at the receiver, the receiver will send an acknowledgment to the transceiver to inform that the packet is already received and to continue sending the next packet. However, if the receiver fails to receive the packet, it will inform the transceiver to resend the packet. The total retransmission is up to a maximum of 3 times, meaning a total of 4 transmissions. If at the fourth try the transmission still fails, Zigbee receiver will let the transceiver know that something is not right and that rightful actions must be done to overcome this problem.

**Uses Mesh Networking**

Mesh networking is very useful in the sense that it is able to have the range extended through multi-hops and form a network by ad-hoc. It can also discover route automatically and self-heal (Gislason 2008). In Zigbee networking, irrespective of distance, the nodes can communicate with one another when there are enough radios between the nodes for passing of message. Fig 3 shows an illustration of Zigbee’s mesh networking. Node 1 cannot reach node 6 directly due to lack of radios between them. Zigbee will find the best path for the message to be passed. In this case, node 1 needs to pass the message via node 4 and node 5 before finally reaching node 6. Throughout time, there may be some obstructions in the route. When this occur, Zigbee will be able to detect the error caused by this obstruction automatically and will figure out a way to overcome the problem by rerouting. Fig 4 shows the rerouting of route from node 1 to node 6 when an obstacle is found between nodes 5 to node 6. Besides mesh networking, Zigbee is also able to do broadcasting whereby a message can be distributed to various nodes within the network and send a message by using the multicasting technique so that a group of nodes can receive it (Gislason 2008).

**Uses End-to-End Acknowledgement**

The use of end-to-end acknowledgement can assure the reliability of the network in the sense that it is able to learn if a packet has been received by each node. If there is any retransmission of packets, Zigbee will be able to detect the duplicated packets and filter out those packets.

**Very Cost Efficient**

Zigbee is made to be low cost due to its unique characteristics. As what is already known, Zigbee is license-free globally in the 2.4GHz band and it is a patent-free technology by Zigbee Alliance. With this, developers do not need to spend a lot of money for licensing and to acquire the patent.

**Low Power Consumption**

There are many nodes available to make up the entire Zigbee network. However, these nodes do not need to be kept active to be in constant contact with the network in order for the network to work. What happens is that these nodes are often in sleep mode when not in used and can be immediately switched to active mode as and when it needs to be used. Therefore, Zigbee consumes very little power when compare with other technologies.

**Secure**

Zigbee is highly secure technology as it adapts the Advanced Encryption Standard (AES) by the National Institute of Standards and Technology (NIST) (Reddy 2004). The AES-128 functions as a block cipher. Therefore, packets can be encrypted and decrypted in such a way that it is very hard to crack (Gislason 2008). The encrypted packets cannot be understood by any third party if they do not know the key. Zigbee also has authentication whereby malicious nodes are not able to place fake packets into the network.

**Have Low Data Rate**

In the 2.4GHz band, Zigbee is able to communicate with a maximum bit rate of 250kbps. However, when services such as retransmission, encryption or decryption and mesh acknowledgement takes place, the bit rate will drop, giving the throughput of about 25kbps.

**An Open Global Standard**

“Zigbee is global, Zigbee is open, and Zigbee is standardized” (Gislason 2008). The main foundation for the architecture of Zigbee is based on the MAC and PHY layer of the IEEE 802.15.4 standards. The IEEE standard
is free for download and also specified that the radios in the 2.4GHz band are reliable and licensed-free globally.

**Architecture of Zigbee**

The Zigbee architecture consists of several layers – Physical, Media Access Control, Network, Security and Application. However, Zigbee Alliance is only accountable for the Network, Security and Application Layers. The Physical and Media Access Control layers are not the responsibility of Zigbee Alliance and are based on the IEEE 802.15.4 standards (Rahmani 2005). The architecture of Zigbee is properly illustrated in Fig 5 above. The physical layer is in-charge of translation of packets into bit streams and vice versa. It uses different modulation schemes for this process depending on the different ISM bands. For 868MHz and 915MHz, it uses DSSS to spread the bits into bit streams of 15 bits. It then uses Binary Phase Shift Keying (BPSK) to modulate the bit streams. As for the 2.4GHz band, an advanced version of DSSS called the 16-ary orthogonal modulation is used to spread the bits to bit streams of 32 bits. These bit streams will then be modulated using either the OQPSK or MSK modulation technique.

The media access control layer is used to control the flow of traffic in the network. CSMA-CA technique is used to ensure that the data sent is uncorrupted while the acknowledgement and retransmission technique is to ensure that the receiver have received the data. It also provides the security for the overall network by using AES-128 encryption as mentioned previously.

The network layer is in control of anything that is involved in forming the network. This includes configuration of network and discovery of devices (Le 2005). Zigbee network is able to support 3 types of network topology, namely, star, mesh and cluster tree as seen in Fig 6. The various types of devices used are shown in Table 3.

There can be only one coordinator node in each network. The coordinator’s role is to act as a router to the other networks. All the information about the network is stored in the coordinator and serves as the root of a network tree. Full Function Devices (FFD) functions as intermediary routers that are used for transmission of data from other devices. It uses less memory when compared to the coordinator node. These devices can also act as a coordinator and is able to operate in all the topologies. As for the Reduced Function Devices (RFD), they require the least memory due to their function in the network. All of these devices does talk in the network, particularly to the coordinator only and does not relay any sort of data from other devices. The use of RFD can be found in a star topology. Besides being accountable for formation of the Zigbee network, the network layer also consists of some commands that are used to ensure the security of the network. These commands are usually used for devices to join or rejoin the network securely. Since the overall payload of the network frame encrypted, it can be said that the whole network layer in a Zigbee network is secured (Gislason 2008).

All running applications in the network are the responsibility of the application layer. It acts as a filter to remove duplicated messages that are sent by the network layer (Gislason 2008). It also obtains a local binding table which is used to keep a record of whichever group or nodes in the network that the node wants to communicate with.

**EXPERIMENTS AND DISCUSSION**

Some researches on performance of Zigbee with the existence of Bluetooth and WLAN have been conducted by other researchers. They have used the MatLab/Simulink software to perform their experiments. In the following paper (Shuaib et al. 2007) they have analyzed the interference effect on Zigbee performance by implementing Simulink models as seen in Fig 7. The simulations were performed based on different modulation schemes and frequency bands as seen below.

**OQPSK for Frequency Band of 2.4GHz**

A random numbers between 1 to 16 are randomly generated by the random number generator. Once the random integer is generated, this integer will be fed into the spreader block.

In the spreader block, the integer will be spread into a 32 bits chip value based on below. This table which consists of the mapping of symbols to chips is based on the Zigbee standard. Once conversion to chip value is done, the 32 bit stream was inputted into the OQPSK modulator for modulation. Then, the modulated stream was added with Additive White Gaussian Noise (AWGN) before being handed off to the OQPSK demodulation block. Demodulated stream will then be passed to the de-spreader block where the 32 bit stream is being converted back into an integer. Before calculation of the Bit Error Rate (BER), the integer will first have to be converted into a 4 bit stream by using an integer-to-bit converter. BER is calculated by comparing the new 4 bit stream with the original one (Shuaib et al. 2007).

The result obtain for this experiment can be seen in Fig 8. The BER values decreases when the SNR increases. From the graph, data rates of 64kbps and 128kbps tend to have the same pattern as the SNR is increased. However, for data rate of 250kbps, the BER value decreases much more when SNR value is 5.

**Enhanced OQPSK for Frequency Band of 868MHz/915MHz**

This experiment is almost identical to the first experiment. The difference is the mapping function of the spreader and de-spreader. The modified codes can be seen in Table 5. Fig 9 shows the results obtained from
experiment 2 when the 2.4GHz band from the previous experiment is being replaced with 868MHz/915MHz band. Data rate of 64kbps tend to have the lowest BER value when the same SNR is applied. As for data rate of 128kbps and 250kbps, the pattern of the graph is almost similar, with data rate of 128kbps slightly lower in terms of BER value.

**BPSK for Frequency Band of 868MHz/915MHz**

For this experiment, BPSK modulator and demodulator are being used in replacement of the OQPSK ones. The code chip mapping at the spreader and despreaders are also being replaced. The 2 values of 15 bits ones. The code chip mapping at the spreader and demodulator are being used in replacement of the OQPSK BPSK for Frequency Band of 868MHz/915MHz of BER value.

**Summary of Past Researcher’s Work**

All these results gotten by past researchers were based on Simulink models which were theoretically calculated on MatLab. Different modulation schemes and frequency bands produced different BER versus SNR graph. Therefore, physical implementations of Zigbee in different conditions need to be done in order to verify the results. It is only by comparing both theoretical and physical result that we can have proper understanding of the performance of Zigbee.

**Types of Interferences**

Today, more and more manufacturers are producing new technologies that run on the 2.4GHz ISM band. This is because this band is license-free globally and it occupies a large portion of the overall ISM band. Since there are so many technologies running on this band, there ought to be interference among signals in this band. Fig 11 shows the diagram of the electromagnetic spectrum.

**Microwave**

Microwave is a kind of electromagnetic wave with frequency of 0.3GHz to 300GHz, having wavelength of 0.001m to 0.3m. Microwaves are usually propagating in a straight line and does not diffracts around obstacles (Rouse 2006). However, when it passes through obstacles, attenuation will occur – drop in signal strength. Normally, radio frequencies of longer wavelength are affected less in terms of attenuation. Microwave band is used for transmission of signals with bigger bandwidth wirelessly. The bigger the bandwidth, the higher the data speed.

**Ultra Wide Band (UWB)**

In wireless networking, UWB is used in order to obtain high bandwidth connections while using very low power consumption. Large amount of data can be transmitted by the UWB. However, the distance of transmission is very short – up to a maximum of 230 feet (Rouse 2008). On the other hand, UWB has the capability of transmitting signals through obstacles that reflects signal with more limited bandwidths and high power. This band is widely used in WPAN and devices which uses wireless connectivity (Mitchell NA).

**Near Field Communication (NFC)**

NFC is an evolution of radio-frequency identification (RFID). Similar to Zigbee, functions in short range networks and serves as a wireless link that with low power consumption. When devices are placed in distance as short as a few centimeters, transmission of data can occur between the devices. It is a very convenient way of data transfer as devices do not need to be tapped. Instead, when NFC on both devices is on, with just a single wave, those devices can connect to one another for data transmission.

**Rain**

Interference of rain to the signals in the wireless communication is known as rain fade. Rain fade occurs when the wavelengths of the signal is almost similar to the separation of the rain droplets (Rouse 2005). During rain fading, the signal will face a tremendous drop in signal strength. Since rain is a part of nature, the downpour of rain is not constant. Therefore, rain fade will not last long as it is dependent on the downpour of rain.

**Noise Models**

**Additive White Gaussian Noise (AWGN)**

In a wireless communication, AWGN is the most common type of statistical noise that happens randomly in the background. These random noise are represented by a broad range of frequency with relation to the signal in the communication channel. This model is frequently used in communication systems as it can be easily computed and can be applied to a vast class of physical channels.

**Bit Error Rate (BER)**

Bit error rate is usually used to test the performance of the system. The testing of system performance is based on how good the system is to transfer the data bits from one end to another. In simple words, BER means the capability of the system to receive information that is free from error at the receiver (Alam 2008). High BER indicates slower data rate may be able to improve the transmission time of the overall transfer of data. When BER reduces, the number of packets for retransmission will also reduce, causing a better performance of the system.

The general formula for probability of bit error:

$$P_b = Q(\sqrt{SNR})$$  

(1)
Where,

\[
SNR = \text{Signal} - to - \text{Noise ratio}
\]  

\[
Q = \text{Standard error function}
\]  

\[
Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt
\]

For OQPSK modulation, the formula for SNR is:

\[
SNR = \left(\frac{2E_b}{N_0 + 4(K - 1) E_b / 3N}\right)
\]

The Signal-to-Noise ratio in a system is very important to ensure the performance of the system. When the SNR is high, this means that the medium which is used to transmit data bits is good as the bit error rate is small.

Formula for BER:

\[
\text{Bit error rate} = \frac{\text{Number of errors}}{\text{Total number of bits sent}}
\]

The Eb/No can be used to identify the factors that affects the BER. Eb/No is the ratio of energy per bit over the noise power spectral density. The factor affecting BER can be seen in Table 7.

**Propagation Loss of Wave**

**Rayleigh Fading**

Rayleigh fading occurs in environments that do not have path that are of line of sight and has numerous scatters. All the reflected signals at the front end of the receiver will become uncorrelated in amplitude and phase which is distributed evenly between 0 to 2\(\pi\) when there is adequate amount of multiple scatters in the environment (Mathurathan 2012).

The probability density of Rayleigh fading is:

\[
f_{ray} = \frac{r}{\sigma} e^{\frac{-r^2}{2\sigma^2}}, r \geq 0
\]

Rayleigh Distribution can be seen in Fig 12. It is defined by a single parameter, \(\sigma^2\), which is the time average signal power.

This distribution has characteristics of:

\[
\text{Mean, } \mu_{mean} = 1.25\sigma
\]

and

\[
\text{Variance, } \sigma^2 = 0.4292\sigma^2
\]

With a cumulative density function of:

\[
P(R) = \Pr(r < R) = 1 - e^{-r^2/(2\sigma^2)}, R \geq 0
\]

**Rician Fading**

Unlike Rayleigh fading, Rician fading occurs when there are multipath and line of sight components in the environment. These components are called ‘specular’ and ‘random or scatter’ component respectively (Mathurathan 2012). For specular component, the amplitude distribution has mean that does not equal to zero. As for random or scatter component, the amplitude distribution will have mean of zero.

**Interference That May Affect Zigbee**

The Wireless Local Area Network (WLAN) is based on the IEEE 802.11 standard. It is based on the DSSS spreading sequence and operates in the 2.4GHz ISM band. It occupies 14 channels in this band, namely channel 1 – 14. Each of these 14 channels has a bandwidth of 22MHz and a 5MHz bandwidth separation in between each channel (Atmel 2013). Since Zigbee and Wifi are in the 2.4GHz band, interference between co-channels will occur. The overlapping of channels can be seen in Fig 13. Bluetooth is a wireless technology that adopts the FHSS technique. A total of 79 channels are supported by Bluetooth, with each channel having a bandwidth of 1MHz. The frequency hop is about 1600 times every second. Zigbee adopts the DSSS technique which is not a type of frequency hopping system. Therefore, out of 79 times, only one time a channel will overlap. Despite this, both Zigbee and Bluetooth can coexist well together in the same ISM band. When Bluetooth is transmitting in a frequency that happens to be overlapping the Zigbee channel, Zigbee will back off randomly. In the meantime, Bluetooth will quickly hop to another frequency, thus, not disturbing Zigbee.

Similar to Zigbee, wireless USB uses DSSS. It has the same number of channels as Bluetooth – 79 channels, with each channel having a bandwidth of 1MHz. Both these technologies supports frequency agility and the use of CSMA-CA and transmission duty cycle that is small helps to maintain good transmission of data with low data rate loss. These 2 mentioned technologies are able to coexist well with one another (Atmel 2013). This can be explained when the wireless USB has a frequency that overlaps Zigbee, the wireless USB will change to another new channel.

Cordless telephones functions in the 2.4GHz band. However, it does not adopt any standards like other technologies. Some of the cordless telephones use DSSS while most of them use FHSS. By using DSSS, users can change its channel manually. As for FHSS, channel cannot be changed manually as the frequencies on the telephone keeps changing. Cordless telephones based on FHSS causes a great interference to the Zigbee personal area network (PAN). It is able to back out the whole Zigbee communication system and it transmits a large amount of power that can affect the transmission of data.
To avoid interference from occurring, Zigbee channels must be placed very far apart from the cordless telephones based on FHSS. As for DSSS cordless telephones, Zigbee channels must not overlap the channels of these telephones (Atmel 2013).

Microwave is the most dispersed and unpredictable radio frequency source that functions in the 2.4GHz band. When transmission of data in the Zigbee system is more than 1m in length the interference of microwave oven can be considered negligible (Atmel 2013).

CONCLUSION

It is clear that Zigbee is based on the IEEE 802.15.4 standards, with its unique features of having low data rates, low power and low cost. Since it is license-free globally in the 2.4GHz ISM band, Zigbee is not alone during the communication with one another as there are many other wireless technologies that are operating in the same frequency band. Thus, interferences between these technologies ought to occur. An occurrence of interferences between technologies simply depends on the distance between each devices and the channel that the technologies are travelling in.

Besides interferences from wireless technologies in the same frequency band, there are other external interferences that can affect the performances of Zigbee. Some examples of these interferences are the existence of obstacles such as buildings or trees, and even the human body.

The Simulink results of the performance of Zigbee obtained from researchers of paper (Shuaib et al. 2007) were based on different modulation schemes and frequency bands. To further verify the results, physical scenarios under different conditions need to be done.

REFERENCES


Atmel 2013, 'Atmel AT02845: Coexistence between ZigBee and Other 2.4GHz Products', September 2013, pp. 1-12.


Gislason, D 2008, Zigbee Wireless Networking, Newnes, United States of America.


Mathurvanathan 2012, Eb/N0 Vs BER for BPSK over Rician Fading Channel, Gaussian Waves, viewed 2 December 2013, http://www.gaussianwaves.com/2012/07/ebn0-vs-ber-for-bpsk-over-rician-fading-channel/.


Rahmani, E 2005, ZigBee/IEEE 802.15.4, ECE Department – University of Tehran.


Raylman, JS 2004, ZigBee Security, ZigBee™ Alliance.


Shuaib, K, Boulmalf, M, Sallabi, F & Lakas, A 2006, Co-existence of ZigBee and WLAN, A Performance Study, thesis, College of Information Technology, UAE University, Al-Ain, UAE.


FIGURES AND TABLES

<table>
<thead>
<tr>
<th>WL AN 802.11</th>
<th>Bluetooth/WPAN 802.15.1</th>
<th>LR-WPAN 802.15.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centralized Wireless Networking (WLAN) in the office environment</td>
<td>Cable replacement for consumer devices in the personal operating space</td>
<td>Low-cost wireless link for industrial/commercial sensor and actuator devices</td>
</tr>
<tr>
<td>Embedded Sensors</td>
<td>Wireless LAN</td>
<td>Voice/Video Real-Time</td>
</tr>
<tr>
<td>$$$</td>
<td>$</td>
<td>$</td>
</tr>
</tbody>
</table>

Fig 1: Comparison of the Applications of Various Technologies (Gutierrez 2005)

Table 1: Specification of Wireless Technologies

<table>
<thead>
<tr>
<th>LR-WPAN</th>
<th>Bluetooth™</th>
<th>WLAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>10–30 m</td>
<td>~100–1000 m</td>
</tr>
<tr>
<td>Data Throughput</td>
<td>&lt;0.25 MBPS</td>
<td>1 MBPS</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>&lt;BT/10</td>
<td>BT</td>
</tr>
<tr>
<td>Size</td>
<td>Smallest</td>
<td>Smaller</td>
</tr>
<tr>
<td>Nodes/Net</td>
<td>&lt;BT</td>
<td>BT</td>
</tr>
<tr>
<td>Cost</td>
<td>~$1</td>
<td>~$10–$15</td>
</tr>
</tbody>
</table>

Table 2: Modulation Parameters of IEEE 802.15.4 (Callaway et al. 2002)
Table 3: Type of Devices in Zigbee Network (Le 2005)

<table>
<thead>
<tr>
<th>Physical device type</th>
<th>Logical device type distinguishes the physical devices deployed in a Zigbee network.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinator (FFD)</td>
<td>Network establishment and control.</td>
</tr>
<tr>
<td>Router</td>
<td>Support data routing functionality. Can talk to other routers, coordinators, and end devices.</td>
</tr>
<tr>
<td>End Device</td>
<td>Can only talk to routers or the coordinator.</td>
</tr>
</tbody>
</table>

Table 4: Mapping of Symbol to Chip for OQPSK - 2.4GHz Band (Shuaib et al. 2007)

<table>
<thead>
<tr>
<th>Data symbol (dec.)</th>
<th>Data symbol (bin.)</th>
<th>Chip values (c0 c1...c31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0000</td>
<td>110111001111100011010100011001</td>
</tr>
<tr>
<td>1</td>
<td>1000</td>
<td>111011101101001110001101001001</td>
</tr>
<tr>
<td>2</td>
<td>0100</td>
<td>001110111011111001101010011101</td>
</tr>
<tr>
<td>3</td>
<td>1100</td>
<td>001001011100011011010110011001</td>
</tr>
<tr>
<td>4</td>
<td>0010</td>
<td>010110001010110011011010011001</td>
</tr>
<tr>
<td>5</td>
<td>1010</td>
<td>010110101001100110100110011001</td>
</tr>
<tr>
<td>6</td>
<td>0110</td>
<td>110001111010100111101101010101</td>
</tr>
<tr>
<td>7</td>
<td>1110</td>
<td>100110001011000011101101100110</td>
</tr>
<tr>
<td>8</td>
<td>0001</td>
<td>100011001101010011011111111111</td>
</tr>
<tr>
<td>9</td>
<td>1001</td>
<td>101111111110100111101101110111</td>
</tr>
<tr>
<td>10</td>
<td>0101</td>
<td>011111111111111111111111111111</td>
</tr>
<tr>
<td>11</td>
<td>1101</td>
<td>011111111111111111111111111111</td>
</tr>
<tr>
<td>12</td>
<td>0011</td>
<td>000001111111111111111111111111</td>
</tr>
<tr>
<td>13</td>
<td>1111</td>
<td>011111111111111111111111111111</td>
</tr>
<tr>
<td>14</td>
<td>0111</td>
<td>100111000001111111111111111111</td>
</tr>
<tr>
<td>15</td>
<td>1111</td>
<td>111111111111111111111111111111</td>
</tr>
</tbody>
</table>

Table 5: Mapping of Symbol to Chip for Enhanced OQPSK–868MHz/915MHz Band (Shuaib et al. 2007)

<table>
<thead>
<tr>
<th>Data Symbol (dec.)</th>
<th>Data Symbol (bin.)</th>
<th>Chip Values (c0 c1...c15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0000</td>
<td>0011111110001101</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>0101110001001001</td>
</tr>
<tr>
<td>3</td>
<td>0100</td>
<td>0110100111010100</td>
</tr>
<tr>
<td>4</td>
<td>1100</td>
<td>1001010011101001</td>
</tr>
<tr>
<td>5</td>
<td>0010</td>
<td>0011010110110100</td>
</tr>
<tr>
<td>6</td>
<td>1010</td>
<td>1000101101011000</td>
</tr>
<tr>
<td>7</td>
<td>1110</td>
<td>1111000110101010</td>
</tr>
<tr>
<td>8</td>
<td>0001</td>
<td>0111010101110000</td>
</tr>
<tr>
<td>9</td>
<td>1001</td>
<td>0001101111011100</td>
</tr>
<tr>
<td>10</td>
<td>0101</td>
<td>0001101110111110</td>
</tr>
<tr>
<td>11</td>
<td>1101</td>
<td>1100011101110101</td>
</tr>
<tr>
<td>12</td>
<td>0011</td>
<td>0111100101111001</td>
</tr>
<tr>
<td>13</td>
<td>1111</td>
<td>1111111101110111</td>
</tr>
<tr>
<td>14</td>
<td>0111</td>
<td>1001110000011110</td>
</tr>
<tr>
<td>15</td>
<td>1111</td>
<td>1011011111111110</td>
</tr>
</tbody>
</table>
Fig 9: BER versus the SNR for different data rate values for Enhanced OQPSK – 868MHz/915MHz Band (Shuaib et al. 2007)

Table 6: Mapping of Symbol to Chip for BPSK – 868MHz/915MHz Band (Shuaib et al. 2007)

<table>
<thead>
<tr>
<th>Input bits</th>
<th>Chip values (c0...c14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>11110101001000</td>
</tr>
<tr>
<td>1</td>
<td>00010100110111</td>
</tr>
</tbody>
</table>

Fig 10: BER versus the SNR for different data rate values for BPSK – 868MHz/915MHz Band (Shuaib et al. 2007)

Table 7: Factors affecting BER (Poole NA)

<table>
<thead>
<tr>
<th>Factors</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interference</td>
<td>Due to external factors and cannot be altered</td>
</tr>
<tr>
<td></td>
<td>Reduce the system bandwidth in order to reduce the interference</td>
</tr>
<tr>
<td></td>
<td>Setback: When bandwidth is reduced, the throughput of the data is limited</td>
</tr>
<tr>
<td>Increase in Power of</td>
<td>Increase in power level for transmission will increase the data power per bit</td>
</tr>
<tr>
<td>Transmission</td>
<td></td>
</tr>
<tr>
<td>Lower Order Modulation</td>
<td>Can be used with the expense of throughput of data</td>
</tr>
<tr>
<td>Reduce Bandwidth</td>
<td>When bandwidth is reduced, so does the BER and the level of received noise</td>
</tr>
<tr>
<td></td>
<td>When level of received noise is reduced, the SNR will improve</td>
</tr>
<tr>
<td></td>
<td>Setback: The throughput of the data is reduced</td>
</tr>
</tbody>
</table>

Fig 11: Electromagnetic Spectrum (Electromagnetic spectrum 2012)

Fig 12: Rayleigh Distribution
ABBREVIATIONS
AES - Advanced Encryption Standard
AWGN - Additive White Gaussian Noise
BER - Bit Error Rate
BPSK - Binary Phase Shift Keying
CSMA-CA - Carrier Sense Multiple Access - Collision Avoid
CRC - Cyclic Redundancy Check
DSSS - Direct Sequence Spread Spectrum
FCS - Frame Checksum
FFD - Full Function Devices
FHSS - Frequency Hopping Spread Spectrum
IEEE - Institute of Electrical and Electronics Engineers
ISM - Industrial, Scientific and Medical
LR-WPAN - Low rate-wireless personal area network
MAC - Media Access Control
NIST - National Institute of Standards and Technology
NFC - Near Field Communication
OQPSK - Offset Quadrature Phase Shift Keying
PAN - Personal area network
PHY - Physical
PSD - Power Spectral Density
RFD - Reduced Function Devices
RFID - Radio-frequency Identification
SNR - Signal-to-Noise ratio
USB - Universal Serial Bus
UWB - Ultra Wide Band
WLAN - Wireless Local Area Network
WPAN - Wireless Personal Area Network
ABSTRACT

IEEE 802.15.1 and IEEE 802.15.4 namely Bluetooth and ZigBee respectively, are operating in the unlicensed 2.4 GHz frequency band, which are becoming more and more popular. The numbers of devices equipped with Bluetooth are growing drastically in numbers reaching into the billions. ZigBee’s sensing capabilities have been identified and found niche in smart home technology. The result is the number of co-located devices, say within 10 meters, grown to a limit, which may cause interference issues in the 2.4 GHz radio frequency spectrum. In this paper, we have performed a detailed survey on the interference issues of 2.4 GHz frequency band and their subsequent effects on throughput of Bluetooth.

INTRODUCTION

The IEEE 802.15 standard has been developed to establish a roadmap for development of Personal Area Network, which is meant for modest-sized geographical areas. The IEEE 802.15.1 is the standard made for Bluetooth technology and it’s the most widely used in short range data sharing technology boasting more than 2 billion devices worldwide. It is also a continuously evolving technology. The latest Bluetooth 4.0 with its high energy efficiency and high data rates of up to 24 Mbps, provides the best data sharing experience yet. The IEEE 802.15.4 standard offers fundamental lower network layers of wireless personal area network (WPAN) with low-cost, low-speed pervasive communication between devices. There is no underlying infrastructure to emphasize its low cost and low power consumptions. The basic framework conceives a 10-meter communications range with a transfer rate of 250 Kbit/s. The physical medium is accessed through a Carrier Sense Multiple Access Collision Avoidance (CSMA/CA) protocol.

In this paper, we have involved an understanding of the function of Bluetooth technology and make an analysis of previous work has been done to understand the interference effect in different environmental conditions.

PROBLEM FORMULATION

The 2.4 GHz is the unlicensed spectrum, which means that communications taking place over this spectrum are monetarily not charged. Because of this, most of the personal area network related technologies operate in this spectrum and Bluetooth and ZigBee are no exceptions. There are many other technologies that are operating in this spectrum, such as Wi-Fi, Cordless phones, Microwave ovens, Wireless USB and etc., are just an example. As a result the interference levels on this spectrum are very high and some studied showed that this spectrum is also susceptible to rain due to absorption of parts of this spectrum and due to moisture. Radio frequency waves have also been seen to manoeuvre really well in an area covered with trees. In this research project, we want to investigate the interference effects of environmental conditions on Bluetooth and ZigBee.

LITERATURE REVIEW

Comparison of different IEEE 802 Standards

Initially, we have performed a comparison of different technologies in the IEEE 802 standards and have discussed their similarity and differences with our findings. There are some differences and similarities between the IEEE 802.11, 802.15.1, 802.15.4 and 802.15.6 wireless standards which emphasis on the physical layer.

- The IEEE 802.11 is the IEEE standard for Wi-Fi
- The IEEE 802.15.1 is the IEEE standard for Bluetooth
- The IEEE 802.15.4 is the IEEE standard for ZigBee when application layer is incorporated along with data link and physical layers.
- The IEEE 802.15.6 is the IEEE standard for Wireless Body Area Networks BAN

Modes of operation

Wireless networks have distinct modes of operation such as the ad-hoc networks and the infrastructure networks. The Infrastructure wireless networks usually have kind of base station, which acts as a central node to connect with wireless terminals. The base station has a fixed location and usually provides
Internet access to an intranet or other wireless networks, therefore; it causes the problem of central point of failure, meaning if central point stops working none of the wireless terminals can communicate with each other. Whereas, ad-hoc networks formed without help of a base station. In ad-hoc networks the wireless terminals may communicate directly with each other, while in terminal infrastructure networks have to use the base station to relay with their messages. These different standards have different capabilities when it comes to modes of operation. Table 1 contains an overview of which different standards support which modes. (Tjensvold, 2007).

Medium Access

In order to conduct a two-way wireless communication, a detailed protocol on how the wireless terminals should access medium must need to be established. The 802.11 standards specifies a method called Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) to solve above issues. The CSMA/CA technique allows wireless transmissions to be completed without interference from other terminals. If the communication was one-way only the communication channels could be better utilized because the source terminal would not have to take into consideration transmissions being sent from other terminals. In order to deal with the hidden terminal problem 802.11 can also use a request to send; clear to send (RTS/CTS) scheme in order to reserve the channel for communication. (Tjensvold, 2007).

Frequency, data rate and range

The Table 2 shows several standards operate in the 2.4 GHz band. 802.11g is designed to compatible with 802.11b in order to co-exist. However, when 802.11g base station is operated in 802.11b mode, the data rate will reduce for all connected terminals. 802.15.1 also occupies the 2.4 GHz band with a different signalling method to avoid heavy interference. Along with these two standards, some others common household appliances like microwave ovens and wireless phones also pollute the 2.4 GHz frequency spectrum. The 802.11n does have a high data of 248 Mbps in the same frequency band as the other standards. So far the major large increase in data rate and range is achieved by Multiple-Input Multiple-Output (MIMO). MIMO can handle more than one sender and receiver antennas and combines this with special coding techniques in order to squeeze even more data through the same frequencies. (Tjensvold, 2007).

Modulation and signal coding

When transmitting binary data over the wireless medium, it needs to go through a digital to analogue modulation process; however, when the analogue signal is received, it will be converted back into the digital form to retrieve the binary data. The standards use several methods to accomplish this seemingly straightforward task. The modulation techniques for a standard are decided, based on the design goals of the standard and the targeted deployment environment.

Each technique has different characteristics when it comes to:
- Noise resistance (multi path, fading, etc.)
- Power consumption
- Hardware complexity
- Spectral efficiency (efficient frequency use)

64-QAM () is used by 802.11a/g/n messages to map 6 data bits into each transmitted symbol. This is a high order modulation, which provides high spectral efficiency. The disadvantage of using 64-QAM is that the receiver requires a higher Signal to Noise ratio (SNR) to demodulate the signal than comparable systems using BPSK and QPSK. This means that general high order modulations require higher transmission power to achieve the same Bit Error Rate (BER) as a lower order modulation like BPSK and QPSK. The use of BPSK and OQPSK (similar to QPSK) in 802.15.4 reflects this fact, as the design goal of 802.15.4 was to create devices with low power consumption, which can reduce the power requirements to a minimum. (Tjensvold, 2007).

Bluetooth: History & Background

The name "Bluetooth" comes from the 10th century "Danish King Harald Blåtand or Harold Bluetooth" in English. The King was instrumental in uniting in the warring factions in Northern Europe or more commonly known as the Nordic countries, of what are now Norway, Sweden and Denmark. Similarly, Bluetooth technology was created as an open standard to allow connectivity and collaboration between disparate products and industries.

A cheap, power-efficient radio chip that provides seamless local connectivity, and creates a (worldwide) micro-scale web. Hence, in order to connect mobile phones and their accessories Ericsson Mobile Communications AB in Lund, Sweden, initiated a study to investigate the feasibility of a low-power, low-cost radio interface. The main idea was to rid ourselves from wires or more specifically RS-232 cables and to provide human accessories like wireless handsets and no wired connection between a mobile phone and computer and so on. The study was part of a larger project that investigated multi-communicators connected to the cellular network via cellular telephones. And we have seen their results in today's market. We have delve some wireless connectivity technology that took place almost two decades ago in 1994, in order to facilitate our research (Haartsen and Mattisson, 2000).

Prior to Bluetooth, Infrared was mainly used for short distance data sharing. But due to its very limited range, sensitivity to direction, requirements of direct line of sight and its limitation to only two devices made it out of
favourite, after Bluetooth came into the commercial market. Bluetooth being a radio communication technology provides much greater range, can propagate around objects and through various materials, and connect many devices simultaneously. Bluetooth, thus helped get around the problems that came with infrared systems.

**Bluetooth SIG**

The Bluetooth Special Interest Group (SIG), is the pioneer in the telecommunications, computing, consumer electronics, automotive and network industries. These are at the helm and driving the development of Bluetooth wireless technology and bringing it to market. The Bluetooth SIG, Inc. headquarters is located in Bellevue, Washington, U.S.A. The Bluetooth SIG Promoter group companies include Ericsson, Intel, Lenovo, Microsoft, Motorola, Nokia and Toshiba, along with over 15,000 Associate and Adopter member companies (Bluetooth.org, 2012).

**Different Bluetooth Versions**

Bluetooth has been an evolving technology, for the past two decades since its came into the market. It evolves much evolution in the form of data rate enhancements and higher power efficiency.

The older Bluetooth 1.0 standard was the first version to come out from the Special Interest Group, with a maximum transfer speed of 1 Megabit per Second (Mbps). However, the following version could manage up to 3 Mbps. Bluetooth 2.0 is backward-compatible with 1.0 devices. Thus, the older devices did not become obsolete and were also useful. The big development in Bluetooth was Bluetooth 3.0 standard, which can support a massive data rate of 24Mbps. The Bluetooth 3.0 specification was an update from the Bluetooth 2.1 protocol, which Bluetooth SIG adopted in 2007. This high spike in data rate for Bluetooth 3.0 is attributed to the inclusion of the 802.11 radio protocol. In particular, 802.11 Protocol Adaptation Layer (PAL) is incorporated, which provides for increased throughput of data transfers at the approximate rate of 24 Mbps. Including this radio feature of 802.11 gives the developers more freedom to use the functions of the two radio’s in the device. The specifications of Bluetooth 3.0 include backward compatibility, enabling both the enhancement and expansion of the technology. However, there is a drawback, version 3 has higher power consumption, and this is due to the 802.11 radio, which is combined with it in order to facilitate high data rates. But it is incumbent to note that when the device is idle this feature is not part of the radio anymore, thus idle power usage by the device becomes remarkably low (Bennet, 2012).

Bluetooth 4.0 was later released adopted in 2010, which provided much needed improvement in the area of power. This version was very energy efficient and primarily designed for devices that collect information frequently, either as much as five times a second or just once every hour, the standard is optimized for long battery life. Longevity was increased from 5 to 10 times depending on individual usage. In wireless keyboards and mouse, it would take five to seven years before needing a fresh battery--likely outlasting the life of the product. (Bennet, 2012).

Another convenience newer form of Bluetooth brings is compatibility with NFC chips. Both Bluetooth versions 3.0 and 4.0 can talk to NFC hardware in phones and laptops to make pairing a simple process of tapping the two devices together. For example, Motorola's new Elite Sliver headset already has this NFC pairing ability. (Bennet, 2012).

**License free spectrum usage**

The Bluetooth is operated in a free of cost frequency band such as 2.45 GHz, the Industrial-Scientific- Medical (ISM) band, which ranges from 2,400 to 2,483.5 MHz in the US and Europe (only parts of this band are available in France and Spain), and from 2,471 to 2,497 MHz in Japan. Consequently, the system can be used worldwide, given that the radio transceivers cover the frequency band between 2,400 and 2,500 MHz and that they can select the proper segment in this band. (Haartsen and Mattisson, 2000, pp. 1652).

**Frequency Hopping**

The ISM band is open to everyone and thus radio systems operating in this frequency must cope with several unpredictable sources of interference, such as baby monitors, wireless keyboards and mouse, cordless phones and microwave ovens. Interference can be avoided using an adaptive scheme that finds an unused part of the spectrum, or it can be suppressed by means of spectrum spreading. In the United States, radios operating in the 2.45-GHz ISM band are required to apply spectrum-spreading techniques if their transmitted power level exceeds 0 dBm (Arrl.org, 1999). Bluetooth radios use frequency-hop (FH) spread spectrum, since it better supports low-cost, low-power radio implementations. In addition, they better cope with near-far problems. The frequency band is divided into several hop channels using FHSS. After a connection has been established, radio transceivers hop from one channel to another in a pseudo random fashion. The hops are small but they are spread over the entire spectrum. This results in low-cost narrow-band transceivers with maximum immunity to interference. Rarely does it happen that interference jams a hop channel, which leads to faulty reception. When this occurs, error-correction schemes in the link restore the bit errors. (Haartsen and Mattisson, 2000, pp. 1652).

**Channel Definition**

Bluetooth channels use a FH/time division duplex (FH/TDD) scheme (Fig. 4). It splits the 2.4 GHz ISM
band into 79.1 MHz channels. The devices hop among the 79 channels 1600 times per second in a pseudo-random pattern. The channel is divided into consecutive slots, each slot lasting 625 μs; a different hop channel is used for each slot. One packet can be transmitted per slot. Subsequent slots are alternatively used for transmitting and receiving, which results in a TDD scheme as it better matches the peer communication concept. Two or more units sharing the same channel form a piconet. One unit acts as a master, while the other units are slaves and it is the master that controls the traffic on the piconet. The FH channel is determined by the FH sequence (the order in which the hop channels are visited) and by the phase in this sequence. In Bluetooth, the sequence is determined by the identity of the piconet master and phase is determined by the master unit’s system clock. In order to recreate the master clock in the slave unit, the slave may add an offset to its own native clock. The FH sequence is very long and its repetition time exceeds 23 h. If every participant on a given channel uses the same identity and clock as input to the hop-selection box, then each unit will consistently select the same hop channel and remain synchronized. Every piconet has a unique set of master parameters, which create a unique channel. The channel makes use of several, equally spaced, 1-MHz hop channels. With Gaussian-shaped frequency shift keying (FSK) modulation, a symbol rate of 1 Mb/s is achieved. In the 2.45-GHz ISM band, 79 hop channels have been defined. On average, the FH sequence visits each carrier with equal probability. (Haartsen and Mattisson, 2000, pp. 1652).

Physical Link Definition

Two types of links have been defined to support multimedia applications that mix voice and data:

- Synchronous connection-oriented (SCO) link
- Asynchronous connectionless (ACL) link

SCO links support symmetrical, circuit-switched, point-to-point connections typically used for voice. These links are defined on the channel by reserving two consecutive slots (forward and return slots) with a fixed period. Reservation is carried out by the master and the slave when the link is set up. ACL links support symmetrical or asymmetrical, packet-switched, point-to-multipoint connections typically used for burst data transmission. Master units use a polling scheme to control the ACL connections and to prevent collisions on the channel when multiple slaves should transmit. All SCO and ACL traffic on the channel is scheduled by the master. Local slave addresses in the packet header indicate the recipient and only this recipient is allowed to respond in the next slave-to-master slot. The ACL link is constantly present between the master and the slave as long as the piconet exists; the ACL link conveys both control information and asynchronous data services. In contrast, SCO links can be set up and released dependent of the need for synchronous services. . (Haartsen and Mattisson, 2000, pp. 1653).

Ad hoc Connectivity

Bluetooth is an example of ad hoc connectivity. That means a device with Bluetooth radio on it can communicate with device that also has a Bluetooth connectivity enabled on it. There is no wired infrastructure with base stations or access points that can support the call setup or can provide low-power modes. Bluetooth does all this with short setup time and boasts a low-power standby mode. (Haartsen and Mattisson, 2000, pp. 1654)

The 2.4 GHz band being an unlicensed free band is a host to many different wireless applications. Thus an ample source of interference to Bluetooth is available in this spectrum. However, as we have seen how Bluetooth with the help of FHSS avoids interference. But guaranteeing a hundred percent throughput without any loss and error is not possible. Hence, there have been many studies carried out to analyse the throughput of Bluetooth devices in different interference environments. The following are the most prominent researchers, which has been carried out on interference effects on Bluetooth performance.

EXPERIMENTS AND DISCUSSION

First Experiment Reviewed

In the following paper the researcher (Punnoose and Tseng et al., n.p) has published experimental results on interference between Bluetooth and IEEE 802.11b DSSS system.

A piconet was set up between the two Bluetooth laptops and an IP based network connection was established. UDP data was transferred from the Bluetooth transmitting laptop to the receiving laptop using the simulation program. Each transmitted packet contained 289 bytes of user data. The packets were chosen to be able to efficiently use the largest available Bluetooth frames, the DH5 frames. The packets were generated at a rate equivalent to a 250 Kbit/sec stream (user data). For each measurement, 10 Bytes of data were sent by the transmitting node. Statistics were collected at the receiving Bluetooth laptop. Interference was provided with an 802.11b laptop sending UDP traffic at a bit rate of 5 Mb/sec. (Punnoose and Tseng et al., n.p). Figure 3 shows the Bluetooth performance under a range of interference levels. Since the Bluetooth does not have acknowledgements, retransmissions, and does not perform carrier sensing before transmission, the attained throughput is directly related to the percentage of lost packets. A packet loss rate of 20% results in a throughput of 80% of the maximum. Since each UDP packet that was transmitted from the Bluetooth was contained in a single
DH5 frame, one would expect the Bluetooth to overlap in frequency and time with the 802.11b at most 1/3 of the time, and therefore the packet loss rate would not exceed this. However, our laboratory measurements with fixed-frequency CW interference signals suggest the much larger packet loss rate results from the limited frequency selectivity of the Bluetooth receivers. Bluetooth performance starts to degrade rapidly when the interfering 802.11b signal is compared to the desired signal, and reduction in data throughput tracks the packet loss rate since the Bluetooth protocol does not attempt carrier sensing. (Punnoose and Tseng et al., n.p)

Second Experiment Reviewed

In the following experiment, the channels on the interfering devices were chosen so that their spectrum co-inside with each other. Channel 6 operating at a central carrier frequency of 2437 MHz was chosen on the IEEE 802.11g AP, and Channel 17 operating at a central carrier frequency of 2435 MHz was chosen on the ZigBee/Bluetooth devices. Three test cases were conducted: the first test case was run with the IEEE 802.11g and ZigBee/Bluetooth devices. The second test case was run with the ZigBee/Bluetooth devices placed one in a cubical and the other in cubical approximately 6 meters apart while the IEEE 802.11g client was in the reference cubical. The third test case was the second one, with the ZigBee/Bluetooth devices placed approximately 12 meters apart. The results of these experiments are summarized in Table 3 and Table 4. (K. Shuaib, 2006).

We can analyse from the results in Figure 3, there was no significant effect on the performance of the IEEE 802.11g due to interference from the Zigbee for all conducted test cases. However, that was not true for the Zigbee devices since the throughput was affected in all test cases with case three being the worst. Figure 10 shows how Bluetooth, as expected, greatly affected the performance of the IEEE 802.11g client and vice versa the performance of Bluetooth was also greatly affected by the presence of an IEEE 802.11g network in close proximity. (K. Shuaib, 2006)

Second Experiment Reviewed

In this paper, the first set of experiments was done to study the interference effect when Zigbee or Bluetooth devices are placed within a close proximity from an IEEE 802.11g client. These tests basically looked at the impact on the IEEE 802.11g down link channel from the IEEE 802.11g AP to the IEEE 802.11g client. In this experiment, we have observed how such devices might affect the performance of IEEE 802.11g when placed near the AP rather than the WLAN client, i.e. affecting the uplink between the WLAN AP and its client. With the IEEE 802.11g client residing in the reference cubical, the Zigbee or Bluetooth devices were placed few meters apart at several positions on the same horizontal line as the IEEE 802.11g AP but on a 50 cm high table. Figure 11 summarizes the results for these test cases. (K. Shuaib, 2006).

To see the effect of Zigbee and Bluetooth on WLAN, in an environment where the IEEE 802.11g signal strength is weak, they emulated positioning of the IEEE 802.11g client at a distance considered far from the AP. This was done by placing the AP behind an obstacle that brought down the signal strength indicator on the IEEE 802.11g card software fluctuating around -80 dbm. At this signal strength level, a test was performed with all interfering devices were placed within the reference cubical in the AP operating on Ch6 and the Zigbee devices operating on Ch17. The baseline performance of the IEEE 802.11g client without any source of interference and at this signal strength level was measured at 6.7 Mbps. Table 6 summarizes the obtained results for this test case. As seen in Figure 12, there was a slight effect on the performance of the 802.11g client for this test case due to interference from Zigbee; however, the effect of Bluetooth was drastic with a 53% drop in the throughput compared to 22% drop when the IEEE 802.11g signal strength was at full strength around -40 dbm as was shown in Table 4. (K. Shuaib, 2006).

**DISCUSSION AND ANALYSIS**

**Bit Error Rate**

Bit Error Rate BER, is the probability of an error occurring in a given set of bits being sent in data transfer.

\[
BER = \frac{\text{number of Errors}}{\text{total number of bits sent}} \quad [1]
\]

BER is used to assess the performance of any given system. It covers full end to end performance of the system, including the transmitter, receiver and the medium between them. Three variables can be used to realize the BER of any communication system such as:
- The error function
- The energy in one bit (Eb)
- Noise Power Spectral Density (No) or noise power in 1Hz Bandwidth

We know from Literature Review, Higher order modulations like 64QAM carry higher data rates but are not robust in the presence of noise, whereas lower order modulations like QPSK offer lower data rates but are more robust.

**Interference**

A thorough study is carried out to understand the interference in 2.4 GHz Frequency Band in order to understand the spectrum constraints that Bluetooth technology had to deal with.
The two methods for radio frequency modulation in the unlicensed 2.4 GHz ISM band are frequency-hopping spread spectrum (FHSS) and direct-sequence spread spectrum (DSSS).

Wi-Fi uses DSSS, with each channel being 22 MHz wide, which allows three evenly-distributed channels to be used simultaneously sans overlapping.

Bluetooth uses FHSS and splits the 2.4 GHz ISM band into 79 1 MHz channels. Bluetooth devices hop among the 79 channels 1600 times per second in a pseudo-random pattern.

WirelessUSB uses DSSS. Each WirelessUSB channel is 1 MHz wide, allowing WirelessUSB to split the 2.4 GHz ISM band into 79 1 MHz channels similar to Bluetooth. WirelessUSB devices use a "fixed" channel, but dynamically change channels if the link quality of the original channel becomes suboptimal.

ZigBee also uses the 2.4 GHz ISM band (available worldwide), and has sixteen channels are defined; each channel occupies 3 MHz and channels are centred 5 MHz from each other, giving a 2-MHz gap between pairs of channels.

The cordless phones may use either DSSS or FHSS. Most 2.4 GHz cordless phones use a channel width of 5 to 10 MHz; Phones using DSSS and other fixed channel algorithms typically have a "channel" button on the phone allowing users to manually change the channel. FHSS phones do not have this feature because of the nature of FHSS which changes channels dynamically (Gerrior and Woodings, 2006).

The DSSS systems have the most to lose because of the danger of overlapping with another DSSS system. But there are things DSSS systems can do to obtain the frequency agility by network monitoring. If the DSSS system uses a polled protocol (where packets are expected at specified intervals) then the master can switch channels after a number of failed transmit attempts or bad received packets.

Another approach is to take a reading of the energy level on the air if the radio has this capability. A receive strength signal indicator (RSSI) can be used to proactively measure the amount of energy in the air and if that level is too high over a period of time switch to a clearer channel. A period of time is taken into account so as not to change channels if a FHSS system is passing through. (Gerrior and Woodings, 2006)

Collision avoidance

Wi-Fi's collision-avoidance algorithm (CSMA/CA) listens for a quiet channel before transmitting. This allows multiple Wi-Fi clients to efficiently communicate with a single Wi-Fi access point. If the Wi-Fi channel is noisy the Wi-Fi device does a random back off before listening to the channel again.

Interference from Bluetooth is minimal due to the hopping nature of the Bluetooth transmission. A single active Wi-Fi network can cause heavy interference on 25% of the Bluetooth channels. Packets lost due to overlap have to be retransmitted on quiet channels, thereby greatly reducing the throughput of Bluetooth devices. Bluetooth uses Adaptive frequency hopping (AFH) algorithm. This algorithm allows devices to mark channels as good, bad, or unknown. Bad channels in the frequency-hopping pattern are then replaced with good channels via a lookup table. The AFH algorithm allows Bluetooth to avoid channels occupied by DSSS systems such as Wi-Fi and WirelessUSB. However 2.4 GHz FHSS cordless phones may still cause interference with Bluetooth since both systems are hopping over the entire 2.4 GHz ISM band, but since the Bluetooth signal is only 1 MHz wide the frequency of collisions between the FHSS cordless phone and Bluetooth is significantly less. (Gerrior & Woodings, 2006).

CONCLUSION

In this paper a survey of experiments and measurements was done to quantify the interference effect of Bluetooth devices on the throughput performance of the IEEE 802.11g and vice versa. The results show how IEEE 802.11g is greatly more affected by Bluetooth than ZigBee. Our intended future work will be performing practical tests under different conditions to further characterize the interference effect of Zigbee/Bluetooth/WLAN on each other and to provide a performance model for an environment where all three technologies can co-exist. Overall these experiments will provide a better understanding on simulation and physical implementation of Bluetooth.

REFERENCES


FIGURES AND TABLES

Table 1: Modes of Operation for different Wireless Standards

<table>
<thead>
<tr>
<th>Standard</th>
<th>Ad hoc</th>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.11a/b/g/n</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>802.15.1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>802.15.4</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>802.15.6</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

1 Specified for outdoor environments with few obstructions.
2 802.11n device with two streams (four antennas).
3 Bluetooth version 2.0.
4 Bluetooth class 1 device.
5 All 802.15.6 values are unconfirmed.

Table 2: Comparison of Different wireless standards

<table>
<thead>
<tr>
<th>Standard</th>
<th>Frequency</th>
<th>Data Rate1</th>
<th>Range Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.11g1</td>
<td>5 GHz</td>
<td>54 Mbps</td>
<td>120m, LAN</td>
</tr>
<tr>
<td>802.11g1</td>
<td>2.4 GHz</td>
<td>11 Mbps</td>
<td>140m, LAN</td>
</tr>
<tr>
<td>802.11g1</td>
<td>2.4 GHz</td>
<td>54 Mbps</td>
<td>140m, LAN</td>
</tr>
<tr>
<td>802.11n</td>
<td>2.4 GHz</td>
<td>248 Mbps²</td>
<td>250m, LAN</td>
</tr>
<tr>
<td>802.15.1</td>
<td>2.4 GHz</td>
<td>3 Mbps³</td>
<td>100m², PAN</td>
</tr>
<tr>
<td>802.15.4</td>
<td>868/915 MHz</td>
<td>40 kbps</td>
<td>75m, PAN</td>
</tr>
<tr>
<td>802.15.6²</td>
<td>2.4 GHz</td>
<td>250 kbps</td>
<td>10m, BAN</td>
</tr>
</tbody>
</table>

Figure 1: Hop channel selection scheme: In the selection box, the master identity selects the sequence and the clock selects the phase to result in the current hop channel to be used.
Table 3: Result for the three test cases

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Percentage drop in IEEE 802.11g throughput</th>
<th>Percentage drop in Zigbee throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Insignificant</td>
<td>10% (from 100% to 90%)</td>
</tr>
<tr>
<td>2</td>
<td>Insignificant</td>
<td>10% (from 100% to 90%)</td>
</tr>
<tr>
<td>3</td>
<td>Insignificant</td>
<td>22% (from 83% to 65%)</td>
</tr>
</tbody>
</table>

Table 4: Result for three test cases with Bluetooth replacing ZigBee

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Percentage drop in IEEE 802.11g throughput due to Bluetooth</th>
<th>Percentage drop in Bluetooth throughput due to IEEE 802.11g</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12% (from 9.8 Mbps to 8.6 Mbps)</td>
<td>21% (from 554 kbps to 440 kbps)</td>
</tr>
<tr>
<td>2</td>
<td>6% (from 9.8 Mbps to 9.2 Mbps)</td>
<td>36% (from 512 kbps to 328 kbps)</td>
</tr>
<tr>
<td>3</td>
<td>4.6% (from 9.8 Mbps to 9.35 Mbps)</td>
<td>17% (from 365 kbps to 303 kbps)</td>
</tr>
</tbody>
</table>

Table 5: Test Results for IEEE 802.11g weak signal

<table>
<thead>
<tr>
<th>D2 (meters)</th>
<th>Percentage drop in 802.11g throughput due to Zigbee</th>
<th>Percentage drop in 802.11g throughput due to Bluetooth</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>11% (from 9.8 to 8.7 Mbps)</td>
<td>19% (from 9.8 to 7.9 Mbps)</td>
</tr>
<tr>
<td>6</td>
<td>6% (from 8.8 to 9.2)</td>
<td>17% (from 9.8 to 8.1 Mbps)</td>
</tr>
<tr>
<td>8</td>
<td>Insignificant</td>
<td>20% (from 9.8 to 7.8 Mbps)</td>
</tr>
</tbody>
</table>

Table 6: Test Results

<table>
<thead>
<tr>
<th>Percentage drop in IEEE 802.11g throughput when using Zigbee</th>
<th>Percentage drop in IEEE 802.11g throughput when using Bluetooth</th>
</tr>
</thead>
<tbody>
<tr>
<td>6% (from 6.7 to 6.3 Mbps)</td>
<td>52% (from 6.7 to 3.2 Mbps)</td>
</tr>
</tbody>
</table>

Figure 2: Bluetooth frequency-hop/time division duplex channel

Figure 3: Performance of Bluetooth in the presence of interference.

Figure 4: Signal Comparison of wireless systems in 2.4GHz Band
SOLUTION GENERATION IN CONCEPTUAL DESIGN USING ENVIRONMENT-BASED DESIGN: ALGORITHM AND CASE STUDY

Ronaldo Gutierrez, Yong Zeng, Xuan Sun, Suo Tan, Xiaoguang Deng
Concordia Institute for Information Systems Engineering
Concordia University
Montreal, Quebec, Canada, H3G 1M8

Fayi Zhou
Design Services, Drainage Design and Construction
The City of Edmonton
Edmonton, Alberta, Canada, T5M 3B8

ABSTRACT

Environment-based design (EBD) methodology has three activities: environment analysis, conflict identification, and solution generation. This paper focuses in solution generation. An algorithm for solution generation is presented and used in a case study. Such algorithm and application is missing in the literature. The steps in the algorithm are: decompose the environment, find design knowledge and generate solutions. The case study consists in developing a Total Quality Management System (TQMS) for Area Development Planning (ADP), a subsection of the City of Edmonton. The TQMS must comply with ISO 9001:2008 requirements.

INTRODUCTION

Environment-Based Design (EBD) methodology includes three activities: environment analysis, conflict identification and solution generation (Zeng, 2011). The scope of this paper covers solution generation, which is after environment analysis and conflict identification activities. The output of the two last activities are a Product-Environment System (PES) in Recursive Object Model (ROM) representation, and existing conflicts to satisfy the design problem requirements. Most specifically, a conflict in EBD is a lack of available resources for a system to meet a request on its actions or its properties (Zeng, to appear 2014). A conflict appears when there is a potential failure in a PES. The failure in a system may be attributed to: 1) absence of a necessary action on an object; 2) opposite actions on the same object; and 3) lack of resources to accommodate an object and its actions (Zeng, to appear 2014). Conflicts are implied in a PES in ROM representation. Procedures for building ROM representations can be found in the paper (Zeng, 2008). Guidance to infer a PES from a ROM representation is provided in the reference (Wang, et al., 2013). The PES is composed of a product, its environment components and their mutual relationships (Wang, et al., 2013). The PES denotes the environment in which a desired product is to work.

“A Conceptual design process begins with design requirements and ends with product descriptions of design concepts” (Zeng, 2001). The output PES of the environment analysis implies design requirements and partial solutions to satisfy the requirements (Zeng, 2004). Conflict identification strives to identify lack of available actions and resources for the system to satisfy unmet requirements. These actions and resources lay the foundation for solutions generation. Solution generation will solve the identified conflicts leading to the creation of design concepts.

The rest of the paper contains two sections. One section introduces an algorithm for solution generation. At the same time, a case study using the proposed algorithm to develop a TQMS system is presented. The remaining section concludes about the proposed algorithm and the case study.

SOLUTION GENERATION: ALGORITHM AND CASE STUDY

Solution generation activity in EBD solves conflicts implied in a PES. Theoretically, conflict identification finds conflicts and dependencies between them. The dependencies provide the direction to follow for solving the conflicts. After solving a conflict, the solution becomes part of the PES. Subsequent conflicts are constrained by previous generated solutions.

Fig. 1 shows the proposed algorithm for solution generation. The inputs to the algorithm are a PES and conflicts with dependencies from environment analysis and conflict identification activities in EBD. The algorithm includes three steps: decompose the environment, find design knowledge and generate
solution. The output is solution generation to solve any existing conflict.

Fig. 1 Algorithm for solution generation

Each of the steps in the algorithm are presented in next sections. The case study is used to illustrate each of the step. The case study was conducted in collaboration of Area Development Planning (ADP), a subsection of the City of Edmonton. The case study goal is to develop a TQMS for ADP.

**Decompose the environment**

The inputs to decompose the environment are a PES and conflict with dependencies. Conflicts with dependencies show the direction to follow for solving the conflicts. The PES is used to decompose the environment related to the conflicts.

Environment decomposition follows a structure operation, denoted by $\mathcal{S}$ (Zeng, 2001). Eq. (1) shows a general structure operation, where $U$ is a union of environment components and $\mathcal{X}$ represent interactions between environment components.

$$\mathcal{S} = E_1 \cup E_2 \cup (E_1 \mathcal{X} E_2) \cup (E_2 \mathcal{X} E_1)$$  (1)

After conducting environment analysis and conflict identification in the case study, the resulting PES and conflict dependencies and are shown in Fig. 2 and Fig. 3 respectively. Environment analysis followed a question asking approach using generic and domain specific questions. Generic questions clarify and extend the meaning of a design problem whereas domain specific ones identify implicit design information related to both the life cycle and natural, built and human environments of a product/service (Zeng, 2011). Once the question asking approach was completed, the output is the PES shown in Fig. 2. Conflict identification started with the PES as an input. The conflict identification activity followed three main steps: construct a performance network, identify active and reactive conflicts, and identify critical conflicts (Zeng, to appear 2014). The outputs of the activity are the conflict dependencies in Fig. 3, which indicate that $C_1$ has to be solved before $C_{12}$. Therefore, $C_{12}$ depends on the solution for $C_1$. The indices in the conflicts, i.e., $C_1$ and $C_{12}$, correspond to the parent interactions $I_3$ and to the pair $(I_1, I_2)$ respectively, which can be found in Fig. 2. $C_3$ is an active conflict while $C_{12}$ is a potential reactive conflict.

Based on the ROM representation in Fig. 2, $I_3$ is triggered by “instructions to comply with ISO 9001:2008 requirements”. Two instruction sources are needed for $I_3$: 1) an existing work handover procedure, and 2) ADP’s process, ADP’s resources and ADP’s supporting documents complying with ISO 9001:2008 requirements. The first source is already in place. The second source has to be further analyzed. Therefore, environment decomposition has to be applied for the second source by breaking it down into basic components. Basic components include each ISO 9001:2008 requirements and their sub-requirements, ADP’s processes including individual tasks, internal members’ knowledge, skills and technologies, ADP’s external stakeholders and existing supporting documents. There are several existing documents - such as Design and Construction Standards (The City of Edmonton, 2012), Drainage Services Stormwater Quality Strategy (The City of Edmonton, 2006), Erosion and Sedimentation Control Guideline (The City of Edmonton, 2005b), Erosion and Sedimentation Control Field Manual (The City of Edmonton, 2005a), Low Impact Development Best Management Practices Design Guide Edition 1.0 (The City of Edmonton, 2011), The City of Edmonton Drainage Services Master Plan 2004-2014 Implementation and Strategies (The City of Edmonton, 2004), Environmental Management System (Drainage Services Branch & Asset Management and Public Works Department, 2010), Terms of reference for the preparation and amendment of residential area structure plans (ASP) (The City of Edmonton, 2010a), Terms of reference for the preparation and amendment of residential neighbourhood structure plans (NSP) (The City of Edmonton, 2010b), and so forth - to support ADP’s technical work. However, they are called existing supporting hereafter in the paper for simplification purposes.

Environment decomposition is a technique used in EBD to understand the environment components related to interactions corresponding to a conflict. The understanding approach is systematic based on the analysis of all the possible relationships between basic environment components related to the interaction. This approach helps to uncover important interactions. The input to find the design knowledge is the identified critical conflict $C_3$. Relevant environment components were extracted from the ROM representation in Fig. 2. They are shown in Fig. 4, which also includes additional relationships between the interactions.
Fig. 2 PES in ROM representation after conducting environment analysis

Fig. 3 Algorithm for solution generation

Fig. 4 ROM representation containing only the environment components corresponding to $I_3$
Using the ROM representation in Fig. 4, the relevant environment components \( E_1 \) are defined as shown in Eq. (2). The environment components are simplified. For instance, \( E_1 \) is defined as ADP’s processes for simplification purposes; nonetheless, \( E_1 \) includes Stormwater management, Green infrastructure & environmental compliance and Drainage planning for land development processes with detailed member tasks. The same simplification applies for \( E_2, E_3 \) and \( E_4 \), where they refer to each requirement in ISO 9001:2008 and its sub-requirements, ADP’s members and its domain knowledge, skills and technologies; and the previous described supporting documents. The product “TQMS guideline” is not considered in the environment decomposition process. However, the acquired design knowledge through environment decomposition will help to identify the parts to be developed for the TQMS guideline to comply with ISO 9001:2008 requirements.

\[
E_1: \text{ISO 9001:2008 requirements} \\
E_2: \text{ADP’s stakeholders, knowledge, skills and technologies} \\
E_3: \text{ADP’s processes} \\
E_4: \text{ADP’s supporting documents}
\]

The environment decomposition for \( E_3 \) is carried out as shown in Eq. (3):

\[
\Theta E_3 = E_1 \cup E_2 \cup E_3 \cup E_4 \cup (E_1 \otimes E_2) \cup (E_2 \otimes E_1) \cup (E_2 \otimes E_3) \\
\cup (E_3 \otimes E_2) \cup (E_3 \otimes E_4) \cup (E_4 \otimes E_3) \\
\cup (E_4 \otimes E_2) \cup (E_4 \otimes E_1) \cup (E_2 \otimes E_3) \\
\cup (E_3 \otimes E_4) \cup (E_3 \otimes E_1) \cup (E_4 \otimes E_2) \\
\cup (E_4 \otimes E_3) \cup (E_4 \otimes E_1) \cup (E_1 \otimes E_3) \\
\cup (E_1 \otimes E_2) \cup (E_1 \otimes E_4) \\
\cup (E_2 \otimes E_1) \cup (E_2 \otimes E_3) \cup (E_3 \otimes E_2)
\]

where \((E_1 \otimes E_2), (E_2 \otimes E_1), (E_3 \otimes E_2)\) and \((E_4 \otimes E_3)\) mean updates of the environment components defined in Eq. (2), and \((E_1 \otimes E_2)\) refers to interactions between the environment components. The interactions between the environment components \( E_3, E_1 \) and \( E_4 \) are already contained in existing ADP workflows. For instance, \((E_2 \otimes E_3)\) are interactions between ADP’s stakeholders, knowledge, skills and technologies and ADP’s processes: Stormwater management, Green infrastructure & environmental compliance and Drainage planning for land development. The missing interactions to be evaluated are \((E_1 \otimes E_2), (E_2 \otimes E_1), (E_3 \otimes E_2), (E_3 \otimes E_4), (E_4 \otimes E_2), (E_4 \otimes E_1)\) and \((E_1 \otimes E_4)\). For example, \((E_1 \otimes E_2)\) is an interaction between ISO 9001:2008 requirements and ADP’s stakeholders, knowledge, skills and technologies. This interaction is given by a requirement from the ISO 9001:2008 to ADP’s stakeholders, knowledge, skills and technologies. \((E_2 \otimes E_1)\) will be a response from ADP’s stakeholders, knowledge, skills and technologies to comply with the ISO 9001:2008 requirement.

The environment decomposition in previous paragraph is the input for the next step in the proposed algorithm: find design knowledge.

Find design knowledge

Design knowledge is divided into synthesis and evaluation knowledge (Zeng & Gu, 1999). Synthesis knowledge maps from design requirements to design concepts. Design concepts include actions or physical structures to meet or exceed the design requirements (Zeng & Gu, 1999; Zeng & Gu, 2001). Evaluation knowledge assesses created design concepts against design requirements to conclude whether the design concepts satisfy the design requirements, they do not, or the assessment is undetermined (Zeng & Gu, 1999). Undetermined assessment arises when the created design concept is abstract, so it cannot be concluded whether it satisfies or it does not the design requirements. A more detailed design concept should be created to enable the designer to evaluate the undetermined assessments. Fig. 5 shows the relationships between design requirements, design knowledge, design concepts. Finding synthesis knowledge to satisfy the design requirements is the main goal of this step. The synthesis knowledge is instantiated in high level of abstraction candidate solutions. The rest of the process in Fig. 5, i.e., after candidate solutions, corresponds to the next section.

![Fig 5 Basic design process for synthesis and evaluation knowledge (Zeng & Gu, 2001)](image)

The case study in this paper followed the basic design process for synthesis and evaluation knowledge in Fig. 5. Generally speaking, this step in the case study evaluates Eq. (3). If something is not complying with the ISO 9001:2008 requirement, it is considered a gap. The step ends with synthesis knowledge, i.e., candidate solutions, at high level of abstraction to close the identified gaps.

The systematic gap evaluation was conducted between previously defined basic environment components. The evaluation was supported by existing ADP’s supporting documents, interviews with ADP group members and external stakeholders, and the design team knowledge about ISO 9001:2008 requirements. The inputs of the gap evaluations are the missing interactions to be evaluated: \((E_1 \otimes E_2), (E_2 \otimes E_1), (E_3 \otimes E_2), (E_3 \otimes E_4), (E_4 \otimes E_2), (E_4 \otimes E_1)\). Gaps, in this context, are defined...
as actions or resources missing to comply with the ISO 9001:2008 requirements. \((E_2 \otimes E_1)\), \((E_3 \otimes E_2)\) and \((E_4 \otimes E_3)\) are ISO 9001:2008 requirements to regulate ADP’s stakeholders, knowledge, skills and technologies, ADP’s processes and ADP’s supporting documents respectively. \((E_2 \otimes E_1)\), \((E_3 \otimes E_2)\) and \((E_4 \otimes E_3)\) are actions or resources from ADP’s stakeholders, knowledge, skills and technologies, ADP’s processes and ADP’s supporting documents to comply with the ISO 9001:2008 requirements.

After executing the systematic gap evaluation, the found synthesis knowledge is summarized in Table 1. Existing gaps between ISO 9001:2008 requirements and current condition of ADP are shown in the body of the table. Main ISO 9001:2008 requirements (ISO, 2008) are shown in the first column content in the left side of the table. These requirements regulate ADP’s stakeholders, knowledge, skills and technologies, ADP’s processes and ADP’s supporting documents. Therefore, this part of the table structure corresponds to \((E_2 \otimes E_1)\), \((E_3 \otimes E_2)\) and \((E_4 \otimes E_3)\). The second, third, and fourth columns in the table display existing gaps from ADP’s stakeholders, knowledge, skills and technologies, ADP’s processes and ADP’s supporting documents to comply with the ISO 9001:2008 requirements. Hence, this part of the table structure corresponds to \((E_2 \otimes E_1)\), \((E_3 \otimes E_2)\) and \((E_4 \otimes E_3)\). The fifth column provides synthesis knowledge a high level of abstraction to comply with ISO 9001:2008 requirements.

<table>
<thead>
<tr>
<th>ISO 9001:2008 requirements</th>
<th>ADP’s processes</th>
<th>ADP’s stakeholders</th>
<th>ADP’s supporting documents</th>
<th>Synthesis knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. General requirements</td>
<td>ADP is being</td>
<td>Stakeholders’ tasks are being redistributed now; therefore, tasks need to be balance accordingly.</td>
<td>KPIs* to monitor processes, stakeholders satisfaction, and efficiency and effectiveness of ADP have to be documented.</td>
<td>1. A plan has to be developed to deal with the current restructuring phase of ADP.</td>
</tr>
<tr>
<td>2. Documentation requirements</td>
<td>restructuring now; therefore,</td>
<td></td>
<td>Management reviews, internal audits, nonconforming results analysis and improvement actions procedures have to be documented and recorded every time they are used. A TQMS has to be created.</td>
<td></td>
</tr>
<tr>
<td>3. Management responsibility</td>
<td>existing workflows and documents</td>
<td>Stakeholders need to be provided with required resources due to new tasks balance.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Resource management</td>
<td>have being affected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Service realization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Measurement, analysis and improvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* KPIs: Key Performance Indicators

**Generate solutions**

This is the last step of the proposed algorithm. The inputs to this step are the candidate solutions. The step generates design concepts at a desired level of abstraction. After that, the generated concepts have to be evaluated against the design requirements. The evaluation decisions given in the “find design knowledge” section have to be followed.

The goal of this step for the case study is to develop the described solutions in the synthesis knowledge column in Table 1. This step was iterative until meeting or exceeding the design requirements. Each of the environment components in Eq. (2) were broken down in this step. The environment components were also divided into sub-components, for instance \(E_3\) includes Stormwater management, Green infrastructure & environmental compliance and Drainage planning for land development processes with detailed member tasks. The solutions were generated at the level of sub-components. The solutions were composed and finally evaluated against the ISO 9001:2008 requirements until satisfaction was achieved. While generating solutions, an effort was made to avoid the potential reactive conflict \(C_{12}\) to happen. This conflict
is concerning to a potential overlap of resources while implementing and improving continuously the developed TQMS.

A summary of the generated solutions corresponding to each of the synthesis knowledge in Table 1 includes:

1. Procedures for redefining ADP’s tasks distribution. The procedures comprise redefinition of work scope, redefinition of boundaries in overlapping activities between stakeholders (internal and externals), redefinition of expected deliverables, clarification of responsibilities, workload balance approach, follow up actions, provision of training and other resources, management involvement, tasks to create/update affected documents and records, and process to communicate changes.

2. A document defining KPIs, management reviews and internal audits. Furthermore, the document is complemented with a framework to measure the KPIs while conducting management reviews and internal audits. The document also contains templates to record each management review and internal audit.

3. A procedure document to analyze the results of the management reviews and internal audits. This document also includes a framework to record each analysis. A procedure document to implement corrective and preventive actions when nonconforming analysis results are obtained. Moreover, this procedure document contains steps for following up actions, templates for recording the analysis results, improvement and follow up actions; and tasks to update affected documents and communicate to concerning stakeholders.

4. A TQMS outline was created based on the ISO 9001:2008 requirements. The outline has sections for documenting the quality policy, objectives, scope of the TQMS, exclusions, ADP’s processes and interactions at all levels in the organization, work handover procedure, existing supporting documents, management reviews, internal audits, nonconforming result analysis and improvement actions, procedures for controlling documents and records, record templates, and records. The outline provides directions to be filled out and states some exclusions to the requirements. The exclusions arose because some requirements cannot be applied due to the nature of ADP. The exclusion are limited to requirements within Clause 7 in ISO 9001:2008 requirements (ISO, 2008) and they do not affect the ADP’s ability, or responsibility, to provide services that meet customer and applicable statutory and regulatory requirements.

CONCLUSION

This paper presented an algorithm for solution generation in conceptual design using EBD. The inputs to the algorithm are a PES and conflicts with dependencies created in previous EBD activities: environment analysis and conflict identification. The algorithm has three main steps: decompose the environment, find design knowledge, and generate solutions. The solutions are evaluated to conclude whether the generated concept meet the design requirements, they do not, or it is undetermined. In the case that the generated concepts meets the requirements, the algorithm stops. In the other cases, a designer has to go back to the second step, find design knowledge, and continues the algorithm until the design requirements are satisfied.

A case study with the collaboration of ADP, a subsection of the City of Edmonton, was presented using the proposed algorithm. The goal after applying the algorithm was to develop a TQMS for ADP. The algorithm breaks down the main environment components of the TQMS and leads to identify missing tasks, procedures, documents and records for ADP to comply with ISO 9001:2008 requirements. Solutions were developed for the identified tasks, procedures, documents and records. The generated solutions were evaluated against the ISO 9001:2008 requirements iteratively and the algorithm ends until design requirements were met. A TQMS outline was created for ADP including existing documentation and the one generated. Both documentations are needed for the TQMS for ADP to comply with ISO 9001:2008 requirements.

REFERENCES


ABSTRACT

The paper presents persuasive nature of Social Network Sites (SNS) using Facebook as a case study for augmenting Behavior Change Support Systems (BCCSSs). The prime objective of this paper is to gain richer understanding of SNS usage, acceptance and users’ actual behaviors. We have attempted to understand selected constructs from available literature. We propose that understanding phenomenal success of Facebook could provide valuable information for developing BCCSSs by bringing in the components of entertainment and sociability. We carried out qualitative study to understand SNS users’ online behaviors with an explicit focus on enjoyment, entertainment, desire to be connected, reciprocation, information quality, need for admiration and influence of obtrusiveness. We performed quantitative analyses with special focus on Value-based constructs because values are motivational constructs could help predict users’ behaviors. Results indicate that enjoyment, entertainment, reciprocation and admiration have the strongest influence on individuals’ behaviors.

KEYWORDS

SNS = Social network sites
FB = Facebook
SI = Social learning
SF = Social facilitation
BCSS = Behavior change support systems

INTRODUCTION

Humans are predominantly driven by intrinsic or extrinsic motivations. These motivations could be a blend of varying subtleties ranging from living a healthy life, financial stability, having social connections and so forth. Further, as humans we have different attitudes and beliefs that gradually develop our characters. Intriguingly, not all of us act in accordance with our beliefs (Langrial, 2012). For instance, most of the people approve physical exercise still only a few follow regular exercise routine. The same goes for junk food, despite its detrimental consequences, high majorities of people consume junk food leading to physical disorders such as obesity. Humans have a natural tendency to behave in contradiction to their beliefs. In other words, there is a gap between what an individual believes and what she essentially does. This varying behavior gives rise to a distinctive state of cognitive unease (Festinger, 1962). It was this particular state of cognitive dissonance that predetermined a manifesto for persuasive technologies (Fogg, 2002) developed to support people in overcoming differences between their beliefs and actual behaviors.

Fogg (2002) wrote his formative book “Persuasive Technology: Using Computers to Change What we Think and Do”. Developing technologies that aim to change peoples’ behaviors is a demanding task. Studying human behavior and developing behavior change interventions exposes several gaps in terms of descriptions, development processes and implementations. Lately, digital interventions with persuasive features have received growing attention. Largely, there are three categories of such interventions with subtle however significant differences. The categories include Digital Interventions, Persuasive Systems and Behavior Change Support Systems (Langrial, 2012; Oinas-Kukkonen 2010). Digital interventions have largely been focused on preventive health, persuasive systems emerged as interactive IT artifacts with a specific focus on behavior.
change using functionalities comprising reminders, rewards, social learning etc. Persuasive systems have gained noteworthy achievements over the years. However, researchers have rather failed to demonstrate actual and enduring change in users' behaviors or attitudes (Oinas-Kukkonen, 2010). We propose that for developing effective behavior change interventions, there is a need for superior understanding of socio-technological features in the process of abstraction, development and implementation. Building upon the existing research, Oinas-Kukkonen (2010) conjectured BCSSs as a general platform incorporated with carefully selected persuasive techniques and software features.

Broadly persuasive systems are interactive IT artifacts (Langrial, 2012). Emergence and pervasive acceptance of social web and ubiquitous computing devices have created opportunities for shaping peoples’ behaviors in general health, psychological well-being, sociability and economic contexts (Oinas-Kukkonen and Harjumaa, 2008). Several researchers have defined persuasive systems nevertheless Fogg’s (2002) definition stands out amongst all where he describes such technologies that are designed to bring a desirable change in peoples’ behaviors and/or attitudes. Lately, persuasive systems have been developed to tackle a range of problem domains such as promoting physical activity (Toscos et al. 2006), smoking cessation (Walters et al. 2006), promoting sensible usage of energy (Midden and Ham, 2009), healthy aging (Intille, 2004) and managing depression (Langrial et al. 2013).

Oinas-Kukkonen (2010) has added to the research field of behavior change technologies by introducing the theoretical framework of BCSSs. His work takes contemporary research a step further. He emphasis on developing interactive IT artifacts that are augmented with software features for extended user-system interaction (cf: Langrial, 212). It is interesting to note that the concept of BCSSs develops from Fogg’s (2002) original work on persuasive technologies. Nonetheless, it brings in new research directions including comprehensive systems analysis and implementation of software features (tunneling, reduction, personalization, reminders, rewards, social influence etc.) The prime objective of a BCSS is to engage users into a continuing interaction with the system that would lead to a long lasting change in behaviors and/or attitudes.

We suggest that designers of behavior change technologies have thus far failed to develop successful IT artifacts that could establish permanent change in peoples’ behaviors. Designing effective behavior change technologies is a difficult task for several reasons consequently users’ requirements and expectations need to be judiciously understood while conceptualizing such interventions (Langrial, 2012). Available literature reveals that there is a lack in understanding models that are aimed at changing human behavior (Or et al. 2011). According to Benbasat (2010), interventions that are reviewed through the lens of Human Computer Interaction (HCI) are neutral in nature, however it is a subjective argument (Langrial, 2012). On the contrary, we suggest that enhanced human computer interactivity is fundamental for developing successful behavior change technologies. Some of the common most techniques used to improve human-system interaction are prompts, reminders or notifications (Walji and Zhang, 2008), positive feedback (Arroyo et al. 2005), virtual rewards (Harjumaa et al. 2009) and social facilitation (Gasser et al. 2006). The next section discusses SNSs.

The main objective of this paper was to examine and report subtleties of SNSs and draw design implications for abstracting and developing successful health-related Behavior Change Support Systems.

SOCIAL NETWORKING SITES

Social networking sites including Facebook, LinkedIn, and Twitter have gained mounting popularity over recent years. According to a report, there were 1,110 million active Facebook users alone by March 31, 2013. Several studies have been carried out to assess and cognize reasons behind the recognition and acceptance of Facebook (Lin and Lu, 2011; Pinho and Soares, 2011). It is evident from existing literature that majority of the studies performed on Facebook usage focus on theories such as the Technology Acceptance Model (TAM) (Davis et al. 1989) or constructs relating to social connectivity (Cheung et al. 2011). Livari (2012) has highlighted a recent addition to the social media literature where he has outlined a novel construct of Perceived Sociability of Use (PSOU). In short, Livari (2012) suggests that PSOU explains users’ desire to create and sustain social connections through Facebook. We propose that in addition to the studied and documented theoretical explanations relating to the success of SNSs, there are under-studied details that need to be adopted.

These subtleties and persuasive features of SNSs could provide valuable information to develop BCSSs by integrating elements and features that have made Facebook such a phenomenal success. Conventionally, usage of interactive technologies has been well-studied research area (Yousafzai et al. 2007). Present research provides evidence that some of the key factors behind the success of social media are related to enjoyability (Davis et al. 1992). Researchers have pointed out other factors such as playfulness, pleasure, fun and flow as contributory factors for acceptance and popularity of SNSs. Instead, we are interested in analyzing the design implications of social networking platforms for developing BCSSs. It is also suggested that massive data gathered through such

---

A BCSS is an object of study within the field of Persuasive Technology (Oinas-Kukkonen, 2012). It is a comprehensive socio-technical platform that facilitates the process of behavior / attitude change through long-drawn-out support for users through combination of persuasive software features and socio-psychological theories (Oinas-Kukkonen, 2012). We, therefore, propose that sociability is one of the key features for developing effective BCSSs. According to Oinas-Kukkonen (2012), emerging technologies particularly social web have created openings for developing information systems that facilitate the process of designing, retrieving and sharing information in novel ways. He outlines that social web will continue to create opportunities for developing improved and effective information systems to influence users (Oinas-Kukkonen, 2012). According to Fogg and Iizawa (2008), SNSs are among the most popular information platforms. They further add that such sites are most influential in nature. Persuading individuals is key to the phenomenal success of SNSs (Ferebee and Davis, 2009). Keeping this approach in mind, it becomes vital to investigate the core socio-technical subtleties of Facebook in persuading people especially in terms of enjoyment, desire, reciprocity, trustworthiness and (social) admiration. Further we advocate that efforts should be made to ascertain if and why exclusive dynamics influence Facebook users’ online behaviors. Cialdini’s (1984) work also provides comprehensive persuasion principles that have previously helped researchers better understand the persuasion process.

Online social networking sites create an opportunity for researchers in the area of behavior change information systems. Enabling users to create and sustain personal web presence is among several distinctive features of such sites (Ellison et al. 2007). Typically, users of Facebook use it for a variety of reasons including sharing information, staying in touch with friends and family, viewing published content and communicate through instant messages (Ellison et al. 2007; Maia et al. 2008). In other words, Facebook as a SNS offers an opportunity to connect with others, share information and converse on a regular basis (Gangadharbatla, 2008). It is argued that Facebook gives its users independence and openings to relish if not undergo a unique user experience (Hart et al. 2008).

Most of the scientific research on Facebook and other SNSs has largely been focused on issues related with confidentiality (Dwyer et al. 2007). Indisputably, it is a delicate research area nevertheless the fact is that Facebook is gaining popularity on an ever-increasing basis. According to Alexa’s, Facebook was among the top 7 most visited web sites. Several studies have attempted to assess the reasons for SNSs’ approval (Gangadharbatla, 2008). Among the reported reasons, need for being connected, need for knowledge, socializing and perceptions of acknowledged behaviors have been summarized (Ridings and Gefen, 2004). However, to the best of our knowledge, no study has attempted to analyze the influence of holistic values such as desire, reciprocity, trustworthiness, admiration, and frustration on the usage of Facebook leading to a potential pattern in their online behaviors. A variety of psychological theories and the Persuasive Systems Design model (Oinas-Kukkonen and Harjumaa, 2009) have been used to have a better understanding of the aforementioned examples.

Social Networking Sites and Entertainment

Facebook is one of the most popular SNS and was developed to support users establish online connections. According to Tosun (2012), different features of Facebook serve users’ personal and social desires, for example, staying in touch, posting photos, making friends, socializing and getting entertained. Several studies have been carried out to find out reasons for Facebook usage. Among other identified motives, entertainment has emerged as one of the key reasons according to a study conducted by Boyd (2008). Tosun (2008) concludes that entertainment is the second key motivators for Facebook users following social connectivity. On similar lines, Shi et al. (2010) have outlined entertainment as a primary motive for Facebook usage that pertains to enjoying free time, playing games and having fun. They further propose that Facebook provides an opportunity to entertain oneself and pass time when feeling bored. Facebook offers plentiful tools for entertainment, for example growing and sharing gifts, IQ tests and other online games embedded within the SNS. Research has shown that Facebook has proven to be highly engaging. Entertainment applications do play a central role in the phenomenal success of Facebook. For example, social games that are available have had a huge success and received tremendous popularity. It is estimated that games such as Mafia Wars, Farm Ville and City Ville have more than 60 million daily active users (Adolph, 2011). Having said all that, one of the major challenging issues in incorporating entertainment-based applications for BCSSs include privacy that calls for further research.

---

2 www.alexa.com Web information Service (traffic ratings) in the year 2008
Position of Socio-psychological Theories

In depth understanding of motivational theories is indisputably fundamental for envisioning health Behavior Change Support Systems. Extensive work has been done in the area of persuasive technologies and Behavior Change Support Systems using motivational theories such as the Goal-setting theory (Locke and Latham, 2002), positive and negative reinforcement (Dey and Abowd, 2000), the Social Identity theory (Turner and Reynolds, 2011) and the Cognitive Dissonance theory (Festinger, 1957). Use of socio-psychological theories in developing health Behavior Change Support Systems is a promising sign and as the research field of behavior change interventions is expanding, we notice studies that are more often theory driven. Theory-driven practices for abstracting and developing behavior change interventions has been strongly supported by scholars, for instance, Consolvo et al. (2009) who have employed the Transtheoretical Model and Presentation of Self in one of their projects. We propose that health Behavior Change Support Systems researchers should exploit the theory of Self-efficacy (Bandura, 1977). We firmly believe that an effective Behavior Change Support System should to be incorporated with persuasive techniques and distinctive software features that could boost users’ self-confidence, which is essential to overcome undesirable behaviors.

Aforementioned summary provides evidence about the materialization and gradual expansion of digital interventions for behavior change to Behavior Change Support Systems. These innovative technologies are evolving at a swift pace yet weaknesses in prevalent research are noticeable as outlined by Oinas-Kukkonen (2012). Scholars have advocated critical and rigorous evaluation of behavior change technologies, for example, Consolvo and Walker (2003) argue that if evaluations were to be carried out on incessant basis, there will be greater possibilities to improve such technologies by identifying users’ necessities and expectations. Evaluation of behavior change interventions has received a great deal of attention both from IS and HCI researchers where different models and frameworks have been presented for the purpose. For example, Consolvo and Walker (2003) have promoted the use of Experience Sampling Model (ESM), Scholtz and Consolvo (2004) have conceptualized a framework for evaluating ubiquitous applications, Fogg (2009) has proposed Behavior Model (FMB), Wiefe et al. (2012) have come up with the 3D-RAB model and Oinas-Kukkonen and Harjumaa (2009) developed the Persuasive Systems Design Model (PSDM). Although existing research in the field of behavior change technologies is enthusing and valuable contributions have been made. Yet we argue that the prime focus has been on development and more precisely evaluation of such systems. We, therefore, suggest that there is a need for researchers and scholars to think out of the box. We strongly believe that research on social networking sites (SNS) can provide valuable knowledge that will assist behavior change technologies’ researchers better understand human behavior and attitudes. We propose that research in the field of behavior change interventions has thus far failed to exploit subtleties and varying dynamics of social network sites.

Literature Background

Oinas-Kukkonen (2012) promotes categorical application of socio-psychological theories in designing and developing BCSSs. He adds that research on users’ attitudes and/or behaviors has more or less been a part of computing sciences. The significance of meticulous conception of socio-psychological theories is that they provide valuable knowledge that helps researchers better understand users’ behaviors and/or attitudes. A deeper understanding and pertinent application in analyzing outcomes will help researchers to design and evaluate better persuasive user experiences (Oinas-Kukkonen and Oinas-Kukkonen, 2013). According to Gibb (1972), several persuasion concepts could be applied to determine the success and influence of Facebook. For example, trust, openness, realization and interdependence. Gibb (1972) calls it as the TORI theory. Although Cialdini (1984) suggests that persuasion takes place for the reasons such as sense of commitment, expectancies and imitation of others. Gibb (1972), argues that persuasion takes place through confidence and directness. Not on the same lines, but interestingly Oinas-Kukkonen and Harjumaa (2009) add that psychological theories alone might not be enough for successful persuasive attempts. They argue that it is a combination of pragmatically selected socio-psychological theories and particular software features that are the key constituents of a successful persuasive system and BCSSs.

Consolvo et al. (2009) have proposed theory driven design strategies for technologies that support behavior change. Their proposed design strategies include: data abstraction, presenting information in an unobtrusive manner, empowering users to share private information, aesthetics of the user interface, positive reinforcement, user empowerment, credibility, provision of historical data and comprehensive presentation of data relating to target behavior change. It is evident that some of the strategies proposed are also being addressed in this paper. However, we have adopted a different approach by studying users’ behaviors on Facebook in order to formulate hypotheses and draw design implications based on statistical findings. We propose that understanding Value-based constructs from social networks’ outlook could improve design and implementation of successful BCSSs. Primarily drawing the constructs from social and psychological theories, we outline six potential design constructs i.e. enjoyment, desire fulfillment, reciprocation, quality of information, addressing human need to be
admired and creating non-obtrusive systems. Table 1 represents the constructs and related theories. A variety of research has been carried out to evaluate persuasion in social networks and several designs have been reported. These patterns outline persuasion dynamics in Facebook (Ferebee and Davis, 2009). Below is a brief account of the examined constructs:

HYPOTHESES

Keeping in mind the holistic values, we examined six constructs and developed the hypotheses accordingly.

Enjoyment

Davis (1982) claims that enjoyment is about feeling pleasure where pleasure is classified with simultaneous joy, in terms of belief, desire and feelings. Based on the Davis’s (1982) theory, we propose that Facebook users enjoy and experience pleasure. Not everyone using Facebook might enjoy her experience at the same level. In fact, the degree of enjoyment by being socially interactive through Facebook depends on how much one relishes. Nonetheless, enjoyment is a vital feature of Facebook experience. We developed several questions relating to enjoyment, for example, “I use Facebook because it is pure fun”. Keeping in mind Davis’s theory, we hypothesized:

H1. Enjoyment has a positive influence on Facebook users actual behaviors.

Desire

People have a natural urge to belong, be valued and accepted (Baumeister and Leary, 1995). We term this phenomenon as desire to belong and be loved. It is a fundamental characteristic of humans that they make an effort to form and sustain social contacts. Facebook, just like other social networking platforms enables its users to satisfy their desire to stay connected with their family and friends. For example, among a set of question relating to Desire to belong, one question posed was, “I use Facebook because it keeps me connected with my family. Because humans have an inherent desire to belong and be treasured by others, we hypothesized:

H2: The desire to seek admiration has a positive influence on Facebook users actual behaviors.

Reciprocity

According to Cialdini (1984), people have a natural inclination to return favors. He supplements his claim by referring to the generality of free marketing sample. Based on this theory, we propose that reciprocity plays a key role in users’ online behavior on Facebook. For example, when a user likes a post of her Facebook connection, more than often, “The later reciprocates by liking a post of the earlier”. Based on the above mentioned, we hypothesized that:

H3: Facebook users feel obligated to reciprocate positive comments.

Trustworthiness

According to Kelton et al. (2008), trustworthiness is a critically important topic in the area of Internet research. Trustworthiness can be assessed in several forms such as security, privacy and reliability. We aimed to examine how Facebook users perceived information shared by others. For example, “The information on Facebook is reliable”. Assuming that trustworthy information has a vital role in Facebook usage, we hypothesized that:

H4: Users are positively influenced by the information shared in Facebook.

Admiration

This is about efforts made by Facebook users to be appreciated. According to Goffman (1959) it is all about self-presentation. Goffman (1959) uses the metaphor of a theatric play. Here the performer is the Facebook user who self-presents her. The performer interacts with the audience at the front end with an effort to portray a character that will be admired by others. While offline, the person is at much ease as there is time to loosen up some of the “self-imposed restrictions from the front area”, according to Goffman (1959). Nevertheless, it is humans’ inherent motivation to be liked and admired and we argue that Facebook users are not an exception to this. For example, one of the questions that we asked was, “Being a Facebook user, I do not care of what others might think about me”. It was our assumption that Facebook users care about their image. Therefore, their online behavior is positively influenced by the need to be admired. We hypothesized that:

H5. Facebook users actual behavior is significantly influenced because they want to be admired by others.

Task Obtrusiveness

According to the PSD Model (Oinas-Kukkonen and Harjumaa, 2009), persuasive system should aim at being unobtrusive (cf. Postulate 6). The idea is that the system should not intrude or frustrate users while they are engaged in their primary tasks (Drozd et al. 2012). This way, the system would meet users’ positive expectations. We propose that nudging or reminding at inappropriate times might end up in undesired consequences. For example, receiving a notification on a smartphone while in a meeting might not be a suitable moment for the user. Keeping in mind the postulate, we asked a set of questions. For instance, “While I am working and receive
feedback notification on my mobile phone, PC or laptop, such notifications annoy me”. Keeping in mind Postulate 6 of the PSD Model [9], we hypothesized that:

H6: Facebook users would feel frustrated by obtrusive nature of Facebook notifications.

RESEARCH METHODOLOGY

This section presents recruitment and data collection, basic demographics about the participants and the measurement model. Our target group was university students and staff. We decided to conduct a qualitative analysis by inviting people to voluntarily take part in the study. Participation invitations were sent via university’s internal email as well as by placing printed posters on campus. In all, we received 108 paper-based responses out of which three were excluded from the study because they were incomplete. Hence, the total number of reliable responses was 105. There were no monetary inducements offered to the participants. The minimum sample size can be calculated by applying the heuristic ten times per the largest number of independent constructs influencing the dependent construct (Chin et al. 1996). By doing so, the minimum sample size of our study would be 10 times 6 (i.e. 60). Several responses were discarded because the respondents did not return the questionnaires in sealed envelopes. Basic demographics of the reliable respondents are presented below:

Profession. We had 8 categories of profession/occupation with the eighth being an open ended question i.e. “Other”. There were 6 (5.7%) engineers, 5 (4.8%) teachers, 87 (82.9%) students, 3 (2.9%) researchers and 4 (3.8%) did not specify their occupation.

Computer Expertise. We had 6 categories for computer skills i.e. competent, very good, good, average, fairly weak and weak. In response, 42 (40%) reported to be competent, 34 (32.4%) very good, 26 (24.8%) good, 2 (1.9%) average and 1 (1%) reported having fairly weak computer expertise.

Frequency of Facebook Usage. We had 4 categories for this question i.e. rarely, seldom, often and regularly. In response, 12 (11.4%) reported to rarely use Facebook, 8 (7.6%) reported that they seldom use it, 24 (22.9%) used Facebook often and 60 (57.1%) stated that were regular Facebook users.

THE MEASUREMENT MODEL

The data analysis was conducted using Smart Partial Least Squares (PLS) software, which is used for predictive applications. Previously, Drozd et al. (2012) have also employed the structural model in their research. It allows modeling of non-normal variables and small-to-medium samples (Grewal et al. 2004). In the analysis, we used a two-step approach as recommended by Anderson and Gerbing (1988). Convergent validity was ensured by assessing the factor loadings and by calculating variances extracted. As Table 3 exhibits, all the model items loaded well, exceeding 0.50 (Anderson and Gerbing, 1988), except in the case of PSD (PSD1), which was dropped. Calculating Cronbach’s alpha assessed internal consistency reliability among the items. Table 3 shows that the suggested value of 0.60 was exceeded for all constructs, except in the case of behavior and desire, which was slightly under the threshold suggested by (Anderson and Gerbing, 1988). Nunnally (1970) suggests that reliability ranges of 0.5 or 0.6 can be regarded as sufficient in exploratory studies. The variance extracted of all the constructs exceeded 0.5 (Hair and Anderson, 1998). The composite reliability of all the constructs exceeded the suggested value of 0.7 (Nunnally, 1978).

Computing the correlations between all pairs of the constructs assessed discriminant validity. All the correlations were below the threshold value of 0.90 (Anderson and Gerbing, 1988). The square root of the variance extracted was greater than the correlations of the constructs (see Table 4). Cross-loadings of the items on their assigned latent variables were larger than any other loading. Hence, the reliability and validity of the constructs in the model are acceptable.

STRUCTURAL MODEL AND RESULTS

The results of this study are exhibited in Figure 1, which shows estimated path coefficients and the significance of the path, which is indicated with asterisks (*). Tests of significance were performed using the Bootstrap procedure. Beta coefficients show that Enjoyment (β = -0.525), PSD (β = 0.160), reciprocity (β = 0.233) and Admiration (β = -0.223) have a strong significant influence on actual behavior therefore H1, H4, H6 and H7 are supported. Desire (β = 0.108) and Trustworthy (β = 0.028) have an insignificant influence on actual behavior. Enjoyment (β = -0.189) has an insignificant impact on mediating variable Desire therefore H2, H3 and H5 are not supported. Overall, the research model accounts for 43.1% (\(R^2 = 0.431\)) of the variance in actual behavior.

RESULTS AND DISCUSSIONS

First, we hypothesized, enjoyment, desire to stay connected, reciprocation, trust in information quality, need to be admired and obtrusive nature of SNSs’ significantly influences Facebook usage and users’ actual online behaviors. In terms of significance, Beta coefficients show that Enjoyment, Reciprocation and Admiration have the strongest influence on actual behaviors. This suggests that Facebook users’ actual
online behaviors are mostly influenced by need for enjoyment, reciprocation and the need for admiration. Surprisingly, the effect of desire and trust on the users’ actual online behavior was found to be relatively insignificant. One possible reason could be that an overwhelmingly majority of the participants were students who are frequent users of Facebook. For most of them enjoyment, reciprocation and desire to be admired might be important. Second, we confirm that enjoyment, need for admiration and reciprocating positive comments have a positive influence on Facebook users’ actual behaviors. The effect of trust and obtrusiveness were observed to be less significant. The finding relating to trust is in contrast with mostly reported findings. We hypothesized that high degree of information quality would lead to users’ trustfulness. It could be argued that youthful Facebook users are less concerned about security, privacy and reliability.

FUTURE RESEARCH

Presented work offers evidence that underlying subtleties of Social Networking Sites could potentially enhance the effectiveness of Behavior Change Support Systems. However, specific focus on the utilization of socialiability and entertainment in BCSSs is evidently understudied with a few exceptions. For example, a recent study by Rao (2013) outlines preliminary evidence that gamified approach for behavior change motivates individuals with depressive symptoms to perform outdoor activities and develop social connections. Paredes et al. (2013) worked on design principles for abstracting gamification for behavior change. Behavior change is a process that involves a prolonged human-system interaction. We propose that SNSs could provide a better understanding about engaging users for an extended period of time. One challenge that cannot be overlooked is individuals’ sensitivity towards sharing personal issues (e.g. health) over online social networks. To overcome this critical issue, we suggest that designers, researchers and practitioners need to utilize identified constructs from SNSs in abstracting and designing novel BCSSs that would empower the users to socialize and connect with fellow users if they wish.

Further research is also called for in the field of gamification and its implications on developing BCSSs. There is evidently a role for entertainment, joy, social connectivity and gamification for BCSSs. We propose tightly controlled trial studies of such systems that mimic SNSs. This would help researchers better understand and employ vital constructs from online communities and SNSs for mass dissemination as pointed out by Bennet and Glasgow (2009).

CONCLUSIVE REMARKS

Using Facebook as a case study, we have presented statistical findings from a qualitative data with an aim to understand characteristics and software features that influence peoples’ behaviors both on and offline. Based on the statistical findings, we propose that social network platforms such as Facebook could provide insights for behavior change researchers in order to design and develop effective BCSSs. It is rather disappointing to note that social influence and social facilitation as software functionalities are the least incorporated features in prevailing BCSSs especially for general well-being (Langrial et al. 2012). For developing successful BCSSs, we propose following design strategies:

Engaging user experience. Develop interventions that engage users through feeling of joy, delight and fun. This is in accordance with Davis (1982) and the Technology Acceptance Model (TAM) (Venkatesh and Bala, 2008). A BCSS should be designed and implemented in a way that users not only find it easy to use but equally importantly, they should have a feeling of joy and fun.

Connect with family and friends. Humans have an inherent desire to stay in touch especially with their loved ones. Software features that facilitate users to stay connected with loved ones would motivate them to continue using BCSSs thereby increasing the chance of behavior change. According to Stibe et al. (2013), software functionalities that facilitate social learning, normative influence and social comparison play a significant role in influencing users behaviors.

 Reliable information. The quality of information has significant influence on users’ trust in BCSSs. It has already been argued that trustworthiness is a critical area of research Kelton et al. (2008). Reliable information will bring in the elements of trustworthiness and reliability, for instance, through the use of expertise, 3rd party endorsements and verifiability (Oinas-Kukkonen and Harjumaa, 2009).

Public admiration. Incorporate software features that facilitate public interaction. Designers of BCSSs need to understand that human beings (users) have an inherent need to be admired (Goffman, 1959). Generally, people act in accordance with social norms so that they are accepted and admired by others. For illustration, Recognition as a persuasive software feature (Oinas-Kukkonen and Harjumaa, 2009) allows users to share information such as success stories. Therefore, admiration can play a vital role in changing behaviors. As argued by Oinas-Kukkonen and Harjumaa (2009), facilitating public recognition can potentially increase the chance of users’ adopting desirable behaviors. In addition, Stibe and Oinas-Kukkonen (2012) report that there is evidence of persuasiveness of recognition in designing Behavior Change Support Systems.
**Unobtrusiveness.** A BCSS should be designed in such a way that vigilant attention is paid to potential impact of obtrusive reminders, nudges and notifications. This brings use and user context into discussion. Reminders and notifications should be sent out at opportune moments. The system should not disrupt users while they are engaged in primary tasks (Oinas-Kukkonen and Harjumaa, 2009). This proposition is also supported by (Consolvo et al. 2009).

Limitations of the study include relatively narrow sample size however it meets the statistical criteria as mentioned earlier. Another limitation was a smaller number of female participants making it hard to perform detailed analyses comparing results based on gender. Nevertheless, the model and design implications can be further studied and applied in several contexts. Our work contributes to the existing literature on behavior change field because to the best of our knowledge, this is the first exploratory study that examines Facebook users’ behaviors and draws design implications for developing effective BCSSs.

Analysing Facebook users’ online behavior is manifestly relevant in improving the effectiveness of BCSSs. Recent research is at general level and there is a pressing need to think outside of the box in terms of research approach (Oinas-Kukkonen, 2012). This paper reports specific dynamics of Facebook usage and Facebook users’ behaviors in terms of enjoyment, desire, reciprocity, trustworthiness, admiration and obtrusiveness. A theory-based model was drawn using PLS-ESM. The results indicate BCSSs’ researchers and designers could gain valuable knowledge from SNSs such as Facebook. Another contribution of this paper includes the PLS-ESM-based drawn research mode that would supplement prevalent knowledge base in the emerging research area of BCSSs.

**Acknowledgements.** The author would like to thank Prof. Seppo Pahnila for his guidance in developing the structural model and Anton Oinas-Kukkonen for his help in developing the initial data set. Lastly, special thanks to SalWe research Program for Mind and Body (Tekes – the Finnish Funding Agency for Technology and Innovation grant 1104/10).

**REFERENCES**


Table 1 Proposed constructs and relevant theories

<table>
<thead>
<tr>
<th>Construct/Theme</th>
<th>Related Theory/Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyment</td>
<td>A Casual Theory of Enjoyment (Davis, 1982)</td>
</tr>
<tr>
<td>Desire</td>
<td>Need to Belong (Baumeister and Leary, 1995)</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>The Psychology of Persuasion (Cialdini, 1984)</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>Trust in Digital Information (Kelton et al. 2008)</td>
</tr>
<tr>
<td>Admiration</td>
<td>The Presentation of Self in Everyday Life (Goffman, 1959)</td>
</tr>
<tr>
<td>Task Obtrusiveness</td>
<td>The PSD Model – Postulate 6 (Oinas-Kukkonen and Harjumaa, 2009)</td>
</tr>
</tbody>
</table>

Table 2 Respondents’ profiles

<table>
<thead>
<tr>
<th>Measure</th>
<th>Items</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>61</td>
<td>58.1%</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>44</td>
<td>41.9%</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>Null</td>
<td>Null</td>
</tr>
<tr>
<td>Age</td>
<td>&lt;22</td>
<td>4</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>22-31</td>
<td>87</td>
<td>82.9%</td>
</tr>
<tr>
<td></td>
<td>32-41</td>
<td>10</td>
<td>9.5%</td>
</tr>
<tr>
<td></td>
<td>&gt;42</td>
<td>4</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>Null</td>
<td>Null</td>
</tr>
</tbody>
</table>

Table 3 Validity, internal consistency and reliability

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items</th>
<th>Factor Loading</th>
<th>Average Variance Extracted (Ave)</th>
<th>Cronbach’s alpha</th>
<th>Composite reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual behavior</td>
<td>Beh1</td>
<td>0.800</td>
<td>0.675</td>
<td>0.521</td>
<td>0.806</td>
</tr>
<tr>
<td></td>
<td>Beh2</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSD</td>
<td>PSD1</td>
<td>Dropped</td>
<td>0.797</td>
<td>0.749</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>PSD2</td>
<td>0.836</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PSD3</td>
<td>0.921</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>Enjoy1</td>
<td>0.834</td>
<td>0.702</td>
<td>0.787</td>
<td>0.875</td>
</tr>
<tr>
<td></td>
<td>Enjoy2</td>
<td>0.897</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enjoy3</td>
<td>0.777</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desire</td>
<td>Desire1</td>
<td>0.788</td>
<td>0.541</td>
<td>0.541</td>
<td>0.812</td>
</tr>
<tr>
<td></td>
<td>Desire2</td>
<td>0.864</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reciprocity</td>
<td>Respro1</td>
<td>0.859</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>Trustw1</td>
<td>0.891</td>
<td>0.814</td>
<td>0.773</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>Trustw2</td>
<td>0.914</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admiration</td>
<td>Admirl</td>
<td>0.900</td>
<td>0.772</td>
<td>0.706</td>
<td>0.871</td>
</tr>
<tr>
<td></td>
<td>Admirt2</td>
<td>0.857</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4 Correlations of the constructs; the diagonal bolded elements are square roots of the average variance extracted

<table>
<thead>
<tr>
<th>Construct</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Actual Behavior</td>
<td>0.822</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. PSD</td>
<td>0.286</td>
<td>0.893</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Enjoyment</td>
<td>-0.546</td>
<td>-0.239</td>
<td>0.838</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Desire</td>
<td>0.209</td>
<td>0.170</td>
<td>0.540</td>
<td>0.827</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Reciprocity</td>
<td>0.111</td>
<td>-0.163</td>
<td>0.168</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Trustworthiness</td>
<td>0.091</td>
<td>0.135</td>
<td>-</td>
<td>0.098</td>
<td>0.008</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>7. Admiration</td>
<td>-0.214</td>
<td>-</td>
<td>0.028</td>
<td>0.325</td>
<td>0.325</td>
<td>0.879</td>
<td></td>
</tr>
</tbody>
</table>
TRACKING FUNCTIONAL DECLINE USING AMBIENT INTELLIGENCE FOR ALZHEIMER'S PATIENTS

Sin Wee Lee\textsuperscript{1}, Usman Naeem\textsuperscript{1}, Richard Anthony\textsuperscript{2}, Abdel-Rahman Tawil\textsuperscript{1}, Muhammad Awais Azam\textsuperscript{3} and David Preston\textsuperscript{1}

\textsuperscript{1}Schools of Architecture, Computing and Engineering, University of East London, UK
\textsuperscript{2}School of Computing & Mathematical Sciences, University of Greenwich, UK
\textsuperscript{3}Department of Computer Engineering, University of Engineering & Technology Taxila, Pakistan

ABSTRACT

In a ubiquitous world we are surrounded by context sensitive information and smart devices that are able to capture information about our surroundings unobtrusively. Making use of such rich information can enable recognition of activities conducted by elderly users, and in turn can allow the possibility of tracking any functional decline. This paper describes an approach on making use of such rich information to enable recognition of activities conducted by elderly users, and in turn facilitate the possibility of tracking any functional decline. It will highlight the current methods for unobtrusively recognising activities of daily living within a home environment for Alzheimer’s patients, which is followed by a proposed taxonomy of the key characteristics that are required for robust activity recognition within a smart environment.

INTRODUCTION

In the beginning and closing stages of a person’s life they have a high level of dependency on others. In today’s working world with its emphasis on the nuclear family there are fewer problems with looking after babies, but there are significant problems in looking after the elderly. In Britain, in common with most western societies and now in Asian societies (particularly China with its single child policy), there has been an increase in the proportion of elderly people and many find it hard to look after their parents because of life commitments and distance, or just not wanting to look after their parents when needing care. Adversely, many old people want to remain independent for as long as possible. However, existence of Alzheimer’s disease (Jeffery & Cummings, 2004) among the elderly is also seen as a concern, as this disease gradually destroys the elderly person’s memory and their abilities to learn, communicate and carry out everyday activities. These elderly patients are usually sent to care homes where other people look them after. In many cases this approach is not completely successful due to issues concerning isolation or even abuse. Therefore the introduction of smart homes is gradually becoming an alternative promising solution, as it aims to provide the ability for elderly people to lead an independent life until the disease reaches a severe stage.

Studies have indicated that in the year 2000 the number of people with Alzheimer’s Disease (AD) in USA was an estimated 4.5 million, and it has been estimated that if there were no advances in therapy this figure of 4.5 million could rise to 13.2 million (Hebert et al, 2003). A study led by Knapp in 2007 on the social and economic impact of dementia in the UK has discovered more than 700,000 people suffer from dementia in the UK. This figure is expected to increase and it is predicted that by year 2025 there will be one million people in the UK with dementia. Cost is something that needs to be considered. Currently Alzheimer’s disease alone costs the UK £17 billion, which is equal to £539 per second. In the USA, it is estimated that 5.2 million Americans of all ages has AD in 2013, which included 5 million people who are aged over 65 (Association, 2013).

Table 1 shows the measure of global and cognitive dysfunction that is associated with each stage of Alzheimer’s disease (Gauthier, 2002). The measures for the mild stage indicate that independent living is possible in this stage, therefore it is important that the elderly Alzheimer’s sufferers remain in this stage for as long as possible. Conducting general ADLs such as bathing, grooming and eating; instrumental ADLs (IADLs) (Lawton & Brody, 1969; Katz, 1983) such as maintaining the household and preparing meals; and enhanced ADLs (EADLs) (Rogers et al. 1998) for example, using the Internet to shop online, can be used to assess the levels of functional decline among patients with AD’s.

It has been suggested that smart homes contribute to a better future for elderly people who are in the early stages of AD disease. It is important for the wellbeing of these elderly people that they can independently perform day-to-day tasks such as dressing, cooking, and toileting.
This has been recognised by gerontologists, who developed a detailed list of activities in 1963 (Katz et al. 1963) which should be carried out by the elderly person, known as Activities of Daily Life (ADL). They are also referred to as Activities of Daily Living (Philippose, 2005). Being able to monitor these ADLs is seen as a key approach for tracking functional decline among elderly people (Ogawa et al. 2000). Caregivers in the USA prescribe these ADLs to the elderly in order that they carry them out. Information regarding the ADL would then be collected on each visit from the caregiver, via interaction with elderly person. This collected information is important as decisions on medicine allocation depend on it. However the way this information is collected can often lead to inaccurate data, as elderly people can misinterpret facts and in many cases are forgetful. Additionally, the size of the window used for collecting such data is narrow in comparison to the period being evaluated. This shows that manual data collection regarding ADLs can be long and tedious, thus imposing further workload and burden on caregivers. Accordingly, there is a need for techniques and algorithms to be developed which are able to discriminate between different ADLs and determine the intentions of old people as they carry out such everyday tasks. One way to support elderly people who aspire to live an independent life and remain safe in their home is to find out what activities the elderly person is carrying out at a given time and provide appropriate assistance or institute safeguards.

**ACTIVITY OF DAILY RECOGNITION**

There has been a significant amount of research carried out that is focused on efficient and reliable ADL identification. Reliable ADL recognition relies on three main subcomponents (Choudhury et al. 2006):

i. **Feature Detection:** is usually a sensing level that collects appropriate information about activities that are being executed. The gathering of information can be carried out with non-intrusive ubiquitous sensors (Tapi et al., 2004) such as RFID (Fishkin et al. 2005; Stikic et al. 2008) technologies to collect activity information rather than using any visual equipment. Also the use of anonymous binary sensors such as: motion detectors, break-beam sensors, pressure mats, and contact switches can aid the process of tracking an individual around the home and complement the whole activity recognition process (Wilson & Atkeson, 2005). Utilizing the built-in kinematic sensors of a smartphone is another approach for feature detection purposes (He & Li, 2013).

ii. **Feature Selection:** is when raw sensor data from the sensing level component is manipulated into features that can help differentiate between activities. These features can correspond to high-level or low-level information. The high level information could range from information related to specific objects detected to the number of people detected in a room at the time when an activity is conducted. Low-level information could be as simple as frequency content or correlation coefficients between activities (Choudhury et al., 2006).

iii. **Models for recognition:** This component can be in the form a computational model (e.g. Hidden Markov Models, Bayesian Models), which makes use of the features from the feature-selection component for a more informed decision about which activity the person is engaged in.

In addition to the above three sub-components, semi-supervised classification is an interesting approach to ADL recognition. The semi-supervised classification approach is based around the idea of making use of labelled and unlabelled data for training and learning, where the volume of unlabelled data is greater than the labelled one. In ADL recognition context, the employment of semi-supervised learning can be used for dealing with unlabelled data generated by feature detection components.

**WEARABLE SENSORS**

Wearing different types of sensors around your body is a technique for feature detection (Kern et al., 2003; Huýnh & Schiele, 2006; Stikic & Van Laerhoven, 2007). These types of sensors are known as wearable sensors, which can range from accelerometers to audio microphones that provide data about body motion and its surroundings where the data has been collected. Wearable sensors can also be in the form intelligent gadgets, which can be reconfigurable, and scalable smart objects that can be embedded into the personal everyday goods that are used by the person to be monitored. The embedded smart object generates data that is used to log and recognise a person’s activities. Jeong et al (Jeong et al., 2008) uses these smart objects to obtain two levels of data. At the low level it is concerned with body movement and hand movement, and here the wearable device is attached around the wrist and one is attached around the wrist. At the high level it is concerned with associated predefined rules to interpret the low level information. For example a rule may be: if the body movement is fast and the wrist is being used fast then it is likely that the person being monitored is running.

Previous work (Luokwicz, et al., 2004; Ravi et al., 2005) has also shown that a variety of activities (Ermer et al., 2008) like climbing stairs and working in a workshop have been determined from similar techniques. For instance, Bao et al (Bao & Intille, 2004) carried out feature detection based on data collected from five biaxial accelerometers that were worn by 20 subjects while they conducted activities. The motivation of this work was to
conduct activity detection in a naturalistic environment as opposed to in a laboratory environment. The use of accelerometers provided data that were then labelled by the subjects themselves, this labelling was carried out without any researcher supervision. This data can then be used by supervised learning classifiers for training purposes. One of the benefits of using labelled data for training is that a collection of training data can be generated by the learning classifiers that can then allow different users to train algorithms for recognition of the activities they conduct themselves. However labelling data can be a long and tedious process.

An alternative approach was by Wang et al (Wang et al, 2007), who used wearable personal sensors to detect fine-grained arm actions like ‘drink with a glass’, ‘chop with a knife’. These were then combined with object-use data to achieve accurate activity recognition. A distinct feature of this approach was the low level of labelling required in comparison to existing approaches. This was because the accurate recognition was based on a joint probability model of object-use, physical actions and activities. It is based on a combination of generative and discriminative models and referred to as ‘common sense based joint training’. Similar work was conducted by Petney et al (Petney et al., 2006; Petney, Popescu et al., 2006) on the State Recognition using Common Sense (SRCS) system. This system works in conjunction with dense sensing as it provides a common sense interpretation of the world by forming a bond between the dense sensors to a model that represents the Open Mind Indoor Common Sense (OMICS) database. The OMICS (Gupta & Kochendeerfer, 2004) is a database, which provides basic facts that can be reasoned about. This information is produced by a small dedicated team of humans who add facts, which can then be accessed by users over the Internet for using and adding more facts relevant to the system. The SRCS converts information from this database and data mined with the KnowItAll system (Etzioni, et al., 2004) (used to bootstrap knowledge) into a large dynamic graphical probabilistic model, which is used to interpret real-world activity data (Petney et al., 2008). Conventional approximation techniques have also been applied to the SRCS system, which was used to improve the performance of the SRCS by enhancing the accuracy of SRCS’s prediction state by maximizing the likelihood by using small amount of labelled data (Petney et al. 2007). The mining aspect of the SRCS approach has similarities with mined ontology models from the Internet.

The use of wearable sensors around the body may be seen as intrusive and sometimes can get in the way of completing activities, therefore the need for a single sensing device seems like a suitable approach for feature detection. Hence the use of an RFID reader as single sensing device is a solution as it can focus just on the objects that the person interacts with rather than capturing irrelevant data (Smith, et al., 2005). RFID readers can come in the form of ring-like readers, which reads information from transponders located around the home. RFID reader can also be integrated into everyday devices that people carries with them, i.e. mobile phone (Choudhury et al. 2006). The work conducted by Lester et al (Lester et al. 2006) was in relation to providing users with integrated sensing devices within everyday tools that can be used in conjunction with minimal wearable sensors.

**MODEL INCOMPLETENESS PROBLEM**

Activity recognition based on object usage data generally uses activity models, which are models that map activities to objects that are used to carry out the activity. However, when constructing these activity models, missing information is seen as a common problem, as information from the sensing level or during feature extraction can be unexpected or even misinterpreted. Another reason for this is that it can be difficult sometimes to recognise all the relevant objects that are required to carry out the activities, this can be because of the following:

- Lack of labelled data.
- The volume of labelled data.

Objects that are not modelled are sensed. For example, if a person makes tea in a cup everyday and then decides one day to use mug, which is not encoded in the system. Any realistic system would have to deal with this, as this would be the norm for some people who actually vary in the objects they use.

A stream of natural languages terms can be used to represent the sequence of objects used to conduct an activity. For example, a stream of object data 100110111100012131121093232 can be mapped into: Kettle-Sugar-Bowl-Milk Carton. This representation allows the mapping from these terms (based on the object data) to activity names (e.g. make tea), which makes it possible to mine generic activity models from the web in order to segment the stream of object data into instances of activities. For example, given a stream of object data that consists of “kettle” and “tea bag bowl”, we would segment out many instances of the activity making tea. This is then followed by using the labelled instances to learn custom models of the activity “make tea” from the collected data. During this learning process, it is also possible to learn the variations and the use of different object to perform the activity, e.g. having milk, but no sugar. Perkowitz et al (Perkowitz et al. 2004) was able to carry out classification of activities with this approach to a certain extent, as it worked well with hand-segmented data. However, since these models are generic they are reliant on websites that follow a particular format so that information can be mined from them. This results in
certain activities being unrecognised as the information from the web does not have any relation with the actual objects used to conduct the activity. Wyatt et al (2005) also developed a technique for mining from the web, and these can be applied to segments of unrecognised stream data, as well as being able to label streams of object-usage data. Given a set of activities A (e.g. Make Tea), this technique mines a set of objects O (e.g. Mug) from the web that are used for each activity an A.

The approaches mentioned so far for mining models are mainly focused on being able to deal with unlabelled data, e.g. unsupervised. The problem of missing information or if a model omits an object while incorporating a similar one has been explored by Tapia et al (2006), who have developed an unsupervised approach that uses information mined from an ontology of reference system for English languages, called WordNet. Ontologies have been utilised to construct reliable activity models that are able to match unknown sensor readings with a word in an ontology is related to the sensor event. For example object ‘Mug’ (a sensed event but not associated with the task ‘make tea’) could be substituted for a ‘Cup’ object, as the model make tea recognises ‘Cup’, because ‘Mug’ has not been modelled in the make tea model. In addition a statistical smoothing technique called shrinkage has been applied to this approach. This technique has been used by many researchers for situations where it is not possible to reliably compute parameter values of a given model from training data alone, hence shrinkage is used to improve a given model’s estimated parameter values. In the context of ontologies, shrinkage has been used to improve the probability estimates of leaf nodes that are generated within the ontology.

As well as activity recognition, ontologies have also been deployed for collaborative healthcare experiments to support and enhance the living of elderly people. One such example is where Wang et al (2008) deployed ontologies to manage a collaborative healthcare environment, used on information concerning elderly people are used to present important context aware information, which can be crucial for emergency treatment.

As described above, there are several existing approaches used to recognise ADLs, and the state of the art is such that in many cases it is possible to determine ADLs from the use of objects.

The approaches that have been mentioned may be able to carry out classification and learning for the activities, however they cannot handle interactions, suspension and interweaving well and are unable to predict what ADL may follow from a previous ADL, i.e. do not reason about sequences of ADLs.

**WHAT MAKES A SUCCESSFUL SMART ENVIRONMENT FOR ADL RECOGNITION?**

ADL recognition comprises of several fundamental stages. Firstly the user’s behaviour must be sensed, then the sensed data must be processed to extract key features which identify the specific activities that have occurred, and then the activities must be examined to understand them in the application context – this can include sequencing them and looking for patterns in terms of timing and duration.

Each of these stages can be performed in a wide variety of ways, as the earlier discussion has illustrated. A number of factors influence the design choices for a given application scenario. We are specifically concerned with elderly users who have, or may have some degree of AD and the purpose of the activity detection is to monitor the user in order to detect the symptoms of AD and to track any degeneration that occurs over time. Thus our system constraints are distinctly different to those of monitoring systems where the interest is in the activities themselves. We are more interested in the way a task is performed, in terms of correct sequencing and completeness of the task, and detecting behavioural traits such as repetitive stages or mixing up activities which may indicate confusion or cognitive decline, than we are in the study of what steps are involved in a particular activity and can the process be improved for efficiency (as we would be for example if the smart environment were used for time-and-motion study in a factory).

In a system that is fundamentally concerned with accurate detection of the activities themselves it would be appropriate to discard incomplete activities (whether they are incomplete due to poor sensor placement, sensor failure or because the user really did not complete the activity) as noise. However, incomplete activities can be a source of valuable information when monitoring people with some degree of AD – the ratio of incomplete activities against complete activities could be an indicator of cognitive ability. Similarly, for users with AD, variations of activity duration, and repeated sub-steps within an activity can be vital clues that the person is struggling to complete the activity, which in other systems might simply indicate a false activity detection.

Hence the design of the smart environment must specifically cater for the fact that some user activity may be inconsistent between different users, and for the same user from time to time. The choice of sensor types and sensor placement is a good place to start. Sensors that produce implicitly labelled event data are the most valuable because these provide firm anchors when dealing with incomplete or inconsistent chains of events. For example, suppose that a series of events have been detected that could indicate that the user is making themselves a hot drink, but could also indicate several other similar activities in the kitchen. An RFID sensor that detects the kettle has been touched provides a clue that the kettle is involved in the activity (this could be considered a contextually-labelled sensor – as it is dependent on other
sensors in a sequence to indicate for example that the user is boiling water in the kettle – or they could just be moving it out of the way to perform some other activity). However, a sensor on the power circuit of the kettle senses that the kettle has actually been turned on, and the duration can be measured to confirm that the kettle has indeed boiled some water. This could be classed as an implicitly-labelled sensor and is more valuable in predicting or confirming the making-drink activity. In cases where the sensor detection sequence is incomplete, the presence of the implicitly labelled sensor is highly advantageous and thus such sensors should be given higher weighting in the reasoning process. Unlabelled (or weakly labelled) sensors such as a movement detector that indicates that someone is in the kitchen are useful mainly to reinforce the value of the more specific sensors. Even a number of weakly labelled sensors in combination, such as {kitchen light is on, curtain is closed, movement is detected} do not give much information as to what activity is being performed.

Another key design consideration is the way in which sensing is performed. Generally it will be useful to know the duration of an event rather than just the fact that an event has been detected. Thus, where the sensors (and the underlying event) support it, it is useful to detect two signals {start of event, end of event} and thus to permit additional reasoning about duration of events and concurrent events whose active periods overlap. Some sensors such as a PIR movement detector send a series of signals when movement is detected – the output from sensors in this category needs to be aggregated such that meaningful events are extracted from the raw signals. A burst of signals after a silence period could indicate that someone has entered a room, or perhaps that someone who has been sitting still in a chair for a while has started to move, perhaps they have now stood up – this sort of detail would need to be added by signals from other sensors to build up a context. Clearly though, once the primary detection has been performed (e.g. movement where there was previously none) the value of the continued series of triggers is diminished and if not depreciated may act as noise in the subsequent activity classification. The point at which the movement stops is however very significant and this is an example of an event that is detected by the lack of signal from a sensor – which further reinforces the message that the processing of signal data to extract event data must be carefully designed for purpose.

Aggregation of data at the signal level can also reduce the costs of processing, communicating and storing samples, but this must be performed so as to preserve all of the information needed to detect user-level events. For example if someone is moving to and from between rooms there might be a number of signals detected from a sensor in the hallway that connects rooms (this could be a Passive Infra-Red sensor (PIR), or an under-carpet pressure mat, for example). Trying to make sense of each activation might over-complexify the feature detection, but a higher-level composite feature could instead be detected, named heavy-traffic-in-hall. This could be used to help identify the higher-level scenario without needing to know the specific purpose of each of the journeys. For example, the user might be tidying up the house and thus making a lot of journeys from room to room, or perhaps they are the hall and the adjoining rooms. It could also indicate a confused person wandering about, or even sleep-walking. Additional context would be needed to confirm which of these scenarios was being played out – such context could include the duration of activity, and the time of day. A system that focuses on classification of all of the sub-events in such a scenario (the individual journeys), may fail to detect the higher-level event; thus the need for intelligently aggregating the signals when appropriate. Equally though, when a single activation of the sensor is detected within a given timeframe it must be treated as a separate signal (i.e. the system should not aggregate by default but should intelligently apply aggregation and / or filtering).

In terms of activity detection, a system can be pre-programmed to ‘look for’ a certain set of activities, or can be equipped with a learning capability in which unknown clusters of signals can be detected (by their distance from the known clusters). ADLs can be learned based on the sequence of activities and separation into categories by means such as cluster analysis or Case-Based-Reasoning (CBR). Alternatively, if a fixed set of specific ADLs is to be monitored then the sensor type and placement can be much more directed and a smaller ‘event space’ is needed – i.e. the amount of information required to be collected concerning event start/stop, duration and precise sequence is lower if the goal is to match a limited set of pre-described patterns that the system is purposely set up to detect. However such a system is typically inflexible, so there is a much higher dependency on design-time understanding of the use-scenarios.

In addition to the technical aspects of the sensing and processing, there are also a number of human factors that must be considered very carefully. One such consideration is whether the sensing should be performed entirely passively, or whether active participation of the user is helpful. In some systems, such as a security application at an airport, or a safety shutdown system on a semi-automated factory production system, it might be appropriate that a user provides some specific active input before a certain event is allowed to happen, or is required to confirm an event that the system has detected. Certainly any active inputs will be classified as implicitly-labelled signals and these will always be the highest value signals for classification purposes. However, when detecting functional decline in AD patients active inputs would be unreliable by the very nature of the AD illness, but also it adds a burden onto the user which may not be welcome.
There are a number of valuable extensions that can however be added onto a basic ADL monitoring system. For example, some basic telecare features could be added at very low additional cost once the basic system is in place. A simple example of this could be a voice prompt that is played each morning when the user first enters the kitchen “How are you feeling today – if you are well please press the green button, if you are unwell and need assistance please press the red button”. Of course, even such a simple adaptation such as this raises a challenging question – what to do if no button is pressed? This also serves as a warning not to make systems over-complex (which can be a side effect of an over-featured design in which some features do not add real value).

Another human factor to consider is privacy. This must be balanced carefully with the amount of information needed to correctly classify activities. A completely sensor-agnostic system can be built where the patterns in sensor activation can be learned. Such a system has value in the ability to learn the normal patterns of behavior of a user and thus to detect an abrupt change – which could indicate that the person is unwell, or otherwise needs help – especially in the case that the events stop or reduce in frequency dramatically. Such a system preserves privacy – for example there is no need for the collected data to show that one of the sensor is connected to the lavatory flush, and another is connected to the biscuit tin – the user’s behavior is not revealed. However, this approach is of very little, if any, use if we are trying to detect ADLs and track the performance of these over time. There is a certain level of behavior-revealing information that must be tracked for the system to be successful, but this can be done considerably, so that privacy is not violated. For example a user might consider that a sensor that measures how long they are sat on the toilet crosses the fine line of acceptability.

In any monitoring system, there is the possibility that ADLs will overlap, or will be only partially completed and this of course is much more likely with AD patients. Even with a prescribed activity set, the monitoring system will detect non-matching activations. In such cases it needs to be determined whether to discard such information – or to provide the closest match – this could be a useful indicator of cognitive decline over time.

Perhaps the single largest issue for the analysis of data from the sensors is the variation in the users behavior. Once-precise actions may become vague over time, such as placing objects near to their correct location, or doing things in sequences that become less regular or precise over time. People with AD sometimes experience ‘good’ and ‘bad’ days – in which their mental ‘sharpness’ fluctuates and thus they will interact with their environment more positively on some days and more passively on others. Passive engagement might be detected by increased randomness in the timing of events, and it is beneficial to sense the duration of events as well as sequence of sensor activations, as a person who is confused will take a lot longer to complete a task that on another occasion may be performed efficiently.

There is also the issue that different activities have different sensing requirements – in terms of informational, temporal and spatial aspects. For example, some activities need to be sampled over a short time frame (such as closing the curtains) whilst others span a significant time frame, such as preparing a meal, eating a meal. Some activities have closed spatial bounds, such as activities wholly carried out in a bathroom (washing, using the toilet), or a kitchen (making a drink) whilst others such as reading a book could occur in different spaces, and could continue across spatial boundaries. Some activities, such as flushing the toilet can be detected definitively by a single (implicitly labelled) sensor, whilst some activities may require more information, such as a sequence of several different sensor activations, in order to be correctly detected. One possibility when designing the monitoring system would be to take a pragmatic approach and to focus on the easier-to-detect ADLs, but this must be balanced against the need to collect information concerning a range of different ADLs. It is likely that the toilet flush activity reveals far less about functional decline than an information-rich activity such as preparing a meal. An ideal situation would arise if a suitably diverse range of ADLs could be identified that provide good coverage of the range of human behaviours, whilst all having low-complexity detection requirements.

Another aspect that must be considered when designing a smart environment for ADL detection is that the physical environments themselves are non-uniform. A system that uses a zone concept such as ‘the user is in zone A (bath)’ or ‘the user is in zone B (toilet)’ will work well in an environment where the toilet is in a separate room, but the zones will overlap if there is a single bathroom space which includes the lavatory. Other common differences are that some premises will have a hallway, whilst some street-doors open directly into a lounge. Some homes have a separate kitchen and dining room, whilst others have a joint kitchen-diner. Some users will have a separate fridge and freezer, whilst some with have a combined unit, or only a fridge. The set of sensors that are deployed must be appropriate for the accommodation both in terms of its size and layout. In some small flats a single PIR movement detector will pick up almost all movement and thus will be particularly unhelpful as an activity discriminator. The reasoning system must be designed to be able to work with almost any subset configuration of the supported sensor types.

Finally there are some issues of practicality. One aspect of this is cost effectiveness and availability. As sensor systems become more common, the variety of sensors available is expected to increase whilst their cost falls.
There is also the challenge of ensuring systems are realistic for real-world deployments. Much of the leading research at present has been tested in controlled and semi-controlled environments. The use of RFID sensing is one example technology that works well in a lab or with volunteers who agree to wear the special glove that contains the RFID reader that must pass within a few centimeters of the transponder which may be attached to the various objects such as cups and spoons, but will AD patients be happy to do this over long time frames? Will they remember to keep the glove on? Similarly, wearing a single accelerometer on a wrist-band might be acceptable, but applications where multiple sensors need to be attached to the body might struggle for acceptance by users.

TAXONOMY OF ADL MONITORING SMART ENVIRONMENT CHARACTERISTICS

From the discussion above, a number of decision categories and tradeoffs that arise when designing the ADL monitoring smart environment can be extracted.

Sensor types. Implicitly-labelled sensors are more useful in terms of activity detection and discrimination, but these may be generally more expensive or not always available. Unlabelled data such as signals from PIR sensors that detect any movement are poor discriminators unless used in combination with other sensor data and context.

Sampling method. Using instantaneous samples may be appropriate in some circumstances, but in others it will be necessary to smooth samples over time, for example using an exponential mean to favour more-recent values. For many events, knowing both the event-start and event-end time, and thus being able to determine duration, is important.

Approach to data aggregation. Filtering and aggregation of data can be powerful ways to optimize processing in terms of cost and effectiveness, but must be done such that the key underlying information is retained. A signal that is considered noise in one context may be useful data in another.

Pre-programmed activity detection versus learning capability. The tradeoff is between a simpler, easier to develop system with design-time limited scope, and a more costly system that can learn patterns and activities the designers were unaware of.

Passive monitoring versus active involvement. The former is generally better suited to AD patients, but the latter provides more definitive information. Some compromise middle ground may be found in some systems.

The extent of privacy afforded to the user. Sensing of ADL, by its nature cannot totally preserve privacy, but the sensor systems can be designed such that they only reveal less-sensitive ADLs or obfuscate some aspects of an ADL. For example knowing that the toilet was used may be adequate information in some systems. Revealing the actual time spent sat on the toilet might be a step too far for some users.

Coarseness of ADL set. Fine-grained ADLs lead to a rich information base, but at the risk of greater sensing cost, cluttered environments and activity discrimination problems as scenarios are more likely to overlap. On the other hand, using only a small set of coarsely defined ADLs might not provide sufficient information for the required level of analysis.

ADL detection complexity. Here the tradeoff is between easier-to-detect ADLs requiring a single activation of a single sensor, but providing limited behavioural information, versus complex ADLs rich in behavioural information but having high detection cost in terms of temporal, spatial or informational complexity.

Support for environmental variation. A low cost system might require a standard deployment into an assumed fixed environment design. This could work well for care-homes and hospitals. A more-flexible system is better suited to private homes, allowing for different sets of sensors to be installed as appropriate to suit differences in environment, and the specific needs of individual users.

Means of communication between sensors. Wired systems and wireless systems each have pros and cons that must be carefully considered.

CONCLUSION

Most of the western world has a severe ageing population problem, in which an increasing proportion of the population is elderly and requires care. This care must be provided by, and paid for, the reducing working population. The problem has attracted a lot of research interest, both academic and commercial, because of the urgency on one hand, and because of the scale of the potential benefits of technological solutions.

Rapid advances in technology, especially in terms of sensors and the availability of low-cost embedded systems have occurred over recent years. This is currently being driven in large part by the Internet-of-Things initiative which seeks to interconnect a very diverse set of common objects, has led to a rich technological basis to support applications in areas such as assistive technologies, telecare and home automation.

ADL recognition is a specific challenge which overlaps these broader domains. It has very high potential value to society at a number of levels including understanding the nature of AD and its effects, and tracking the development of AD in individuals.

ADL recognition is quite broadly defined and a variety of experimental systems have been developed. There are certain common characteristics in terms of the
technological requirements and the general types of data that need to be collected.

However, great care is needed in the design of smart sensor systems for the recognition of ADLs. In this regard, the main challenges on these systems must be overcome. A taxonomy which identifies the main design choices and the tradeoffs that arise when making these decisions is proposed.

REFERENCES


**TABLES AND FIGURES**

**Table 1 - Measures Of Global And Cognitive Dysfunction**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Duration, Year</th>
<th>Global Deterioration Scale, *</th>
<th>Mini Mental State Exam, † score</th>
<th>Global autonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>2-3</td>
<td>3-4</td>
<td>26-18</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Moderate</td>
<td>2</td>
<td>5</td>
<td>17-10</td>
<td>Supervision Required</td>
</tr>
<tr>
<td>Severe</td>
<td>2-3</td>
<td>6-7</td>
<td>9-0</td>
<td>Total Dependence</td>
</tr>
</tbody>
</table>

*Scale measures progressive need for assistance in daily activities (e.g., choosing clothes, dressing); scores range from 1–2 (normal) through 6–7 (severe dysfunction).

†This 22-item scale measures cognitive function; scores range from 30 (excellent function) to 0 (severe dysfunction).

**Figure 1 - Four steps for mining activity models**

1. Web pages are identified for each activity being performed in A.
2. Phrases are extracted that describe the objects used to perform the activity.
3. Object-use probabilities are estimated from co-occurrence statistics of the pages and phrases.
4. Hidden Markov Model (HMM) is assembled from the mined information, in order to recognise streams of object usage data.
REAL TIME HUMAN ACTIVITY RECOGNITION FRAMEWORK

ABSTRACT

In this paper, we focus on designing a real time human activity recognition framework to accommodate the requirement for various contexts. This work was motivated by our previous works where a behavior monitoring model for special children were developed for monitoring and preventing physical injury in classroom environment using Kinect sensor. The proposed framework allows developers to develop activity recognition model by assembling the modules we defined easily and flexible. Our framework promote re-usable, flexible, computationally inexpensive, scalable, incrementally improvable, easily maintain and distribute. This framework will benefit researchers or practitioners who intended to develop activity recognition model for various contexts using Kinect sensor i.e. e-Healthcare, monitoring or surveillance system, human computer interaction (HCI), Assistive Technology (AT) and sensor applications.

INTRODUCTION

Human activity recognition is an active field for research (Lara and Labrador, 2013; Chen, Wei and Ferryman, 2013; Iosifidis, Tefas and Pitas, 2013). According to Yang, Lee and Choi (2011), Activity recognition in Information & Communication Technology (ICT) is the ability to distinguish various activities perform by human based on information acquired from various sensors i.e. Camera, microphone, ambience sensor, and wearable device. Many tasks can be revolutionized if machine can interpret human daily activity (Chen, Wei and Ferryman, 2013). Activity recognition can be used in various contexts i.e. in classroom or residential environment, hospital, rehabilitation center, airports and even for healthcare purposes. Researchers i.e. Zwartjes et al. (2010) performed activity recognition for chronic disease management. Sazonov et al. (2011) use it for patient monitoring in rehabilitation center. Pulkkinen (2013) monitor diabetic patient at home with home user activity recognition for health status monitoring (Warren et al., 2010). Besides, activity recognition is also widely used for behavior monitoring for both elderly and children (Chen, Wei and Ferryman, 2013; Stone and Skubic, 2012). Project like Smart Home to Assist Elderly with Alzheimer's Disease (AD) in completing their Activity of Daily Living (ADL) tasks is conducted by monitoring their behavior and environment in real time then respond to their needs (Chen Nugent and Wang, 2012). For example, an elderly will be reminded to brush his / her teeth if “Smart Home System” did not detect any brushing teeth activity is performed after they wake up. Further to this, Robert, et al. (2013) also highlighted there is a need for ICT players to provide solutions that help people with Alzheimer's Disease and their caretakers to better understand Behavioral and Psychological Symptoms of Dementia (BPSD) while providing help with their daily living activity.

Our studies have identified two general ways for machine to recognize human activity which include vision based and wearable device approach (Chen, Nugent and Rafferty, 2013). Works such as Chen, Wei and Ferryman (2013) , Chen Nugent and Wang (2012), and Roshtkhari and Levine (2013) are using vision based approach to perform activity recognition. An example can be found in Chen, Nugent and Wang (2012) where they uses RGB camera in Smart Homes for this purpose. On the other hand, human activity recognition using wearable sensors or specifically device equipped with accelerometer or GPS and tag-sensor can be found in Lara and Labrador (2013), Yang, Lee and Choi (2011), Sazonov et al. (2011) and Alshurafa, et al. (2013). Yang, Lee and Choi (2011) performed activity recognition by attaching "Radio-frequency identification tag" (RFID) on object and sensing these tag using smart mobile device and then interpret these data. Alshurafa et al. (2013) designed activity recognition framework for health and exergaming using wearable motion sensor.

Although activity recognition task by machine can be done by using various sensors, the challenges still exist especially when processing the data acquired. A common challenge for both vision and wearable device based activity recognition system is the data training. A huge
and large dataset is required as input to train the machine to manipulate data acquired and then recognizes the pattern (Lara and Labrador, 2013; Chen, Wei and Ferryman, 2013). Various complex machine learning or classification algorithms i.e. Support Vector Machine (SVM) is employed to perform recognition tasks. On top of that, activity recognition using RGB camera requires complex and intensive data processing i.e. Stream or Motion Capturing, Segmentation, Feature Extraction, Action training or learning and classification (Roshtkhari and Levine, 2013; Alshurafa et al., 2013; Weinland, Ronfard and Boyer, 2011). This causes the human activity recognition become inefficient in real time as it is computationally expensive.

Chen, Wei and Ferryman (2013) initiated a survey and analyze the potential of depth sensor in interpreting human motion or specifically human activity by machine i.e. Kinect sensor. The Kinect sensor (Microsoft Corporation, 2013) as shown in Figure 1 is embedded with infrared projector & CMOS sensor powered by PrimeSense (Chen, Wei and Ferryman, 2013). The Kinect has the capability to measure the depth of every pixel from its view, detect human position and locate 20 human skeletal joints with its coordinate (x, y and depth axis) via the Natural User Interface (NUI) from "Kinect for Windows SDK". However, motion or human activity recognition capability is not available in Kinect SDK 1.8.

PREVIOUS WORKS

We defined 23 algorithms to process the depth data acquired from Kinect sensor for human activity identification previously (Ong, Lau, and Bagha, 2013; Lau, Ong and Bagha, 2013). Each algorithm can recognize one activity such as falling, idle, left and/or right hand waving, seating, standing, walking, running, bending, brushing teeth, drinking, eating, writing, reading, combing hair, climbing, jumping, kicking, punching, slapping, smashing, head knocking, self-hitting, and clapping hands. These algorithms will return true flag if the data received fulfills its condition. We manage to obtain an overall average recognition rate at 89.83% for 23 algorithms where we tested on 10 volunteers.

Thought a series of our algorithm testing and evaluation, we realized that we could branch our research by designing an activity recognition framework where researcher or developer could easily plug in and remove defined algorithm(s) from the model to suit specific purpose or context. This is also in line with the observation concluded by Chen, Nugent and Rafferty (2013) where most of current research is focusing on the improvement on component-level methods, techniques and technologies, e.g. sensor data segmentation, activity modelling and activity recognition. There is lack of contribution in this field in term of system-level solutions, e.g. the integrated architectural or framework and technology infrastructure that are reusable, applicable and scalable to real world problem solving.

In mobile application development, android and IOS do not develop or provide all the apps available in market by themselves (GooglePlay Store – Android; Apple Store - IOS). They are successful and rapidly grow because they provide a standard platform or framework (SDK) to the community i.e. application developers while the community contribute the contents or so called "mobile apps" in return. We call this as "One for All and All for One Ecosystem". Windows 8 are following this trend recently where encouragement is given to application developers all around to world to place their apps in Store (Endler, 2013). Free Windows Mobile App Development tools were also available in Microsoft DreamSpark (2014) portal under student license.

We believe this is the emerging and current trend for software development and thus, we proposed a real time human activity recognition framework to accommodate the requirement for various contexts in human activity recognition either for healthcare, behavior monitoring or even for human computer interaction purposes. We believe the proposed framework will grow and contribute to related research community.

PROPOSED FRAMEWORK

Our framework consist six modules which can be categories into Core Modules (CM) and Optional Modules (OM). CMs are the module that must be included in a model i.e. Setup & Configuration Module, Skeleton Controller Module, and Human Data Module. While OM consists of Guardian Profile Module, Notification Module and algorithms we defined earlier.

The Setup & Configuration Module allows user to preset the setting of assembled model i.e. enabling detection and notification, views and other settings mode. The Skeleton Controller Module consists of features i.e. switching incoming stream from Kinect sensors, and switching human tracking.

Human Data module keeps all information about tracked human i.e. joints' coordinate (in x, y and depth-axis) as well as height (Figure 2). An array of historicalSkeletonData can be found inside Human Data module as well. These collections are useful for algorithm such as walking, running, jumping and etc where the variation of joint(s) over the time can be measured. Apart from that, a joint's acceleration and velocity can be measured by utilizing the historicalSkeletonData array too.

The numbers of Human Data can be plug-in into a model is depending on numbers of connected Kinect sensor(s). For example, two Human Data Modules are
included if a Kinect sensor is connected; four Human Data Modules can be included if two Kinect sensors are connected.

As mentioned before, we defined 23 human activity recognition algorithms and these algorithms can be plug-in and removed easily from a model. Each activity recognition algorithm is represented by an algorithm module. We have defined a standard template for activity recognition module which is shown in Figure 3. Guardian Profile Module is responsible to keep guardian’s contacts i.e. mobile phone number, email address, facebook account and other information. Lastly, Notification Module is responsible to send message or email to guardian(s) when necessary.

Figure 4 show the general flow and illustrate the human activity recognition framework. A Kinect sensor will acquire infrared signal which projected to the scene and passed it to NUI API for manipulation. The NUI API then generate skeletal stream based on infrared signal received. The Skeleton Controller Module will then process the stream acquired from NUI API based on predefined setting and configuration. These processes may include identifying human of interest (HOI) and tracking, measuring HOI’s height, and produce Human Data (1 & 2). Next, The Human Data will be feed into all activity recognition algorithms and identify the activity. An algorithm will eventually send true flag to Notification Module if Human Data received fulfilled its condition.

When Notification Module received true flag from any algorithm, it will send notification in the form of either SMS, email, message or all to guardian as defined in Guardian Profile Module.

EXPERIMENTAL MODELS’ ASSEMBLING,
EVALUATION & DISCUSSION

Two experimental models namely behavior monitoring model for children with special needs and behavior monitoring model for elderly with AD in their resident were assembled. Table 1 summarized specification and modules required for both behavior monitoring model.

The first model were assembled to monitor the behavior of children with special needs and also to detect physical injury i.e. fall. Since the targeted environment was in classroom, we include 17 activities recognition module in this model as these activities are likely to happen in classroom. These activities includes seating, standing, walking, running, falling, idle, left / right hand waving, jumping, climbing, slapping, punching, smashing, kicking, head knocking, self-hitting, reading and writing. The left over 6 algorithms defined i.e. bending, brushing teeth, drinking, eating, combing hair, and clapping hand were not in use in classroom context as these activity are not dangerous and unlikely to happen.

The second was assembled using the same framework to monitor Activity of Daily Living for elderly with AD in their resident. This model were assembled as feeder for external system in clinical and physiotherapy centre where caregivers / nurses could keep track daily activities of an elderly i.e. numbers of hours an elderly walk per day, did the elderly brush their teeth in the morning and before went to bed as well as other daily activities which are included as part of the treatment or clinical assessment. Among activity recognition module included are seating, standing, walking, running, bending, combing hair, brushing, eating, drinking, idle / rest, reading, and writing. All the recognition outcomes can be feed into external devices or system for further analysis or treat as reminder to the elderly.

We performed re-evaluation on both assembled behavior monitoring models with pre-recorded video sequences and manage to acquire overall activity recognition accuracy at 92.12% and 89.17% respectively as shown in Table 2. Overall, we are confident that the framework we proposed could contribute to researchers and practitioners in related field i.e. e-Healthcare, monitoring or surveillance system, human computer interaction (HCI), Assistive Technology (AT) and so forth. Researchers and practitioners who employ this framework for development can define their own activity recognition algorithms based on their needs. This will saves their time and avoid developing the whole model from scratch. Besides, newly defined algorithm is exchangeable among developers as these algorithms can be plug into their model easily via this framework. This indirectly creates the opportunity for researchers or practitioner to contribute their recognition algorithm into our framework and this framework is re-usable by the same community for various purposes. In short, our framework promotes modules re-use, flexible, computationally inexpensive, scalable, incrementally improvabre, easily maintain and distribute.

CONCLUSION

In this paper, we reviews the usefulness of human activity recognition in different contexts and ways for machine to interpret human activities either using visual based or wearable device based approach. We reveal the complexity of using RGB camera (visual based) and the cumbersomeness of wearable device based approach to perform human activity recognition task. We further investigate the potential of depth sensor i.e. Kinect in interpreting human motion or specifically human activity by machine. The outcome of our investigation are convincing where we believe human activity can be recognized by machine using depth sensor easily and
efficient. However, no standard framework or API was found so far and this has motivated us to propose a framework.

A real-time human activity recognition framework for Kinect sensor is proposed in this paper. We defined various modules in our framework which allows researchers and practitioners to assemble activity recognition model based on desire requirements and needs. We also assembled two models based on this framework for re-evaluation and the outcomes are convincing. It shows that this framework is scalable where researchers or practitioners can define their own activity recognition algorithm and exchange with other.

We believe this framework would ease and benefit the community of human activity recognition using Kinect sensor regardless the context, purpose and requirements. Apparently, we are in the mist of wrapping the framework and preparing the framework documentation. We are going to define more activity recognition algorithms and continuously improve our framework in future.

REFERENCES


FIGURES AND TABLES

Fig. 1 Kinect sensor (without casing)

Fig. 2 Class Specification: SkeletonController, HumanData, ActivityRecognitionTemplate

Public Shared Function checkState(ByRef humanData As HumanData) As Boolean

Dim flag As Boolean = False

'''Define local variables

'''Define condition(s) and change flag to True if condition(s) defined is / are fulfilled.

'''Compare the variation between Joint(s) with SkeletonArray (History Data) over time (if any).

'''Checking condition(s) with data available inside HumanData instances i.e. userHeight, userDistance, joint coordinate etc...

Return flag
End Function

Fig. 3 Standard template for activity recognition module (algorithm)

Fig. 4 Proposed framework
Table 1 Experimental Models

<table>
<thead>
<tr>
<th>Components</th>
<th>Behavior Monitoring Model for children with special needs</th>
<th>Behavior Monitoring Model for elderly with Alzheimer's Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Modules</td>
<td>1. Setup &amp; Configuration Module</td>
<td>1. Setup &amp; Configuration Module</td>
</tr>
<tr>
<td></td>
<td>2. Skeleton Controller Module</td>
<td>2. Skeleton Controller Module</td>
</tr>
<tr>
<td></td>
<td>3. Human Data Module</td>
<td>3. Human Data Module</td>
</tr>
<tr>
<td>Optional Modules</td>
<td>1. Guardian Profile Module</td>
<td>1. Notification Module</td>
</tr>
<tr>
<td>No. of ARMs</td>
<td>17</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity Recognition Modules (ARMs)</th>
<th>Behavior Monitoring Model for children with special needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Seating</td>
<td>1. Seating</td>
</tr>
<tr>
<td>2. Standing</td>
<td>2. Standing</td>
</tr>
<tr>
<td>3. Walking</td>
<td>3. Walking</td>
</tr>
<tr>
<td>4. Writing</td>
<td>4. Running</td>
</tr>
<tr>
<td>5. Reading</td>
<td>5. Reading</td>
</tr>
<tr>
<td>6. Writing</td>
<td>6. Writing</td>
</tr>
<tr>
<td>7. Idle</td>
<td>7. Idle / Rest</td>
</tr>
<tr>
<td>8. Failing</td>
<td>8. Bending</td>
</tr>
<tr>
<td>11. Climbing</td>
<td>11. Eating</td>
</tr>
<tr>
<td>12. Slapping</td>
<td>12. Drinking</td>
</tr>
<tr>
<td>13. Punching</td>
<td></td>
</tr>
<tr>
<td>14. Smashing</td>
<td></td>
</tr>
<tr>
<td>15. Kicking</td>
<td></td>
</tr>
<tr>
<td>16. Head Knocking</td>
<td></td>
</tr>
<tr>
<td>17. Self-Hitting</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recognition Output (feeds)</th>
<th>SMS Module</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Email Module</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>Behavior Monitoring Model for children with special needs</th>
<th>Behavior Monitoring Model for elderly with Alzheimer's Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall ARMs Accuracy</td>
<td>92.12%</td>
<td>89.17%</td>
</tr>
</tbody>
</table>
NUMERICAL INVESTIGATION OF FINITE-AMPLITUDE PRESSURE WAVE PROPAGATION IN GAS SUSPENSIONS: APPLICATION TO AEROSOL CHARACTERIZATION

Riady Siswoyo Jo¹, Basil T. Wong¹, Anatoli Vakhguei²
¹Faculty of Engineering, Computing and Science
Swinburne University of Technology Sarawak Campus
Kuching 93350, Malaysia
²Mechanical Engineering Discipline, School of Engineering
Nazarbayev University
Astana 010000, Kazakhstan

ABSTRACT

Aerosol characterization is an important procedure that has been carried out for many applications, ranging from manufacturing, healthcare to environmental monitoring. This paper proposes an alternative method to perform aerosol characterization based on pressure wave propagation. Preliminary numerical study of pressure wave propagation in gaseous media containing suspended particles is presented in detail. A numerical scheme based on Crank-Nicolson finite differencing and trapezoidal quadrature discretization methods is employed to simulate the evolution and energy dissipation of the pressure wave. The solutions can be used to characterize size and volume fraction of aerosol droplets in air. Simulation results are presented and discussed.

NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_p )</td>
<td>Volume fraction of particle</td>
</tr>
<tr>
<td>( \delta )</td>
<td>Radius of particle (m)</td>
</tr>
<tr>
<td>( a )</td>
<td>Thermal diffusivity of particle (m²/sec)</td>
</tr>
<tr>
<td>( C_{part} )</td>
<td>Specific heat of particle (J/kg.K)</td>
</tr>
<tr>
<td>( \rho_p )</td>
<td>Density of particle's material (kg/m³)</td>
</tr>
<tr>
<td>( C_v )</td>
<td>Specific heat of gas in constant volume (J/kg.K)</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>Adiabatic constant of gas</td>
</tr>
<tr>
<td>( \nu )</td>
<td>Kinematic viscosity of gas (m²/sec)</td>
</tr>
<tr>
<td>( \rho_0 )</td>
<td>Density of gas (kg/m³)</td>
</tr>
<tr>
<td>( \tau_v )</td>
<td>Characteristic time for momentum exchange (sec)</td>
</tr>
<tr>
<td>( \tau_t )</td>
<td>Characteristic time for thermal exchange (sec)</td>
</tr>
<tr>
<td>( \hat{P} )</td>
<td>Dimensionless relative pressure deviation</td>
</tr>
<tr>
<td>( \hat{\tau} )</td>
<td>Dimensionless retarded time</td>
</tr>
<tr>
<td>( \hat{x} )</td>
<td>Dimensionless space variable</td>
</tr>
<tr>
<td>( T )</td>
<td>Period of pressure disturbance (sec)</td>
</tr>
</tbody>
</table>

INTRODUCTION

Propagation of Pressure Wave in Gas Suspension

Propagations and structures of pressure (temperature, density and velocity) disturbances in gas suspensions (gaseous media consisting suspended liquid or solid particles; also referred to as aerosols) are common in our environment, e.g. misty and dusty air and in numerous industrial process setups, e.g. automotive spray painting, food and powder processing and acoustic cleaning systems as discussed by numerous studies (Fan and Zhu, 2005; Elperin, et al., 1988; Busnaini and Elsawy, 1998). These disturbances are referred to as pressure waves. It is well known that as a pressure wave propagates through such a medium, it interacts with the suspended particles by exchanging its energy.

There are two important aspects to consider in modelling pressure wave propagation in the gas suspensions. First, the wave evolution plays significant role in describing the change of shape of the wave profile, which in turn might provide information of wave steepening and shock formation. Second, the attenuation of pressure wave is important in describing the energy dissipation of the travelling wave. Several authors (Davidson, 1975; Gregor and Rumph, 1976) investigated this phenomenon by considering the evolution and attenuation of a propagating sonic wave in the gas suspension. The aforementioned works are indeed very useful for modelling pressure wave in the audible amplitudes and frequencies. In this case, the wave is assumed to be almost linear (having a sinusoidal profile). For linear waves, the attenuation can be assumed exponential and is independent of the initial amplitude of the pressure disturbance (Gregor and Rumph, 1976).

However, in many environmental and industrial setups, the amplitude of pressure disturbance involved is usually at the order of tenths of atmospheric pressure or in other words, hundreds of times greater than the audible range.
In this case, transient energy exchange mechanisms are of significant importance and based on that motivation, a more comprehensive mathematical model is necessary.

**Pressure-based Aerosol Characterization**

Due to its significance, aerosol characterization is commonly performed in various applications. In environmental study, for example, it is known that the existence of aerosol in atmosphere is one of the key aspects in analyzing Earth’s short-term weather and long-term climate system (Charlson, et al., 1992). Many renowned methods used to characterize aerosol parameters have been discussed by several authors and compiled by Signorell and Reid (Signorell and Reid, 2011). They include Infrared, Raman and UV Spectroscopy, to name a few. Although these methods are capable of analyzing aerosol parameters in detail, down to each individual particle, it is usually measured over an extensive period of time and a large area.

To this end, we propose a possible alternative method to characterize and measure particle parameters in aerosol based on pressure wave attenuation in localized conditions, for example in automotive spray painting setup. The ultimate aim is to characterize the aerosol parameters in real-time based on the numerical prediction of local pressure responses. Subsequently, a sensor for detecting particle sizes and the volume fraction can be built based upon this principle by solving the inverse problem. Figure 1 shows the setup of the proposed apparatus. This apparatus is expected to be portable and lightweight.

The experimental equipment consists of a flow rig with known cross-sectional area and the length of interest. Compressed air, which will be generated by a compressor, is stored in a reservoir. An electronic-controlled solenoid valve is used to control the amount of pressure disturbance to the fluid, is to be energized and turned off, the valve is energized and the air flows for a short period of time. It is sufficient to create a pressure disturbance in the rig.

In order to observe the wave profile, differential pressure transducers are placed along the flow rig. The pressure fluctuation \( P' \) can be measured directly by differential pressure transducers. The initial pulse, which creates the pressure disturbance to the fluid, is to be observed and detected by the transducer at location \( x = 0 \). The output from the differential pressure sensor placed at this position gives \( P'(0,t) \), which serves as the initial conditions to the numerical simulations. The other transducers are then used to monitor the pressure wave profile at different locations. When the tubes are placed according to the locations proposed in Figure 1, it is possible to directly observe/measure the wave profile at any given time and space covered by the rig. From the wave profile, the wave attenuation profile can also be deduced.

The data acquisition tool is used as an interface between the transducers and the control circuitry that performs data logging. The pressure readings given by the pressure transducers in terms of analog voltages are sampled by the data acquisition tool and fed into the control circuit which also stores the logged data in a memory device for further analysis. By comparing the measured wave attenuation profile and simulated wave attenuation profile, which will be discussed in later sections, it is possible to estimate the particle parameters in aerosol (volume fraction and effective particle size).

**Objective**

The objective of this paper is to propose a numerical method to simulate the propagation of pressure wave in gas suspension and to investigate the possibility of applying this work for the development of a novel aerosol characterization technology.

**EQUATION OF PRESSURE WAVE PROPAGATION IN GAS SUSPENSION**

The mathematical model of pressure wave propagation in gas suspension considered in this paper was based on the equation developed by Vakhguelt (Vakhguelt, 2007) and Siswoyo Jo et al. (Siswoyo Jo, et al., 2013) as follows:

\[
\frac{\partial \hat{P}}{\partial \tau} - \frac{\kappa^2 - \kappa^2 + 4\kappa - 2 \hat{P}}{\kappa^2 (\kappa + 1)} \frac{\partial \hat{P}}{\partial \xi} + \frac{9\phi_r}{2 \kappa + 1} \hat{P} - \frac{9\phi_r}{2 \kappa (\kappa + 1)} \frac{\tau_r}{\tau_r - \tau} \frac{\partial \hat{P}}{\partial \xi} \\
- \frac{9\phi_r}{2 (\kappa + 1) \sqrt{T} \tau_r \left( \frac{dP}{dy} \right)_0} \int \left( \frac{\kappa(\kappa + 1)}{2 (\kappa + 1) \sqrt{T} \tau_r \left( \frac{dP}{dy} \right)_0} - \frac{\kappa(\kappa + 1)}{2 (\kappa + 1) \sqrt{T} \tau_r \left( \frac{dP}{dy} \right)_0} \right) dy = \frac{C_i T \tau_r}{\rho_0 \kappa + 1} \frac{\partial}{\partial \xi} \int \left( \frac{\kappa(\kappa + 1)}{2 (\kappa + 1) \sqrt{T} \tau_r \left( \frac{dP}{dy} \right)_0} - \frac{\kappa(\kappa + 1)}{2 (\kappa + 1) \sqrt{T} \tau_r \left( \frac{dP}{dy} \right)_0} \right) dy
\]

(1)

This model is developed by combining continuity, momentum conservation, energy conservation, and the ideal gas law. These equations are expanded to the second-order by considering a small pressure disturbance. Using the technique introduced by Rudenko and Soluyan (Rudenko and Soluyan, 1977), these equations are reduced to a single equation (Equation 1), which is more readily amenable to physical analysis (Borisov, et al., 1980). Descriptions of variables and coefficients used can be found in the Nomenclature section.

Equation 1 is in dimensionless form where \( \hat{P} = P' / P_0 \), \( P = (P - P_0) / P_0 \), \( \hat{x} = x / \lambda \) and \( \hat{t} = t / T \). Here \( P_0 \) and \( \lambda \) are the undisturbed pressure of the gas in its steady state and the wavelength of the initial pressure disturbance. \( P \) is the local gas pressure at the position \( x \) and retarded time \( t \).
The gas suspension considered in this paper is a gaseous medium with suspended particles of uniform radius \( \delta \) and the fraction of volume occupied by these particles in the gas suspension is described by the volume fraction \( \phi_p \).

The pressure wave of finite amplitude propagates through the medium with the speed of sound in positive direction. During its propagation, the pressure wave loses its energy to the suspended particles through momentum and thermal exchanges. The rate of which the pressure wave loses its energy are related to the characteristic times of the momentum exchange process \( \tau_v \) and thermal exchange process \( \tau_T \) and these two parameters depend on the particle size and the physical properties of the gas suspensions (Equation 2).

\[
\tau_v = \frac{\delta^2}{\nu} \quad \tau_T = C_v \left/ \left(8\pi\delta\phi_nC_{pot}\right)\right.
\]

**NUMERICAL SCHEME**

In order to solve the mathematical model of propagating pressure wave in gas suspension, we propose a numerical scheme based on the Crank-Nicolson finite-difference and semi-open trapezoidal quadrature. Due to its non-linear nature, the discretized form of the equation is solved by incorporating a certain root-finding algorithm.

The equation of pressure wave propagation in gas suspension can be re-written by first grouping the coefficients to ease the discretization procedure as shown below:

\[
\frac{\partial \hat{P}}{\partial \hat{x}} - A \frac{\partial^2 \hat{P}}{\partial \hat{t}^2} - B \frac{\partial \hat{P}}{\partial \hat{t}} - C \hat{P}^2 \\
- D \int_0^\hat{\tau} d\hat{\tau} \hat{P} \left\{ \theta_1 \left[ 0, \exp \left(-\frac{\pi^2 \alpha T}{\delta^2} (\hat{\tau} - \hat{\tau}) \right) \right] - 1 \right\} d\hat{\tau} = 0
\]

Finite differencing is applied to the differential terms with the spatial step size of \( \Delta x = 1/Z \), where \( Z \) denotes the number of spatial grids per wavelength and \( i = 0,1,2,\cdots, Z \). The temporal step size is \( \Delta \tau = \Delta y = 1/N \), where \( N \) is the number of spatial grids per wave period.

We employ Crank-Nicolson finite differencing method at the grid point of reference (see Figure 2). This grid point is an imaginary grid point located at \( \hat{x} = i\Delta x \cdot \hat{t} = (j + 1/2)\Delta \tau \). Here, any points at time \( \hat{t} \leq j \Delta \tau \) are known from initial conditions and prior to computational steps. The aim is to solve for the pressure wave profile at the next time step \( \hat{t} = (j + 1)\Delta \tau \).

We shall begin by first discretizing the terms of the equation that are independent of the integral terms. The Crank-Nicolson differencing is applied as follows:

\[
\frac{\partial \hat{P}}{\partial \hat{x}} = \frac{P_{i+1,j} - P_{i-1,j} + P_{i,j+1} - P_{i,j-1}}{4\Delta x} \\
- \frac{P_{i+1,j} - P_{i-1,j} + 4\Delta x}{4/\Delta x} - \frac{P_{i,j+1} - P_{i,j-1}}{4\Delta x} + \frac{P_{i+1,j} - P_{i-1,j}}{4/\Delta x} - \frac{P_{i,j+1} - P_{i,j-1}}{4\Delta x} \\

\frac{\partial \hat{P}}{\partial \hat{t}} = \frac{P_{i+1,j} - P_{i,j} + P_{i,j+1} - P_{i,j-1}}{\Delta \tau} \\
- \frac{P_{i+1,j} - P_{i,j} + 1}{\Delta \tau} - \frac{P_{i,j+1} - P_{i,j-1}}{\Delta \tau} + \frac{P_{i+1,j} - P_{i,j}}{\Delta \tau} - \frac{P_{i,j+1} - P_{i,j-1}}{\Delta \tau}
\]

\[
P_{i+1,j} - P_{i,j} + P_{i,j+1} - P_{i,j-1} = \frac{4}{1/N} \left( P_{i+1,j} + P_{i,j} + P_{i,j+1} \right)
\]

To carry out the computation, the known terms and unknown terms have to be separated. The terms \textit{Basset} and \textit{W} are introduced to make this possible.

\[
\textit{Basset} = \left\{ \begin{array}{l}
- \frac{1}{2} \sqrt{2} P_0, \quad \text{for } j = 0 \\
\frac{1}{4} \left( P_{i+1,j} - P_{i,j} + P_{i,j+1} - P_{i,j-1} \right) + \frac{1}{2} \sum_{k=1}^{i+1} P_{k,j} - P_{k,j-1} + \frac{1}{2} \sum_{k=1}^{i+1} P_{k,j+1} - P_{k,j}, \quad \text{for } j > 0
\end{array} \right.
\]

\[
\textit{W} = \left\{ \begin{array}{l}
\frac{1}{2} \sqrt{2}, \quad \text{for } j = 0 \\
\frac{3}{8} \sqrt{2}, \quad \text{for } j > 0
\end{array} \right.
\]

\[
I = \sqrt{N} \left( \text{Basset} + WP_{i,j+1} \right)
\]
The same treatment is applied to the heat exchange term:

\[ H = \frac{\partial}{\partial \hat{t}} \int_{0}^{\hat{t}_{j+1/2}} \left[ \sum_{i} \left( \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (\hat{t}_{j+1/2} - \hat{y}) \right) \right] - 1 \right) d\hat{y} \right] = \Delta \tau \psi(k) = \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (k \Delta y) \right) \right] - 1 = \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (k) \right) \right] - 1 \]

\[ \psi(k) = \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (\hat{t}_{j+1} - \hat{y}) \right) \right] - 1 \]

\[ H_{\text{now}} = \int_{0}^{\hat{t}_{j+1}} \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (\hat{t}_{j+1} - \hat{y}) \right) \right] - 1 \right) d\hat{y} = \Delta \tau \psi(k) = \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (k \Delta y) \right) \right] - 1 = \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (k) \right) \right] - 1 \]

\[ H_{\text{prev}} = \int_{0}^{\hat{t}_{j}} \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (\hat{t}_{j} - \hat{y}) \right) \right] - 1 \right) d\hat{y} = \Delta \tau \psi(k) = \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (k \Delta y) \right) \right] - 1 = \partial_{j} \left[ 0, \exp \left( -\frac{\pi^{2} a}{\delta^{2}} (k) \right) \right] - 1 \]

Now we introduce terms HeatNow, HeatPrev and Y to isolate the known and unknown variables.

\[ \text{HeatNow} = \frac{1}{4} \sum_{i} P_{i,j} \psi(j + 1) + \frac{1}{4} \sum_{i=1}^{j} \left( P_{i,j+1} + P_{i,j} \right) \psi(j + \frac{1}{2} - k) + \frac{1}{2} \sum_{i=1}^{j} P_{i,j} \psi(j + 1 - k) + \frac{1}{8} \sum_{i=1}^{j} P_{i,j} \psi(j) \]

\[ \text{HeatPrev} = \frac{1}{4} \sum_{i=1}^{j} \left( P_{i,j+1} + P_{i,j} \right) \psi(j + \frac{1}{2} - k) + \frac{1}{2} \sum_{i=1}^{j} P_{i,j} \psi(j - k) \]

\[ Y = \frac{3}{8} \psi(j) \]

\[ H = \text{HeatNow} - \text{HeatPrev} + Y P_{i,j+1} \]

The final discretized form of the equation is written as:

\[ K_{i} \psi + \alpha P_{i-1,j+1} + \beta P_{i,j+1} + \gamma P_{i+1,j+1} + \Phi P_{i,j+1} = 0 \]

The Jacobian matrix (matrix of partial derivatives) of the non-linear system is also needed, which is constructed as such:

\[ \beta_{i} = \frac{B}{2} P_{i,j} - \left( \frac{DW + E}{2} P_{i,j} - \frac{E}{2} \text{Basset} \right) \sqrt{N} - F \times Y \]

By applying boundary conditions \( P_{0,j+1} = P_{Z,j+1} = 0 \), we can construct the system of non-linear equations \( S \) from \( i = 1 \) to \( Z - 1 \).

\[ S = [S_{1}, S_{2}, \ldots, S_{Z-2}, S_{Z-1}]^{T} = 0 \]

The system of non-linear equations can be solved by employing the Newton-Raphson Method. This method approximates the roots of the system of equations by successively minimizing the differences between the approximated and the real roots (Press, et al., 2007). The algorithm begins by specifying the initial guess to the roots, \( \hat{P} \). We use the wave profile from time step \( \hat{t}_{j} \) to be the initial approximation of \( \hat{t}_{j+1} \) due to the fact that the wave profile must evolve from previous time step. To proceed, the Jacobian matrix (matrix of partial derivatives) of the non-linear system is also needed, which is constructed as such:

\[ \frac{\partial S_{1}}{\partial P_{1}} \quad \frac{\partial S_{2}}{\partial P_{2}} \quad \ldots \quad \frac{\partial S_{Z-1}}{\partial P_{Z-1}} \]

\[ \frac{\partial S_{2}}{\partial P_{2}} \quad \ldots \quad \ldots \quad \ldots \quad \frac{\partial S_{Z-1}}{\partial P_{Z-1}} \]

\[ \frac{\partial S_{Z-1}}{\partial P_{Z-1}} \quad \ldots \quad \ldots \quad \ldots \quad \frac{\partial S_{Z-1}}{\partial P_{Z-1}} \]

\[ J = \begin{bmatrix} \frac{\partial S_{1}}{\partial P_{1}} & \frac{\partial S_{1}}{\partial P_{2}} & \ldots & \frac{\partial S_{1}}{\partial P_{Z-1}} \\ \frac{\partial S_{2}}{\partial P_{1}} & \frac{\partial S_{2}}{\partial P_{2}} & \ldots & \frac{\partial S_{2}}{\partial P_{Z-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial S_{Z-1}}{\partial P_{1}} & \frac{\partial S_{Z-1}}{\partial P_{2}} & \ldots & \frac{\partial S_{Z-1}}{\partial P_{Z-1}} \end{bmatrix} \]
With the inclusion of an initial guess, it can be shown that
\[ \mathcal{J} = J(P) \]
\[
\begin{bmatrix}
\beta_1 + 2\alpha P_1 & \gamma & 0 & \cdots & 0 \\
\alpha & \beta_2 + 2\alpha P_2 & \gamma & \cdots & \\
0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \alpha & \beta_{Z-2} + 2\alpha P_{Z-2} & \gamma & \\
0 & \cdots & \alpha & \beta_{Z-1} + 2\alpha P_{Z-1} & \\
\end{bmatrix}
\]
\[ \mathcal{S} = S - \bar{P} \]
\[ = K_1 + \beta_1 P_1 + \gamma P_1^2 + \Phi P_1^2 \\
K_2 + \alpha P_2 + \beta_2 P_2 + \gamma P_2^2 + \Phi P_2^2 \\
\vdots \\
K_{Z-2} + \alpha P_{Z-2} + \beta_{Z-2} P_{Z-2} + \gamma P_{Z-2}^2 + \Phi P_{Z-2}^2 \\
K_{Z-1} + \alpha P_{Z-1} + \beta_{Z-1} P_{Z-1} + \Phi P_{Z-1}^2 \\
\]
\[ \mathcal{J} \mathcal{A} = -\mathcal{S} \]
\[ \mathcal{A} = (\mathcal{J}^{-1})(-\mathcal{S}) \]
\[ \bar{P} = P + \mathcal{A} \] (17)

The final step of this method is to compute the corrections to the current approximated roots, which are the elements of the deviation vector, \( \mathcal{A} = \bar{P} - P \).

\[ S + J \mathcal{A} = 0 \]
\[ \mathcal{J} \mathcal{A} = -S \]
\[ \mathcal{A} = (\mathcal{J}^{-1})(-S) \]

The simulation results clearly indicate that the radius and the volume fraction of aerosol droplets have a significant impact on the attenuated pressure wave profiles. All these attenuated profiles do not overlap each other as depicted in the figures, which is an encouraging sign to “sense” the droplet sizes and volume fraction in the suspension. An inverse problem can be solved by measuring the local pressure distribution, which can then be matched to the numerical wave profile predicted by the simulation. A sensor for detecting droplet parameters can be built upon this concept.

**RESULTS AND DISCUSSIONS**

Using the technique presented in the previous section, we performed several simulations to investigate the propagation of pressure wave in a gas suspension. The gas suspension chosen was air-water gas suspension (air as the gas phase and water as the suspended phase). The initial temperature of the gas suspension is 20°C. The properties of water and air are evaluated at the initial temperature, which can be obtained from the textbook by Çengel and Turner (Çengel and Turner, 2005). The initial pressure disturbance is an inverted cosine pulse with an amplitude of 0.3P<sub>j</sub> with period T=1s.

Figure 3 shows the evolution of the pressure wave profile for the droplet radius \( \delta = 60 \mu m \) and volume fraction \( \phi_p = 1.2 \times 10^{-4} \). It can be seen that the pressure wave is attenuating as it travels through the medium. The steepening of the wave profile is expected due to the non-linear interactions between the two phases.

The dissipation of wave energies are shown in Figures 4 and 5 in terms of the relative wave energy density, \( \Delta \Delta / \Delta \Delta \), \( \Delta \Delta \) corresponds to the initial wave energy density. The wave energy was calculated by computing the mean square pressure value at a given time \( \hat{\tau} \). The dimensionless retarded time from \( \hat{\tau} = 0 \) to \( 5 \times 10^3 \) corresponds to the real distance of propagation from \( x=0 \) to 1.7156m given that the speed of sound in air at 20°C is 343.12m/s. Figures 6 and 7 show the dissipations of wave energies with respect to the real distance of propagation \( x \) in meters.

Figure 4 shows the wave energy dissipation for a fixed droplet radius of \( \delta = 60 \mu m \) and different cases of volume fraction. By successively doubling the volume fraction from \( 6 \times 10^{-7} \) to \( 4.8 \times 10^{-4} \), it can be observed that the wave energy dissipated at faster rate.

Another set of simulations was performed for a fixed volume fraction \( \phi_p = 1.2 \times 10^{-4} \) and various droplet sizes (see Figure 5). Results show that smaller droplet sizes give rise to faster dissipation of wave energy and it is consistent with Equations 1 and 2, as characteristic times of the exchange process increase with the decrease in the particle radius. This allows the passage of the pressure wave to interact longer with the suspended particles. It is interesting to note that varying the droplet radius not only changes the dissipation rate, but also the behavior of the dissipation. This effect is most significantly noticeable in the last term of Equation 1, as the contribution of the thermal exchange process is exponentially dependent on the droplet radius.

The simulation results clearly indicate that the radius and the volume fraction of aerosol droplets have a significant impact on the attenuated pressure wave profiles. All these attenuated profiles do not overlap each other as depicted in the figures, which is an encouraging sign to “sense” the droplet sizes and volume fraction in the suspension. An inverse problem can be solved by measuring the local pressure distribution, which can then be matched to the numerical wave profile predicted by the simulation. A sensor for detecting droplet parameters can be built upon this concept.

**CONCLUSION**

The evolution of wave profile and the dissipation of wave energy are generated using the numerical techniques described in the earlier section. Using these results, we can observe the behavior of wave propagation in air-water gas suspension.

The proposed numerical simulation method can be applied to particle characterization technology discussed in previous section. By applying a known pressure disturbance to the aerosol and continuously sampling the pressure data in the medium, the recorded dissipation profile of the pressure wave can be compared to the
simulated dissipation profile to characterize the effective aerosol parameters, the volume fraction $\phi_v$ and particle radius $\delta$.

Although the current model is capable of estimating the particle parameters, there are several limiting conditions that have to be considered. First, the current model requires either one of the volume fraction $\phi_v$ or particle radius $\delta$ to be known. This is acceptable in some cases, for example in characterizing aerosol parameters in spray painting, the dispensed volume of the paint is known and hence the volume fraction is determinable. Second, the presence of the suspended particles must be significant enough over the length of the sensor rig in order for the pressure wave dissipation to be detectable. However, this limitation might be minimized by choosing the suitable initial disturbance depending on the given aerosol parameter.

The presented model is a forward problem in which parameters are known a priori. For characterization purpose, an inverse problem involving two governing parameters needs to be solved considering various measured wave profiles.

REFERENCES


FIGURES AND TABLES

![Fig. 1](image1.png)

**Fig. 1** The proposed measurement rig for aerosol characterization

![Fig. 2](image2.png)

**Fig. 2** The computational grid setup used in simulating the evolution of pressure wave profile in gas suspension
Fig. 3  Evolution of the pressure wave profile for volume fraction $\phi_p=1.2\times10^{-4}$ and droplet radius $\delta=60\mu$m

Fig. 4  Energy densities of the pressure wave for droplet radius $\delta=60\mu$m

Fig. 5  Energy densities of the pressure wave for water volume fraction of $\phi_p=1.2\times10^{-4}$

Fig. 6  Energy densities of the pressure wave for droplet radius $\delta=60\mu$m (distance in meters)

Fig. 7  Energy densities of the pressure wave for water volume fraction of $\phi_p=1.2\times10^{-4}$ (distance in meters)
A FLEXIBLE DESIGN FOR SYSTEM CONTROL AND COMMUNICATIONS IN A ROBOTIC MOBILITY ASSISTANT FOR HANDICAPPED PERSONS

Stephan Irgenfried1, Johannes Schneider2, Heinz Wörn1, Wilhelm Stork2, Tim Guhl3, Matthias Schellhase3
1 Institute of Anthropomatics and Robotics IAR
2 Institute of Information Processing Technologies
Karlsruhe Institute of Technology KIT, Germany
3 KUKA Laboratories GmbH
{stephan.irgenfried, johannes.schneider3, woern, wilhelm.stork}@kit.edu
{tim.guhl, matthias.schellhase}kuka.com

ABSTRACT

In this paper we describe a flexible design for system control and communication in a robotic mobility aid, proof-of-concept in a prototype system and experimental results achieved within a multi-day testing session with a group of potential users. The mobility aid combines the functionalities of a walking aid with active fall prevention and a sit-to-stand transfer assistant in a single device, having a smaller footprint than standard walking aids. We focus on the system communication based on a middleware and describe, how this concept introduces robustness and flexibility to assisting devices like this and how the system can easily be extended with additional functionalities by adding new hardware or software. As examples for additional functionality, health monitoring, entertainment and communication applications are described.

KEYWORDS

Mobility Aid; Ambient Assisted Living; Sit-to-stand transfer; Walking Aid; Demographic Change; Robotics; Middleware; Gait Analysis; Geriatrics; Senior Citizens

INTRODUCTION

As the average expectation of life is increasing in many countries in the world, the absolute number of people attaining a great age raises year by year. Being able to live a self-determined life is one of the key expectations of this group of people just as it was to them in the earlier decades of their life. Therefore, preserving mobility in their everyday life is of high demand. (Alwan et al., 2007)

Many industrialized countries are facing a demographic change. With stagnating or falling birth rates, the proportion of retired people with respect to the population overall increases continuously, implicating a shortage of caregivers and hence the need for technical solutions to allow the elderly to stay in their familiar surroundings and avoiding a move to a care home. Technical solutions can help to compensate mobility restrictions, either temporal, e.g. because of a surgery or accident, or permanent, e.g. because of arthrosis or muscle weakness. While passive support tools like canes, crutches or passive walking aids are state of the art, active mobility aids are still under research in the robotic as well as in the medical scientific community. In addition to basic functions like walking or transportation support, such devices can have many additional functionalities like fall prevention, fall detection, health monitoring, liquid intake monitoring, communication, entertainment, indoor- and outdoor-navigation, medication reminder and many more.

A system built of many different hardware and software components requires a robust and efficient communication infrastructure as well as reliable communication protocols. In this paper we describe the system concept and middleware-based component communication infrastructure of an active mobility aid, which supports walking as well as the sit-to-stand transfer, in addition to the work published in (Irgenfried and Wörn, 2014), which focuses on real-times system control by using input signals of a force-torque-sensor.

The rest of the paper is organized as follows: first we introduce the electromechanical system design on a high level basis and describe the core functionalities. Following this chapter the concept and implementation of device communication based on a middleware software is presented. Finally we describe the prototype system implementation and first experimental results from lab and field trials. The paper finishes with a conclusion and an outlook.

RELATED WORK

A survey paper on the current state of research on mobility aids was published by (Dune et al., 2012), which discusses many of the ancestors of our system, like
PAMM, GUIDO or the I Walker. Our work was inspired by the previous works of (Médéric et al., 2004) and (Merlet, 2012). In their work, (Rea et al., 2013) the authors describe a general approach on how to design sit-to-stand assisting devices based on supporting the natural motion of human body with a technical device. The work of (Alwan et al., 2007), (Frizera Neto et al., 2010) and (Wasson et al., 2003) is focused on deriving gait characteristics from a force-torque-signal using passive type walking aids. Adding indoor navigation functionality to an active walking aid is described by (Rodriguez Losada et al., 2005).

**SYSTEM ARCHITECTURE**

As an introduction to the system, we give a short overview on the architecture and the different layers, shown in Fig. 1. On the base layer, we have the sensors and actors in the system. They are connected using standard industry interfaces like e.g. EtherCAT to the low level system control, based on the real-time operating system VxWorks\(^1\) or, in case of non time-critical sensors, directly to the middleware. This multi-layer approach separates concerns inside the system, introducing flexibility in adding new devices and functionalities while ensuring that motion control and security functions, e.g. collision avoidance or fall prevention, are executed meeting hard real-time requirements.

To have a standardized and flexible way of device communication inside the system, we introduce a middleware-layer to the system architecture. This enables high-level applications and the human-machine-interface to receive sensor and system information or to send data to the low level system, without having to implement low level communication protocols. The middleware distributes message following the publish-subscribe pattern, which is explained in greater detail later. The human machine interface makes uses of this infrastructure extensively to collect data from sensors and system control to be displayed on the user interface and to allow the user to give commands to the system using haptic or verbal interaction.

One of the main benefits of this system design is the possibility of adding software functions or services to the system, independently from where they are executed. As described later, only a network connection to a middleware component is required to have access to the underlying communication infrastructure. The middleware also manages access rights to meet privacy criteria.

**CORE FEATURES**

The system functions for walking and sit-to-stand support are described in greater detail in (Irgenfried and Wörn, 2014), so here only a short overview on those two is given.

**Walking Aid with Active Fall Prevention**

The first core feature of the mobility assistant is to be an active, motor-driven walking aid with active fall prevention. Making use of a force-torque-sensor which connects the handle bars with the mechanical body of the system, the user motion intention is detected and motor control signals are calculated within the real-time control.

By classifying the user input command signals, an approaching fall situation can be detected and the active fall prevention smoothly stops the system to give a strong hold to the user and reduce the risk of fall. Fig. 2 shows the user with the system in walking aid mode.

**Sit-to-stand transfer**

The second key feature of the device is to support the sit-to-stand transfer of the user. The system only supports the natural motion of the user, compensating missing muscle strength. The trajectory for the motion of the user induced by the system, is similar to the one described by (Rea et al., 2013). Using the force-torque-sensor signal, the system detects the status of the user body by tracking the center of gravity and stops the stand-up motion, in case the user is about to fall. Fig. 3 shows the user using the system in sit-to-stand transfer support mode.

**Device communication using a middleware**

Using a JAVA-based middleware, all sensors and actors in the system can communicate in a standardized way with low communication overhead and high reliability. The performance level is independent from whether it is connected to the system bus and real-time OS or directly to the middleware. In the following, the middleware architecture and implementation details are presented.

**MIDDLEWARE**

**An Event Based Middleware**

In order to have a flexible and scalable platform, which is also opened to third party suppliers, we placed a middleware layer on top of the hardware and real time system layer. The middleware allows a continuous, dynamic and just in time integration of new functionalities, services and hardware components through well-defined interfaces.

It works as a software bus that enables an event based asynchronous, one-to-many and loosely-coupled push communication between all components. The

---

\(^1\) [www.windriver.com](http://www.windriver.com)
communication is based on the publish-subscribe pattern, a well-known messaging pattern. Senders of messages, called publishers, publish messages without knowing whether anyone is interested. Similarly, receivers of messages, called subscriber, announce their interests in messages, without prior knowledge of publishers. So instead of using a direct communication between components, publishers deliver the messages to a message broker that routes the messages to interested subscribers. Using this pattern offers the possibility of decoupling in time, space and synchronization between message senders and receivers. (Wang et al, 2010), as shown in Fig. 4.

Our implementation is based on the Java OSGI-Framework (Open Service Gateway Initiative)\(^2\), a widespread middleware framework (Wolf et al, 2011; Hanke et al, 2011; Eichelberg et al, 2010) that implements a dynamic component model. The OSGI platform allows the integration and modification of components, called bundles, during system runtime and solves dependencies between different components. Furthermore it offers a natural support of dependency injection and enables event-based communication.

**Middleware Components**

To facilitate the integration of components, we followed the architecture of xAffect\(^3\) (Schaaff et al, 2012), a framework for biosignal processing, and defined three categories of component interfaces that are able to connect to the middleware:

**Data Sources** represent pure publishers. They are able to publish messages to one or more topics but are not able to receive any. This component interface can be used to acquire and publish data from sensors, for instance.

**Data Sinks** represent pure subscriber. In contrast to data sources, data sinks can receive but not publish messages. As an example, this category is used to implement components for logging or visualization of sensor data.

**Data Processors** combine the functionalities of sinks and sources. They can receive messages, process them and publish the results as new messages. For example, data processors can be used to capture higher contexts by merging data from different data sources.

New components can be easily developed by extending these predefined JAVA classes. Super classes manage the connection to the middleware and the inter-component communication. Furthermore, these classes offer access to a central object data base (ObjectDB)\(^4\) where messages can be saved and reloaded. Moreover, to ensure a fast and parallel processing, every component runs in an independent thread.

In addition, all components are supervised by a watchdog. It checks:
- CPU load for each component
- Time required by receivers to process a message
- Publishing rate of message senders

If predefined thresholds are exceeded, the watchdog shuts down and reloads the malfunctioning component.

**Middleware Messages**

To guarantee a smooth communication, we defined a message format compatible with every component. Besides the payload data, which could be any JAVA object, the message format also contains detailed information about the type of the signal or event and the message sender. In addition, the payload can be encrypted using the DES (Data Encryption Standard) symmetric-key algorithm (FIPS PUB 46, 1977). Only receivers that are in possession of the key are able to decrypt the message. This characteristic enables a safe communication between components despite the publish-subscribe pattern.

Referring this main format, we implemented two categories of messages:

**Data Events** for inter-component communication.

**Message Events** are defined to interact with the user. If a component publishes a Message Event, the message is forwarded to the user interface, for example to a GUI (Graphical User Interface) or a text to speech engine.

**Implemented Tools and Components**

Based on the architecture mentioned above, we developed and integrated the following components:

The **Bus Management Tool**, shown in Fig. 5, is a GUI based development tool that visualizes useful middleware information, e.g. states of all components, amount of message traffic, topics of published messages or watchdog information. Moreover events can be visualized by integrated signal viewers or generated by signal generators. Through it, developers can set up closed test environments to simulate and test components.

The **Robot Communicator** connects the middleware layer to the real time system layer and offers access to the robot control. Middleware components are able to collect data from robot sensors and can set requests to influence actuators.

---

\(^2\) www.osgi.org

\(^3\) www.xaffect.org

\(^4\) http://www.objectdb.com/
The **App Communicator** connects the middleware via WiFi to the graphical user interface that is implemented as an Android application running on a tablet computer. The application can act as data source as well as data sink. Message Events are shown as dialogs on the tablet screen and read by a text to speech engine.

In order to extend the multimodal user interface, we implemented a **speech recognition module**. The module uses the .NET Framework and is built on the Microsoft Speech SDK. It implements a continuous speech recognition that transmits recognized speech commands to a middleware component via TCP/IP. The middleware component provides this information to other components by publishing the speech command as a data event on the software bus. Fig. 6 gives an overview on the architecture of the speech recognition component.

**USER INTERFACE**

Since the main target user group of our mobility aid is elderly people, we have to adapt the user interface to the cognitive and haptic skills of seniors. As interaction device, we use a Samsung Galaxy Tab2 10.1 that comes with the Android operating system. A tablet device offers the possibility of multimodal user interaction using graphics, sound, speech and touch. In addition, the Android platform has a big community and plenty of apps that are available in the Google Playstore. Consequently, besides the middleware, this offers a second extendable platform that allows the integration of additional services.

The objective is to advance the user interface from a pure robot control interface to a multimodal, scalable and customizable interface that works as an intuitive robot control as well as social interaction platform or multimedia device.

**An age friendly graphical user interface**

To hide the complexity of Android OS from the user, we implemented an App that works in locked kiosk mode. The App seems to be an operating system on top of the Android system by consistently hiding the Android status bar. Only administrators can to exit the application and return to the default home screen. Furthermore, the application connects automatically via WiFi to the middleware, which runs on the robot platform. Consequently, the application is able to receive and publish events on the software bus. It enables the control of the robot as well as data communication with other middleware components. As mentioned, message events, published by any middleware component, are shown on the tablet screen and read aloud by a text to speech engine. For future evaluation of the user interface, every interaction between user and GUI can be logged in background. Based thereon, we want to evaluate the usability in order to improve the age-friendly design.

Fig. 7 shows the main screen of the implemented app concept. In order to satisfy the requirements of an age-friendly user interface, we followed these design guidelines:

- High contrast
- Big buttons (min 15mm²) that simulate the pressing of real buttons by using force feedback, appropriate sounds and visual effects
- Adaptable font size
- Speech output for important information
- Meaningful symbols in combination with text
- Limited information content per screen (fewer than 10 interaction elements, if possible)
- Continuous repetitive structure concerning screen layout, coloring and forms

**Implemented App Features**

Referring the introduced app design, we implemented services that advance the user interface from a simple control interface to a multifaceted interaction platform. This contains communication (phone and email), multimedia (picture and video viewer, music player and radio), games to improve cognitive and haptic skills, calendar, newsfeed and a weather forecast service. Moreover, we integrated a health management service and connected a blood pressure monitor via Bluetooth to the app. The service records new measurements automatically and creates a blood pressure diary.

Most of the services are synchronized to a cloud account and can be supervised remotely via a web interface. This reduces the complexity of the app and, at the same time, enables relatives to share pictures, videos or music with the elderly users or to support them in the management of their contacts and appointments.

**EXPERIMENTS**

**Lab test**

Fig. 8 shows a test session in the lab. The user wears an age simulation suit which limits agility to give its carrier the impression how task execution may feel from the perspective of elderly people.

To simplify the change of the parameters of the sit-to-stand transfer support mode at runtime, we developed a JAVA application, which makes use of the middleware communication infrastructure. Using this tool, the relevant parameters of system motion can be controlled and changed at runtime, e.g. forward speed, bottom-to-top time, delay times. Fig. 9 shows the GUI we used during the lab trials.
Field trials

The prototype system was evaluated in an assisted living environment in the south of Germany. A total number of 10 users, from the potential target group took part in the tests. Table 1 summarizes the statistical data of the test group.

Every user tested the system for approx. 1 hour in a real housing environment, shown in Fig. 10. Everyday tasks like walking around, stand-up and sit-down, transport of goods and visiting the toilet were executed. The user feedback was recorded with help of a structured interview. User feedback showed that the systems high stability was appreciated and the system was considered to be a trustful companion. A second aspect uttered by the users was their surprise about the mobility assistant being as agile as a standard walking aid despite its’ greater mass. Detailed description of the trials and the results will be published in another article.

CONCLUSION

In this paper we presented our work of implementing a communication infrastructure based on a JAVA middleware in an active mobility assistant for handicapped persons. We showed that this system architecture introduces flexibility and extendibility to this kind of assisting devices, but the idea itself is not limited to this type of devices. Furthermore, we showed that the combination of a real-time OS for system control and security and non-real-time applications can share a common communication infrastructure and that the implementation of the system architecture in a prototype was successfully evaluated in lab and field environment.

FUTURE WORK

Based on the field trial results with the potential group of users, the system currently undergoes a re-design of the electro-mechanical setup. The new system with two separate handlebars reflects the user feedback to make the system look more agile and more similar to a conventional rollator while keeping the additional functionality. Some more work needs to done on how the system design can help rising acceptance. System mode switching (fixed stand, sit-to-stand, walking) needs to be improved by adding context-awareness to make the system recognize the user wish to change modes without manual interaction like pressing buttons. The next generation prototype system will again be evaluated in lab and field.

ACKNOWLEDGEMENTS

The authors would like to acknowledge support from the German Federal Ministry of Education and Research project MAID: Mobility Aid for Handicapped Persons (16SV5815) as part of the program ICT 2020. We also thank our additional project partners in the consortiumMetraX and Evangelische Heimstiftung. We appreciate the work of the students who volunteered for the experiments.

REFERENCES


Schaaff, K., Müller, L., Kirst, M., Heuer, S., 2012, „xAffect – A Modular Framework for Online Affect Recognition and Biofeedback Applications”, 7th European Conference on Technology Enhanced Learning (ECTEL 2012), MATEL Workshop, Saarbrücken, Germany

Wang, T., Xu, W., He, J., Chen, R., Gu, W., 2010 „A Brief Survey of Event-based Middleware”, 2nd International Conference on Computer Engineering and Technology


Wolf, P., Schnidt, A., Otte, J., et al., 2011 „openAAL - the open source middleware for ambient-assisted living (AAL)” AALIANCE conference- Malaga, Spain

FIGURES AND TABLES

Fig. 1: System architecture

Fig. 2: Walking aid, user wears an age simulation suit

Fig. 3: Sit-to-stand support

Fig. 4: Publish-Subscribe Pattern
Fig. 5: Screenshot of the Bus Management Tool showing the ECG signal received from the sensor

Fig. 6: Speech recognition component overview

Fig. 7: Main screen of the user interface

Fig. 8: Sit-to-stand support lab test

Fig. 9: Configuration interface
Fig. 10: Field trial involving the target user group

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>10</td>
</tr>
<tr>
<td>Average age</td>
<td>83</td>
</tr>
<tr>
<td>Min. /Max. age</td>
<td>70 / 90</td>
</tr>
<tr>
<td>Female / Male</td>
<td>9 / 1</td>
</tr>
<tr>
<td>Insecure gait, passive assistive device used</td>
<td>8</td>
</tr>
<tr>
<td>Insecure gait, no passive assistive device used</td>
<td>1</td>
</tr>
<tr>
<td>Problems with sit-to-stand</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1: Test group statistics
INTELLIGENT SELF SUSTAINABLE GREEN ENERGY POWER SUPPLY MODEL FOR RURAL NETWORK ACCESS & HOME APPLIANCES

Siva Raja Sindiramutty¹, Chong Eng Tan²,
Faculty Computer Science and Information Technology,
University Malaysia Sarawak (UNIMAS)
94300 Kota Samarahan, Malaysia.
magan.shiva91@gmail.com¹, cetan@ieee.org²

ABSTRACT

The purpose of this paper is to explore and perform a study on intelligent self sustainable power supply model based on the green energy sources integrated with power generator in one of the most remote area in Sarawak, Bario. The end user equipment deploy in rural areas, have faced issue on power provision as well as power supply which is not consistent. A successful experimental demonstration on the adoption of green energy solution was needed to serve an uninterrupted power supply for highly demanded hardware especially network access equipment. Thus, the proposed power supply will give a big impact in bridging the digital divide in Bario. Intelligent power supply was integrated with the current solar panel system which has been already deployed. The proposed model was composed with effective power supply architecture for end user Wifi network and home appliances to the remote areas. The intelligent power supply model has the intelligence in self sustainable and its nature of modularity enables it to integrate with diesel generator. The intelligent power supply follows the user usage pattern for better use of energy in the end user network access equipment and home appliances.

Keyword – e Bario, Telecenter, rural areas, conventional power supply model, VSAT, critical appliances, non critical appliances

INTRODUCTION

The data or digital age came into its own with the creation of the World Wide Web (WWW) in the year 1989, and gradually after, the internet changed into a global network. Now the internet is commonly recognized as a global platform to fasten the flow of knowledge. Nowadays people more involve in intellectual rather than before because of the presence of internet [1]. Anyhow, the internet utility does not equally exist to be exploited by everyone. There are digital divided gaps between individuals, households, businesses and geographical areas with taking into account to their opportunities to access ICTs and the use of the internet. [2]. In other words, different groups of people have different levels of access and abilities in using the internet especially people in urban and rural area. A total of 85% of the population of the less developed countries and 75% of Asian developing countries live in rural areas. The accessibility and availability of internet facility is even lower in rural areas than the urban areas of the developing countries [3]. Bridging the digital divide at the rural site is always a difficult task as the adoption of modern communication technologies will always be confined due to deficiency of local supporting resources such as power supply, communication gateway connectivity and remoteness of the rural communities [4]. Issue in provision of basic utilities such as power supply and the effort in bridging the digital divide in the remote areas can be several times more costly than in urban area. The power provision is important and acts as the first criteria in order to implement any higher level communication to take place [4].

Telecenter project have been initiated as one of the important plan in bridging the digital divide across the country where there are more than 2000 telecenters that have been implemented to date [5]. Wireless solution has been suggested to be cost effective solution for remote network provision in India’s project of Digital Gangetic Plains (DGP) and Ashwini project. There are five vital aspects in the use of the wireless for rural people which is, network planning and deployment, network management and operations, network protocols, power savings, applications and services that are being discussed. Due to the low paying capacity of user and the low population density in rural areas, deployment of ICT in rural areas is less or does not exist. Thus the cost effective wireless equipment with license free channel is a better choice of network deployment option. DGP project in India is researching the technical possibility of creating long range 802.11 links using high gain directional antennas to connect up to many villages which is surrounding through multi hop mesh network architecture [6].

BACKGROUND

Malaysian government’s pilot project on bridging the digital divide in Bario, known as the eBario project, was initiated by the Centre of Excellence for Rural Informatics (CoERI) from Universiti Malaysia Sarawak (UNIMAS) in the year 1999. Telecenter project has allowed Internet access to one of the most rural community in Malaysia since then. The telecenter project in Bario which has been developed with satellite communication through Very Small Aperture Terminal (VSAT) system, has established a gateway for communication
to the Bario people and the outside world. Since then, Bario people is able to make call to the rest of the world through telephone calls and Internet. The Bario telecenter only act as centralized Internet access for people from nearby villages. People who stay there need to put in a lot of effort to travel few kilometers or travel about 2 hours to reach to the centralized telecenter to use the telecommunication service. The overall cost to develop a telecenter is high. Thus, it is impossible to build another telecenter that are rather closer to each other. Due to this issue, there is a need for a telecenter service to reach to other villages to bring more impacts [4].

**Sustainable Energy Efficient Long Range Wireless Network**

The self-sustainable long range wireless relay that was proposed in previous research [4] has been proven to be effective for the purpose of extending the Internet coverage from the rural telecenter in Bario. The long range wireless relays provide solution for self-sustainable network coverage in order for the users to get connected. The purpose of self-sustainable energy efficient long range wireless network model is to connect nearby villages within 10km radius of distance from the telecenter. The importance of this model are cost effectiveness, self-sustainable and energy efficient long range wireless network which can be easily installed in rural area for the aim of extending the internet access network of the telecenter. The wireless network will enable to overcome the trouble of putting cable along the mountain and thick jungle easily. The proposed model been created based on the rural environment restrain and has been deployed successfully in furthering the Bario telecenter network. The whole wireless connectivity operations are being self sustained through the green energy initiative which is solar energy. This technology has been connected up to six nearby villages in Bario since 2010 [4]. The proposed model is implemented based on the multi hop wireless network topology. This module is the main component of the proposed network model because it relay the wireless connectivity from the internet gateway which is located at Telecenter with broadband VSAT connection.

Self sustainable long range wireless network is a multi hop nature and it can be installed almost at any place with a good Line Of Sight path from the internet gateway with less space occupancy for implementation. It can be installed at high ground such as on small hill top or on the roof of houses to relay the connection to the nearby villages or long houses [7]. This model is not just a wireless relay backbone for furthering the connection from one point to another point but it also act as wireless access point for the end station. The self sustainable long range wireless network model contain directional antenna which relay the connection up to 10 km to another self sustainable long range wireless network model to extend the network. At same time, it also has omidirectional antenna which broadcast the wireless network signal to the nearby villages up to 3 km in distance [4][7]. The proposed model was designed together with small solar panel of ten of watts and attached with a small battery capacity of tens of Amp Hours (AH), which is enough to power up the proposed model. It was invented good enough to sustain the aimed equipment for constant operation for several days even during minimum sunlight condition. The nature of the power supply was designed to optimize the use of the cultivated solar energy and to minimize power wastage [4][8].

**Energy Efficient Adaptive Bandwidth on Demand for Long Range Wireless Relay**

Self Sustainable Energy Efficient Long Range Wireless Network somehow gives problem. People in Bario reported that they experienced slow network as well as the issue of power consumption of wireless equipments because of the unpredictable weather. The installed model relies on power supply from the solar panel in order to operate at all time. Due to this issue, enhancing the current model of sustainable energy efficient long range wireless network is needed. The best solution is monitor and manages the network traffic intelligently, in order to attain an energy efficient adaptive network. The energy efficient adaptive bandwidth on demand for long range wireless relay model used in the current model will be maintained. A two link wireless relay will be made for bandwidth aggregation for the purpose of handling the network efficiently by allowing the additional links during peak hours period and handle the heavy incoming traffic. The energy efficient adaptive bandwidth on demand is able to overcome the insufficient bandwidth by catering different traffic demand at different period of time of the day and to intelligently regulate the power consumption of wireless equipment for energy efficiency purpose [9].

**PROPOSED MODEL**

The long range wireless relay point proven to be self sustainable in Bario but it is only providing the wireless network coverage for the villages. The wireless network access equipment needs to be deployed at the end user premises in order for the network access to be completed. To bear big number of users distributed around villages and long houses, the end user wireless access equipment has to be deployed according to the distribution of users’ premises which can be far apart. The installation of the wireless equipment is pretty much straight forward as they are mobile device but what is immobile is the power supply. If the end user need to access to the internet, the power supply to the network access equipment at user premise needs to be always switched on.

In the rural environment especially the remote rural, the only way for them to have continuous power supply is make sure that their diesel power generator to run at all the time and this is very costly for long run. Thus, it will be a very costly internet access solution, not because of the network access but it is because the running cost of the power provision. According to economic situation in the rural area, not many people can afford to run it. Most of the rural villages are powered by the diesel generator which only operates for very limited hours. Operating hours are only from 7pm to 11pm which is about 4 hours or less. In such agenda, self sustainable power supply
model or architecture is mandatory for enabling mass end user equipment deployment in the rural houses. This research paper is mainly concerned to conduct experiment on Intelligent Self Sustainable Green Energy Power Supply Model for Rural Network Access and Home Appliances model such as the one based on the green sources which is solar energy to power up the equipment in the rural long houses that is not available in commercial market. This proposed experiment is very important as the rural area has got no continuous power supply and the power availability is very limited.

Figure 1.0: Main component of proposed model

Figure 1.0 shows the building block of the Intelligent Self Sustainable Green Energy Power Supply Model for Rural Network Access and Home Appliances. The proposed model is a power supply model for the end user wireless access equipment as well as the home appliances has been made based on the rural context where several environment factors have been taken into consideration in the overall design. The modularity of the proposed model allows the generator to be integrated together to the system for better use of available energy. The environmental factor determines the type of the renewable energy source used to achieve the maximum output and for this case is the solar energy, the size and weight of the equipment for fast as well as easy installation, the modular design of the equipment based on the transportation requirement, and finally the intelligence of energy efficiency based on the Bario user usage pattern. The main advantages of the intelligent self sustainability power supply model are that it greatly cut down cost of network access equipment deployment in the rural with no power availability and with less expenses of generator fuel consumption in order to keep the technology running.

Besides that, the proposed model has the intelligence to monitor the incoming free energy from solar. The free energy from the sun which is collected by the solar panel is not consistent as it always depends on the availability of sun light. Especially during the rainy or cloudy days, power supply output from the solar panel will become very less as the number of hours of the sunlight is greatly reduced. If all appliances drain power from the same source of power supply for this case is battery, the life span of the battery depends on the total power usage of all appliances. For the case of rainy season or at night, many non critical appliances will drain more power from battery quickly and causing the critical equipment such as network access equipment will be switched down due to lack of power in backup battery. The intelligent power supply model regulates the power so that it will make sure that the on demand critical appliances will continue to operate by its allocated sufficient amount of power supply. The proposed model has the intelligence to separate the critical and non critical appliances so that critical appliances are not brought down by the usage of non critical higher power consumption appliances. However, during continuous rainy days, the system will experience a great reduction in producing the renewable energy. In this case, the energy provision will be less for the appliance. Thus, the proposed model intelligently reduces the operating hours of the appliances according to user usage pattern for a better use of energy. The critical appliance such as network access equipment will be switched on during the desirable hours only according to usage pattern.

RESULT & ANALYSIS

To analyze the impact of the intelligent self sustainable green energy power supply for rural in terms of operating hours of the network access equipment and home appliances, a simulation through MATLAB was carried out and outcome is discussed below. The simulation was carried out for 2 types of system, where by one graph is to compare total operating hours between self sustainable green energy power supply model and conventional power supply model. Another graph is to compare the total operating hours between self sustainable green energy power supply model integrating power generator and conventional power supply model. The conventional power supply model is the power supply which came from conventional solar panel system. Before performing a performance testing on the network access and home appliances, an assumption has been made to create a random weather condition. Random generated weather condition data was entered into the simulation so that it acts as an initial raw data.

Random Generated Weather Condition and Solar Power Input

Figure 1.1: Average Weather Condition and Average Solar Panel Input

Figure 1.1 shows the weather condition that was produced randomly which is per hour basis for a 30 set of data
representing 30 days and the solar panel input as well depending on the various weather condition generated. Both parameters were recorded as an average data per hour. Solar panel in this simulation assumed 6 solar panels rated at 100 watts. With maximum sun intensity, the output of the solar panel array is 600 watts or about 42 amps. As seen in figure 1.0 the shape of the graph of the weather conditions and solar panel input shows the same pattern as the solar panel input is depending on the weather condition in order to generate the solar power which is measured in Ampere per hour. This simulation is based on various types of weather condition which includes sunny, cloudy and rainy as the weather is assumed unpredictable by its nature. Thus the graph shows that the weather is uncertain in each day, whereby when the weather graph shows between the average number of 2.5 to 3.4 which means it is more likely a sunny day and 1.5 to 2.5 means it is a normal day which includes cloudy and sunny and finally the range between 0 and 1.5 would be a rainy day.

The weather condition is important in this system as the performance of the network access equipment and home appliances is much depending on this as it decides the power sources that can be conserved by the solar panel and it also decides how much battery capacity of the intelligent power supply can be charged in order to stay active at all time especially for the critical equipment. In this simulation, assumption has been made to observe the result in performance of the hardware which applied with conventional power supply model and intelligent power supply model.

**Intelligent self sustainable green energy power supply model comparing with conventional power supply model**

![Figure 1.2: Proposed model comparing with conventional model for critical equipment](image)

Table 1.0: Total operating hours for critical appliances

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Intelligent Power supply</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telephone</td>
<td>31.3</td>
<td>60.7</td>
</tr>
<tr>
<td>Network</td>
<td>31.3</td>
<td>60.7</td>
</tr>
<tr>
<td>Light</td>
<td>31.3</td>
<td>77.3 (116/150)</td>
</tr>
</tbody>
</table>

Table 1.0 shows the result of the total operating hours for the appliances associated with graph from figure 1.2. The intelligent power supply was kept on monitoring the incoming power from solar panel as well as the power level in the battery subsystem. The conventional power supply which is from conventional solar panel does not have the intelligence to regulate the energy which means that the critical equipment is not interrupted. It also will not be able to priorities the appliances based on critical and non critical appliances. When there is enough power coming out form the solar panel and the reserve energy in battery reach up to comfortable amount, the intelligent power supply model will make decisions to prioritize the equipment as critical and noncritical appliances.

The intelligent power supply only supplies the power for the critical equipment first, in other words, higher priority is given to the telephone, network access equipment and house light. Thus, the total operating hour for the telephone and network access equipment is higher which is, 60.7% compared with those appliances applied with conventional solar panel which is just 31.3%. The intelligent power supply makes sure the light runs according to user usage pattern. For this case, the proposed model will switch on the light start form 7 pm to 11pm. During this hour, the light is highly demand by the users. The conventional power supply will not follow the user usage pattern.
pattern. Thus, it leads to energy wastage during the afternoon hours. Intelligent power supply can save a lot of energy and will use the energy efficiently during night hours. According to percentage, light running with intelligent power supply module, one will experience less black out hours compared with conventional solar panel system. The total operating hours in percentage for the light is 77.3, which is far more efficient compared with conventional power supply which is only 31.3%.

![Figure 1.3: Proposed model comparing with conventional model for non critical equipment](image)

Figure 1.3 shows simulation result for total operating hours for the of non critical appliances running with intelligent power supply model and comparing with the total operating hours for conventional power supply model. The pink line shows that total operating hours for the TV and the yellow line for the DVD player and the red line is for TV as well as DVD applied with conventional power supply.

According to the graph, during rainy days for instance day 13 to day 14, the intelligent power supply will switch of the TV due to the less power provision. Thus, the intelligent power supply will make sure supplying power to the critical appliances which is more important than the TV and DVD in which both are non critical appliance. This enables to sustain the operation for the critical appliances during peak hours. The intelligent power supply will regulate the operating hour for the TV and DVD for certain period of time only according to user usage pattern. For this case, intelligent power supply will switch on the TV or DVD start from 7pm to 12am if reserved power in battery subsystem is in comfortable amount. This is because people in rural area usually will use these appliances at night hours and is not required during office hours. However, during afternoon, if people want to watch any special program in TV, they would able to watch in their PC or laptops because the network access equipment is active during afternoon.

### Table 1.1: Total operating hours for non critical appliances

<table>
<thead>
<tr>
<th></th>
<th>Total Operating Hour (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conventional</td>
</tr>
<tr>
<td>TV</td>
<td>31.3</td>
</tr>
</tbody>
</table>

Table 1.1 shows the result of the total operating hours for the appliances associated with graph from figure 1.3. The total operating hour for the TV 31.3% using conventional power supply whereas using the intelligent power supply the total operating hour for TV is 50.7% which is higher than conventional power supply. The DVD player is the only one appliance which contains less operating hour compared with the DVD player applied with conventional solar panel which is 4.7%. This is because, most of the power supply is channeled to the critical equipment which are, network access equipment, telephone and light as shown previously in figure 1.1.

Intelligent self sustainable green energy power supply model integrated with power generator comparing with conventional power supply model.

![Figure 1.4: Proposed model integrated with power generator comparing with conventional model for critical equipment](image)

Figure 1.4 shows simulation result for total operating hours for the critical appliances running with intelligent power supply model integrated with power generator and comparing with total operating hours for conventional power supply model. The telephone and the network access equipment having same operating hour, thus the black line and blue line overlaps each other. Green line is the total operating hour for light whereas the red line is total operating hour of appliances applied with conventional power supply model. Comparing with the previous system, the output simulation result for intelligent power supply model integrating with generator is much better. Most of the days, the network access equipment and telephone operated for 24 hours except during rainy days. During the rainy days, example for day 14 to day 15, there is no input power from solar panel. Thus when power provision is getting less, the intelligent power supply reduced the operating hours for the network access equipment and telephone for certain hours which from mid-night after 12 am until office hours start. This is because user usually does not use the network during mid night. This will make sure the network access equipment will not be down especially during office
hours. The intelligence power supply is able to regulate the operating time of appliances according to user usage pattern for better use of energy instead of just wasting the energy during idle hours.

The power generator in Bario is operating about 4 hours starting from 7pm to 11pm. Thus it is able to supply power for the most of the appliances for 4 hours, without relying on backup power from battery subsystem. At same time it also will charge the battery. During these hours, the intelligent power supply will use the power from generator to power up the telephone, network access equipment and light. Thus one will not experience any blackout hours in the whole month. The proposed model will make sure the light operates start from 7 pm to 12am according to user usage pattern. After about 11pm when no more power coming from the generator, intelligent power supply will switch the power supply to battery mode so that the appliances will continue to operate. This enables to prolong the operating hour for network access equipment and telephone.

Table 1.2: Total operating hours for critical appliances

<table>
<thead>
<tr>
<th></th>
<th>Total Operating Hour (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conventional</td>
</tr>
<tr>
<td>Telephone</td>
<td>31.3</td>
</tr>
<tr>
<td>Network</td>
<td>31.3</td>
</tr>
<tr>
<td>Light</td>
<td>31.3</td>
</tr>
</tbody>
</table>

Table 1.2 shows the result of the total operating hours for the appliances associated with graph from figure 1.4. The total operating hours for the network access equipment and telephone is almost 100% which is 96.1% comparing appliance applied with conventional solar panel system which is only 31.3%. Total operating hour for network access equipment and telephone is 60.1% by applying intelligent power supply without integrating the power generator. The light operating hour is 5 hours per day. The total operating hour for the light is 100% without any interruption by applying intelligent power supply integrating the power generator compared with the light applied with conventional power supply which is only 31.3% and total operating hour for light applied with intelligent power supply without integrating power generator is 77.3%.

Figure 1.5 shows, simulation result for total operating hours for the of non critical appliances running with intelligent power supply model integrated with power generator and comparing with total operating hours for conventional power supply model. The pink line shows the total operating hour for the TV and yellow line is the total operating hour of DVD player. Both appliances contain same operating hour thus it overlaps with each other. The red line is the total operating hours for TV and DVD applied with conventional power supply model.

Table 1.3: Total operating hours for non critical appliances

<table>
<thead>
<tr>
<th></th>
<th>Total Operating Hour (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conventional</td>
</tr>
<tr>
<td>TV</td>
<td>31.3</td>
</tr>
<tr>
<td>DVD</td>
<td>31.3</td>
</tr>
</tbody>
</table>

Table 1.3 shows the result of the total operating hours for the appliances associated with graph from figure 1.5. Total operating hour for the TV and DVD is 100% if one using with intelligent power supply integrating power generator compared with one using the conventional power supply model which is only 31.3%. Comparing total operating hour for TV and DVD using intelligent power supply model integrating power generator is 100% which is more than the total operating hour for TV and DVD using intelligent power supply model without integrating power generator is 4.7% as shown in figure 1.3.

Battery performance of the intelligent self sustainable green energy power supply model integrated with power generator compare with conventional power supply model

In this section a comparison and evaluation in terms of battery capacity between the conventional power supply model and the intelligent self sustainable green energy power supply model integrated with power generator was made. In this
simulation an assumption has been made for battery model. Battery model with capacity of 1000 Amp per hour with 12.7 voltages is assumed as an initial value for operation which is 100% of fully charged battery capacity. The battery capacity will not go less than 50% in order to sustain the battery lifespan. Charging and discharging the battery depends on the total current consumed by the hardware. If the total current is more than the solar input, then battery will be discharging.

\[
\text{Discharging} = \text{Total current} - \text{Solar input} \\
\text{Battery capacity} = \text{Battery capacity} - \text{Discharging}
\]

If the amount of current produced in solar panel is more than the total current consumed by the hardware then the battery will be charging.

\[
\text{Charging} = \text{Solar input} - \text{Total current} \\
\text{Battery capacity} = \text{Battery capacity} + \text{Charging}
\]

![Figure 1.6: Average Battery Capacity based on Daily basis](image)

Figure 1.6 shows the comparison of the battery performance between the intelligent self sustainable green energy power supply model and conventional power supply model. The blue line shows that the average total battery capacity in daily basis for intelligent self sustainable green energy power supply model. The red line shows the average battery capacity in daily basis for conventional power supply model. As seen in the figure 1.0, during rainy days like 6 to day 7, day 13 to day 14, day 20 to day 21, day 23 to day 24, the intelligent power supply did not show the incoming solar power. Thus, the intelligent power supply gives priority to the critical hardware. Therefore, the battery power will only supply power to critical hardware causing the battery power to be drained. Better power management by the intelligent self sustainable power supply model makes sure that the power is always available to that critical hardware as shown in the figure 1.3.

The battery performance of conventional power supply model, claimed to be less energy efficient than the intelligent self sustainable power supply model. The significance of the intelligent self sustainable green energy power supply model integrated with power generator is that it is able to charge the battery during night hours from 7 pm to 11pm. The current came from the diesel power generator can power up all the appliance and charge the battery so that the power from the battery can be used later after the generator was shutdown or during non presents of solar input.

**CONCLUSION**

The special power condition of the remote areas has motivated the need to invent appropriate and efficient power supply design to get the objective of higher power conservation. The network access equipment and home appliances for the remote deployment should be optimized to take full advantages of the existing green energy and backup battery subsystem. The intelligent self sustainable green energy power supply model, are applicable to the context of rural implementation would have driven the energy conservation into the next stage where by the investment on green energy technology could be scaled down by optimizing the network access equipment and home appliances power consumption behavior. As overall performance evaluation on both systems, the intelligent self sustainable green energy power supply model is claimed to be more energy efficient as it allows the network access equipment and home appliances to consume the power efficiently based on different conditions as well as it is able to improve the operating hours for critical and non critical appliances.

**ACKNOWLEDGEMENTS**

Faculty of Computer Science and Information Technology, University Malaysia Sarawak (UNIMAS) wants to thank the ERGS for sponsoring this project (contract under research grant ERGS/02(03)/855/2012(07)).

**REFERENCES**


A SURVEY OF MAJOR INTER-USER INTERFERENCE MITIGATION TECHNIQUES IN WIRELESS BODY AREA NETWORKS

Soh Brendard Nji, Mujahid Tabassum, Dr. Chua Hong Siang
Faculty of Engineering, Computing and Science
Swinburne University of Technology, Sarawak Campus
Jalan Simpang Tiga, 93350 Kuching, Sarawak, Malaysia

ABSTRACT

Wireless Body Area Networks (WBANs) have the potential to completely revolutionize the healthcare industry. However, inter-user interference is one of the greatest impediments of having this dream realized. This write-up surveys adaptive modulation, adaptive data rate and duty cycling; transmission rescheduling; channel hopping, non-cooperative game theory; genetic fuzzy logic; inter-BAN schedule sharing, and reinforcement learning with approximation as interference mitigation techniques in WBANs. It then proposes a selective transmission and temporal data caching approach as schemes to further curb energy waste and increase throughput in wireless sensor networks.

INTRODUCTION

WBANs constitute wearable low data rate nodes placed in, on, or around the body to measure and transmit physiological conditions such as temperature, motion, ECG, EEG, EMG, SpO2 amongst others, conveying the data to physicians and other stakeholders via coordinators and base stations (Bin Cao et al., 2013). The IEEE 802.15.4 standard is used for reference. Just like for Wireless Sensor Networks (WSNs) (Mujahid et al., 2014) of which WBANs constitute a subset, each node comprises of a sensor, microprocessor, memory, transceiver and battery which is the most expended component. WBANs are low power low data rate devices. Higher reliability requirements, smaller node and battery sizes, mobility and heterogeneous as opposed to homogeneous nodes and traffic are some of the characteristics which are peculiar to WBANs and differentiates them from WSNs (Guowei et al., 2010), (Anup et al., 2012). Mobility increases their susceptibility of interference with other WBANs in close proximity, deteriorating their link quality and packet delivery ratio (Wen et al., 2012). Communication is via a star topology, with transmission consuming 70% of energy (Wen-Bin et al., 2011). A number of inter-user interference mitigation techniques have been advanced to facilitate the effective and efficient implementation of WBAN technology as a boost to the healthcare industry. This paper surveys major approaches to this effect, and proposes a way forward in a bid to improve performance of WBANs.

RELATED WORK

Some considerable work has been done on inter-user interference mitigation in WBANs using different technologies with varied results. However, none as yet has provided a good-for-all solution.

To begin with, adaptive modulation, adaptive data rate and duty cycling were proposed as a means of mitigating inter-user interference in WBANs by reducing transmission power levels (Wen-Bin et al., 2011). For frequency modulation, given each Bit Error Rate (BER), low and high Signal to Interference and Noise Ratio (SINR) thresholds were set. The carrier frequency was switched between least complex BPSK, moderately complex QPSK and highly complex 8PSK modulation types, depending on whether the SINR was below the low threshold, between the low and high thresholds, or above the high threshold values respectively to reduce transmission power. Likewise, adaptive data rates switched to different data rates between low and high thresholds, depending on the obtained SINR compared with the expected SINR. Using the ratio of active to [active + inactive] transmission window of the IEEE 802.15.4 superframe, different duty cycle levels were set based on a comparison of the obtained and expected SINR, with neighbouring WBANs less prone to interference in longer inactive window scenarios, since there was enough space to retransmit in case of failure. An Interference Mitigation Factor (IMF) was used to compare effectiveness of the adaptive schemes based on the difference in transmission power before and after the schemes were implemented. Though these schemes reduced inter-user interference and power consumed, it was at the expense of throughput. The performance evaluation and implementation complexity of the adaptive methods need to be investigated as well.

Secondly, the effect of inter-user interference on WBANs was evaluated using PER (Packet Error Rate), taking into consideration influential factors such as BAN (Body Area Network) density, traffic rate and...
transmission power (Wen et al., 2012). In an experimental setup, the average area per person's BAN was 2m², 3m² and 6m² for peak, moderate and off-peak hours respectively. All BANs transmitted at the same power level, had the same duty cycle, and operated under the same channel. Intra-BAN signals attenuated as per the on-body path loss model, and inter-BAN signals attenuated as per the in-body path loss model. Cumulative distribution functions (CDF) from experimental results depicted an increase in PER with increase in BAN density and duty cycle, and reduction in reliability level (ratio of number of packets received with PER lower than required PER to number of packets transmitted) obtaining 21.9% at peak, and 68.5% at off-peak periods with a duty cycle (DC) of 0.2. PER increased with a reduction of transmission power and yielded a 54% increase in average PER while transmitting at -25dBm relative to when transmitting at 0dBm. The difference in PER caused by different DC at low power was also less because when power is low, loss of packets is caused more by body blockage than by inter-user interference when transmission power is high. The performance evaluations helped determine what magnitude of parameters (BSN density, transmission power, DC) to use during configuration and deployment in order to achieve specific reliability requirements. Only an investigation of inter-user interference in WBANs was made, pending development of a mitigation strategy.

Thirdly, Inter-user interference was mitigated by WBANs selectively rescheduling their transmissions or switching to another channel in a distributed manner, relying on information gotten from neighbouring WBANs by overhearing their beacons (Wen et al., 2012). During the active 802.15.4 superframe duration, coordinators inform their sensors of the channel conditions via a beacon, and then the sensors transmit their data, going back to sleep after transmission. When coordinators couldn’t receive packets from sensors as a result of interference, the Inter-User Interference Mechanism (IIM) was initiated and the coordinator listened for a period equivalent to a superframe duration, not going to sleep. The coordinator of the affected BAN rescheduled transmissions using the Wireless Body Area Network ID, start time and end time. The coordinator of each WBAN maintained its neighbour list. If the number exceeded what the superframe could accommodate, sensors swapped to another channel or rescheduled transmission. Retransmissions began with listening and back-off to prevent further collisions. For retransmissions, if the channel is busy, the sensor waits till it’s free. If it’s free but not enough to accommodate the packet entirely, the sensor again waited till the end of the said transmission. As long as the neighbouring list is complete, rescheduling is good to go. Collisions due to incomplete neighbour list were resolved by carrier sensing and back-off time mechanisms. As long as the number of WBANs in the interference range didn’t exceed the number of available channels, inter-user interference could be mitigated properly by using channel switching. In experiments conducted using 5, 10 and 20 WBANs in a 40m x 40m space, throughput increased by 30% with use of the proposed IIM mechanism. Energy consumption per successful delivery of a packet also increased by 16%, being the cost of listening for the duration of a superframe instead of sleeping in the inactive period. This is because IIM curbs collisions and retransmissions. Implementation of IIM in experiments using physical components to make it more realistic is pending.

Next, Inter-user interference was mitigated by the distributed, non-cooperative and self-organising game theory, in which each BSN measured the inter-user interference of neighbouring BSNs and used it to adaptively select its transmission channel and power without intercommunicating, using the no regret algorithm. Each user constituted a player in the game, and belonged to a set of users. A set of available non-overlapping channels was created as well. The transmission power of the user depended on the type of physiological sensor, and the received power was its signal gain. Each user strived to attain the SINR threshold, and selected a utility-minimising strategy based on the strategy history of other users, but caring less about their attainment of the Nash Equilibrium (NE) which is a stable point. If chosen by all players, the no regret strategy resulted in them attaining a Correlated Equilibrium (CE) which is the product of each player taking a different action, and corresponds to a NE. A user selected a channel randomly with equal probability, calculated the maximum transmission power and utility for that channel, and then the probability of selecting that channel based on its attainment of the probability threshold which symbolises the NE. This was iterated for the available channels based on the probability distribution until the NE was reached. In an experiment conducted with 5 BANs each with 3 sensors and 1 coordinator using 4 non-overlapping channels, a reduction of average system SINR was experienced when the system used DISG as opposed to when it didn’t. The drawback of this method stems from the fact that the NE is just a stable, not an optimum state. NE is attained when each BSN transmits with a different channel. Endless iterations ensue if BSN number exceeds the number of available channels. Long wait times due to many iterations prior to obtaining a solution might lead to packet loss from buffer overflow.

In addition, a fast converging Fuzzy Power Controller (FPC) with feedback which took as input current interference power level, SINR and current transmission power (Interference power level and SINR measured at digital receiver and fed back into transmitter through low throughput error-free control channel) was proposed (Kazemi et al., 2011). The genetic algorithm was used to design FPC, maximising throughput, minimising power consumption and convergence time irrespective of the
number of nodes at the expense of small capacity. Using robust, easily designed and flexible nodes, fuzzy controllers mapped complicated non-linear relationships between input and output spaces. Genetic Algorithms (GA) are stochastic search algorithm based on Darwin’s law of evolution where an objective function called fitness was optimised. Inputs to the function termed chromosomes constituted potential solutions to the problem. Recombination operators such as crossover and mutation were applied to the chromosomes called populations iteratively to obtain new individuals called generations until a stop criterion was reached e.g. a good enough individual was found, or the maximum number of generations was reached. The goal of GA was to select individuals which maximised the objectives of the fitness function, which in this case was to maximise capacity in WBANs while using the least power and smallest convergence time possible. A comparison of FPC with alternative game approaches revealed that ADP (Asynchronous Distributed Pricing) and PCG (Power Control Game) consumed about 55% and 20% more power respectively relative to FPC with high node density. PCG consumed 40% more power in low node density scenarios. ADP and PCG had a 20% capacity increase relative to FPC in low node density scenarios, but this levelled up with high node density. FPC remarkably outperformed both ADP and PCG with regards to the number of iterations hence convergence time to arrive at a stable solution, regardless of node density. Implementation of FPC on a WBAN hardware platform is pending.

Next, a TDMA-based common schedule-creating interference mitigation solution by sensing neighbouring BANs was proposed, enabling cooperation and defining a common time for inter BAN schedule sharing (Malhapatro et al., 2012). Schedule interleaving was used in case of high density deployment to minimise node latency. Each BAN calculated the assigned time and average wait time of the nodes in its network. Each coordinator calculated its Signal-to-Interference-Ratio (SIR) and any below the threshold value triggered the exchange of its schedule with others. Two BANs could simultaneously use a slot if they were not in each other’s interference range. In an experiment to simulate BANs of different density and speeds on a 50m x 50m space, an increase in the number of BANs and BAN speed resulted in a decrease in packet delivery, depicting that the proposed solution worked best in low BAN density scenarios. The latency also increased with higher BAN densities. The drawback of this method lies in the fact that performance increase was achieved at the expense of increased latency.

Additionally, the BAN-BAN Interference Reduction (B²IR) technique rescheduled beacon packets to avoid overlap of active periods between different BANs (Grassi et al., 2012). Using the 802.15.4/Zigbee standard, periodic beacons were sent to demarcate a superframe, a time frame which determines how the sensors communicate with the coordinator. The contention-based CAP (Contention Access Period) used carrier sensing, allowing packets to be transmitted even during overlaps, meanwhile the CFP (Contention Free Period) contained Guaranteed Time Slots (GTS) which are prone to collisions because their transmission isn’t proceeded by channel sensing. In an experiment to quantify the effect of single-channel interference on BANs consisting of 1 coordinator and 2 sensors each by varying number of conflicting BANs & their packet sizes using OMNeT++ Castalia simulator, varying the number of packets transmitted per second from 5 to 60 resulted in more packet loss & collisions due to buffer overflows (when CSMA/CA delays packet transmission due to lack of free channel), while collisions were caused by GTS packets transmitted without channel sensing. The number of available channels needed to be more than number of interfering BANs for channel switching to be effective. The estimated time for packet transmission also needed to be more than the product of superframe slots, base slot duration and superframe order, and another packet’s transmission shouldn’t be in progress, else transmission is rescheduled. With B²IRS, energy consumption increased from 7% with 2 BANS at 5 packets per second to 32% with 4 BANS and 60 packets per second when B²IRS wasn’t used to only 1.6% when it was used. When the SO of 2 BANs was increased from 4 to 5 thereby enabling a longer active than inactive interval, packet reception rate reduced, and energy used increased from the previous 1.6% to 5.3%. One pitfall of this method is that rescheduling can be endless in a high BAN density scenario.

Mindful of the fact that transmission power accounts for 70% of power consumed by sensors, a dynamic, flexible and reconfigurable contention-based power-reduction scheme using reinforcement learning to capture WBAN environment states while using Radial Basis Function (RBF) approximators to reduce number of states and complexity was proposed in (Kazemi et al., 2011). Each agent in a multi agent system performed an action as per the environment state, moving to the next state as a reward. The agent either used exploitation which is selecting a state from its previous knowledgebase for quick immediate action, or exploration to better cope with new situations. Agents used an approximation parameter to learn compacted state-value pairs to curb dimension and complexity. In an experiment run 1000 times comprising WBAN users walking randomly in a 10m x 10m space, Reinforcement Learning Power Controller (RLPC) used less power per node as a function of WBAN density relative to Power Control Game (PCG) and Fuzzy Power Controller (FPC) alternatives by 10dBm and 8dBm respectively. In terms of energy consumption, RLPC recorded above 90% and 80% performance improvement as opposed to PCG and FPC respectively. PCG however
outperformed FPC and RLPC by 15% and 25% respectively with throughput in low node density scenarios, while the discrepancy reduced in high node density scenarios. Node lifetime of RLPC was 5 times that of FPC and 10 times that of PCG. While PCG only relied on the reward function, FPC only captured the environmental condition once, RLPC continuously learned from its environment, being able to capture new discrepancies. Implementation of FPC on physical WBANs is awaited.

Lastly, mobility, body orientation and blockage in experimenting the effect of inter-user interference at different levels of intensity on WBANs; established performance benchmarks for measuring the effect of inter-user interference on realistic WBANs in different environments at different levels of complexity, and proposed frequency hopping with pre-allocation hopping channel using beacon-enabled CSMA/CA to mitigate inter-user interference in an incomplete manner were introduced (Bin Cao et al., 2013). Experiments were conducted to investigate the baseline performance of TelosB motes and the effect of inter-user interference on them without human body effect using one to 5 Wireless Body Area Networks (WBANs). Results reported similar Packet Delivery Ratios (PDR) for both slotted and unslotted CSMA/CA in sparse network distributions, a fairly stable PDR for slotted CSMA/CA regardless of network density, relatively higher throughput for unslotted CSMA/CA in sparse network deployments due to acknowledgment and retransmission overhead in slotted CSMA/CA, steadily reducing PDR and throughput in unslotted CSMA/CA due to the effect of inter-user interference, and relatively high throughput in slotted CSMA/CA due to its scheduling scheme. In a second experiment with a body between two WBANs, throughput increased by 60% and 85% for slotted CSMA/CA and unslotted CSMA/CA respectively, relative to when there was no blocking body. PDR was higher when user backed the wall than when facing the wall, because interference blocking was more in the former. In dense Line Static (LS) deployments, performance reduced steadily with increase in WBANs for unslotted CSMA/CA. However, performance was haphazard for slotted CSMA/CA because different nodes experiencing interference launched frequency hopping and back-offs at different times. The hopping approach maintained a PDR of about 95% despite network density. In another experiment pitting users with WBANs moving randomly, performance of slotted CSMA/CA and hopping approach exceeded that of unslotted CSMA/CA, with hopping approach exceeding slotted CSMA/CA and unslotted CSMA/CA by 144% and 164% respectively. Hopping approach improvement for higher throughput is anticipated.

ANALYSIS AND DISCUSSION

A scrutiny of the inter-user interference mitigation methods covered above reveal that a combination of transmission scheduling and frequency hopping using reinforcement learning to dynamically and continuously capture environmental conditions constituted the best approach in mitigating inter-user interference while maintaining low power consumption which is the scarcest resource in WBANs. However, this was done at the expense of throughput. Considering the criticality of data transmitted by WBANs, a throughput enhancing mechanism is required in order to speed up adoption of the technology. To this end, I propose selective transmission and temporal data caching to append previous transmission scheduling and frequency hopping approaches.

Overview of the selective transmission and data caching approach

To improve throughput, transmission rescheduling is added to the coordinator’s procedures, while temporal data caching and selective transmission are appended to the sensor’s operations, modifying the channel hopping approach. The digital receiver captures input for SINR and interference power level. The SINR model which is an extension of the propagation model (Mujahid et al., 2014) is used to compute SINR from the captured input which forms the basis of transmission power determination and hence inter-user interference mitigation. SINR is calculated using equation (1) below (Bin Cao et al., 2013).

\[
\text{SINR} = \frac{P_{tx}}{(\theta - R)^\alpha + \frac{P_{tx}}{(\theta - D)^\alpha} + \frac{P_{tx}}{(\theta - \Phi)^\alpha} + \frac{P_{tx}}{(\theta + R)^\alpha}}
\]

(1)

where

- \( P_{tx} \) = Received signal strength at receiver
- \( R \) = Receiver – Transmitter distance
- \( \theta \) = Path loss exponent (generally from 2-4)
- \( D \) = Carrier sense range

The proposed scheme functions by carrying out a series of procedures at the coordinator and sensor nodes as depicted by (b) and (d) respectively in fig. 1 below.
Procedures at the coordinator

- Coordinator captures the Body Sensor Network ID (BSNID), transmission channel, backup transmission channel, schedule start time, schedule end time, packet size, SINR, Interference power level; from the beacons of controllers of all neighbouring BSNs.
- The next free transmission time that can be used to reschedule transmission is detected from the schedule start time and schedule end time.
- A transmission channel is selected to forward beacons for synchronization and transmission. A back-up channel for hopping is selected as well, and the sensors in the WBAN are informed of it.
- Initialised to zero, a counter is set which counts the number of times transmission delay exceeds the threshold value.
- The beacon is broadcasted.
- The coordinator receives packets from sensors, calculates their transmission delay, compares it with the threshold and increases the counter if the threshold is exceeded.
- If the counter exceeds the threshold, interference is imminent. The coordinator reschedules the packet transmission to the next free slot within the current superframe if available, or hops to the next available channel in the next superframe.
- If a different channel is selected, channel sensing and back-off algorithms are carried out to prevent possible collisions due to incompletely received neighbor information prior to the coordinator broadcasting a beacon.

Procedures at the sensors

- Sensor scans channels to capture beacon from its coordinator.
- If beacon is sent by its coordinator, the sensor synchronises and receives information such as transmission and hopping channels, BSNID, schedule start time, schedule end time, and transmission power to be used. The beacon is discarded if from another WBAN, and scanning resumes.
- Sensor stores copy of data to be transmitted and compares all subsequent data for transmission to it. If same, it is discarded and no transmission occurs. If different, the packet is transmitted, the stored data replaced, and the comparison continues.
- Sensor transmits data to coordinator and waits for the next beacon for synchronization.
- In the absence of a beacon, the sensor continues scanning in constrained time. If the time is up, interference is imminent, so the sensor switches to the hopping channel since no data can be transmitted in the absence of a beacon.

CONCLUSION

On a conclusive note, this paper analysed the major inter-user interference mitigation approaches, highlighting their modus operandi and shortcomings. Next, the procedures of a proposed inter-user interference mitigation approach are brought to life, capitalizing on extant research as a resource. Rescheduling transmissions prior to switching channels by coordinator, as well as data caching and transmission of new data only by the sensors is expected to reduce bandwidth usage and power consumption respectively, as per the proposed mechanism. Considering the fact that only the preliminary skeleton of the proposed protocol is displayed below, future work involves experimental testing using the NS2 simulator on a Linux platform and implementation on TelosB hardware to ascertain validity of the approach.

REFERENCES


Fig. 1 Procedures at the coordinator and sensor for channel hopping and selective transmission approaches respectively.

(a) Procedures at coordinator for hopping approach

(b) Procedures at coordinator for proposed approach

(c) Procedures at sensor for hopping approach

(d) Procedures at sensor for proposed approach
MAXIMIZING TRUE POSITIVES FOR SIGNATURE-BASED NETWORK APPLICATION IDENTIFICATION

Justin Tharp¹, Jinoh Kim¹, Sang C. Suh¹, Hyeonkoo Cho²
¹ Department of Computer Science, Texas A&M University, Commerce, Texas 75428, USA
² Sysmate Inc., 1290 Dunsan-Dong Seo-Gu, Deajeon, 302-830, Korea
Email: jtharp2@leomail.tamuc.edu, jinoh.kim@tamuc.edu, sang.suh@tamuc.edu,

ABSTRACT: Identifying network applications has become the core of network operations and management for various purposes. For accurate classification of network traffic, an approach using common patterns called signatures is widely studied to compensate the limitations of traditional port-based identification. However, our recent observations with a set of well-known signature generation algorithms indicate that simply using the collected signatures may lead to very poor accuracy less than 60% of true positives even in an optimal case, due to the poor quality of signatures. In this work, we present a signature refining process to create a high quality of signature from the crude signatures initially acquired by a signature generation method. Our experimental results show that using a new set of signatures after refining can dramatically improve the upper bound of true positives, which should be the first step towards an acceptable degree of identification accuracy.

I. INTRODUCTION

Identifying network applications has become the core of network operations and management. Such areas include network planning, capacity provisioning, Quality of Service (QoS) assurance, security policy management, traffic engineering, and so forth [1,2,3,4]. The classical technique for identifying network traffic is using TCP/UDP port numbers. Even though in the past this port-based identification was reliable and accurate, it has become inaccurate due to several reasons. First, many applications including peer-to-peer applications allocate port numbers dynamically or even randomly, which disarms the classical port-based identification assuming static port numbers for every application. Even for applications defining standard port numbers, using non-standard ones could be possible as some applications allow users to configure port numbers to be used. For example, BitTorrent users can choose port numbers outside the given range (i.e., 6881–6889) [5]. Second, a number of applications use HTTP tunneling as a transport method, and in that case, traffic for those applications would simply be classified into HTTP because of the port number 80 [6].

Due to these limitations of port-based identification, alternative techniques have been introduced [7,8,9,10,11,12,13], including signature-based identification that relies on a set of common patterns called signatures observed from flows belonging to that application, where a flow is a set of packets for a single session of communication. In this class of identification, signatures are collected from a training set with a signature generation algorithm such as LCS [10], LASER [7], and AutoSig [8], which are then used to identify an input stream by examining its payload with the collected set of signatures. With the payload inspection, it is known to be highly accurate with respect to identification performance although complexity is also high. However, our observations with existing signature generation algorithms indicate that simply relying on the collected signatures (without refining them) may lead to very poor accuracy less than 60% of true positives even in an optimal case, as we will show in this paper.

The primary reason for the unacceptable identification accuracy basically comes from the poor quality of signatures. In general, signature generation is an iterative process, in which flows in the training data set are repeatedly compared to obtain common patterns, and a certain set of patterns that appear relatively frequently over the training data are chosen as signatures. Considering frequency however does not always confirm the quality of signatures. Suppose two signatures for application XYZ, “HTTP/1.1 200 OK XYZ” and “HTTP/1.1 302 Cache-Control XYZ”, obtained by using LCS. If an input stream actually belonging to that application has a pattern of “HTTP/1.1 304 Not Modified XYZ”, none of the above signatures can identify this flow. However, if we had more a common pattern from the signatures (e.g., “HTTP/1.1 XYZ”), we would expect a better identification result. This implies that the raw signatures after the iterative signature creation process could still be less general and somewhat specific to only a subset of flows in that application. It also suggests that it is necessary to improve the quality of
signatures. We refer to this post-processing as a signature refining process.

In this work, we present a signature refining process to create a high quality of signature from the crude signatures initially acquired from a signature generation method. Our post-processing includes the following two steps: (i) signature merging to consider any possible combination of the raw signatures to relax their specificity, and (ii) signature identification to discover a manageable set of signatures, particularly with respect to the volume of the signatures as the cost of signature-based identification is strongly correlated with the number of signatures for payload examination. Our experimental results show that using a new set of signatures after refining can dramatically improve the upper bound of true positives up to 99% (that was less than 60% before refining), which should be the first step towards an acceptable degree of identification accuracy.

This paper is organized as follows. In section 2, we provide a summary of related studies for this work. Section 3 describes the problem we would like to address and the following section presents our refining processes to create a high quality of signatures with greater true positives. In Section 5, we conclude our presentation.

II. RELATED WORK

Several algorithms for obtaining common patterns for identification have been introduced [7,8,9,10,11,12,13]. LCS (Longest Common Subsequence) [10] is a classical method to search common subsequences. LCS compares two strings together and looks for a common pattern of characters between the two strings. The characters do not have to be in the same location or right next to each other. For example, suppose we have a string \( S_1 = \text{"AABBCCDD"} \), and another string \( S_2 = \text{"ABCDFFEE"} \). When these two strings are compared in the LCS algorithm the resulting signature will be \( \text{"ABCD"} \) because these characters appear in the same order in both strings.

LASER (LCS Application Signature ExtRaction) [7] is a modified LCS algorithm that generates the longest common substring as a signature when comparing two strings together. Unlike the original LCS algorithm, LASER looks for the longest chain of characters that follow each other in both comparison strings. The LASER algorithm uses the same matrix traversal scheme as LCS but it keeps track of how many diagonals it traverse in one instance. There could be several substrings found in the two strings but LASER will take the longest substring found as the signature.

Another well-known algorithm for signature generation is AutoSig [8] that looks for common substrings for signatures but does not use the LCS algorithm as its base like LASER. Instead it tokenizes the packet information into chunks (called “shingles”). Once both input strings are completely split the shingles are compared. If two shingles are matches, they are kept as the starting point of the signature. This continues until the complete signature is generated. Once all the signatures are made from the training set, AutoSig adds another step by next comparing the shingles of all the signatures to make a signature tree. The tree will show all possible signatures that can be made with from certain starting points. In this work, we chose the LCS algorithm for signature creation as it is simple but still comparable to the other two techniques with respect to identification performance observed from our preliminary study [14].

Covering subset [15] is an algorithm to find a subset that covers an entire set with the smallest number of elements. For example we have a set \( U = \{1, 2, 3, 4, 5\} \) and we have a series of subset \( S = \{\{1, 2, 3\}, \{1, 3, 5\}, \{3, 4\}, \{4, 5\}\} \). The covering subset problem would choose the subsets \{1, 2, 3\}, \{3, 4\} and \{4, 5\} or it could choose \{1, 2, 3\}, \{1, 3, 5\} and \{3, 4\}. It can choose either one of these subsets because both of them cover all the instances in \( U \). One of methods we present for selecting a subset of signatures employs this algorithm as we will describe later.

Our prior study [14] addresses the problem of multiple matches arising when more than a single application identifies the data stream in question when using signature-based identification. The main motivation of this previous work comes from the observation that shows a significant fraction of identification resulted in multiple matches. To address the problem of multiple matches, a set of selection methods were presented that help battle multiple matches and identify the input stream accurately.

Google Refine [16] is free software that can be downloaded off the internet. Google Refine has many features that help with cleaning the data such as clustering and merging. There are several clustering algorithms that Google Refine uses, and we used the features for our signature refinement (i.e., merging). The first set of algorithms it has is called key collision. The first key collision algorithm is called Fingerprinting. Fingerprinting tokens the signature at the whitespaces and edits the signature in a certain way to generate a key for the signature. Next is N-Gram Fingerprint that does the same as Fingerprinting but tokens the signatures at certain length intervals based on what the user enters in as \( N \). Another clustering algorithm that Google Refine uses is Nearest Neighbors. The first Nearest Neighbor algorithm is called Levenshtein Distance. Levenshtein Distance groups together the signatures that meet a certain number of edits to the comparison signature. The other Nearest Neighbor is called PPM which uses a mathematical formula to group together signatures based on string comparisons.
Table 1. Applications and the number of flows in the data set

<table>
<thead>
<tr>
<th>Application</th>
<th>Category</th>
<th># Total Flows</th>
</tr>
</thead>
<tbody>
<tr>
<td>BitTorrent</td>
<td></td>
<td>3,676</td>
</tr>
<tr>
<td>Cyworld</td>
<td>Social computing</td>
<td>628</td>
</tr>
<tr>
<td>Daum</td>
<td></td>
<td>1,915</td>
</tr>
<tr>
<td>Naver</td>
<td></td>
<td>1,381</td>
</tr>
<tr>
<td>Wemakeprice</td>
<td></td>
<td>523</td>
</tr>
<tr>
<td>Afreeca</td>
<td></td>
<td>1,029</td>
</tr>
<tr>
<td>GomTV</td>
<td></td>
<td>940</td>
</tr>
<tr>
<td>Pandoratv</td>
<td></td>
<td>731</td>
</tr>
<tr>
<td>Tving</td>
<td></td>
<td>537</td>
</tr>
<tr>
<td>Melon</td>
<td></td>
<td>685</td>
</tr>
</tbody>
</table>

Table 2. Fraction of testing flows with at least a single signature for the application

<table>
<thead>
<tr>
<th>Application</th>
<th>Afreeca</th>
<th>BitTorrent</th>
<th>Cyworld</th>
<th>Daum</th>
<th>Gomtv</th>
<th>Melon</th>
<th>Naver</th>
<th>Pandoratv</th>
<th>Tving</th>
<th>Wemakeprice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afreeca</td>
<td>0.45</td>
<td>0.10</td>
<td>0.58</td>
<td>0.07</td>
<td>0.80</td>
<td>0.42</td>
<td>0</td>
<td>0.64</td>
<td>0.12</td>
<td>0.80</td>
</tr>
<tr>
<td>BitTorrent</td>
<td>0.15</td>
<td>0.48</td>
<td>0.19</td>
<td>0.04</td>
<td>0.28</td>
<td>0.13</td>
<td>0.01</td>
<td>0.13</td>
<td>0.11</td>
<td>0.18</td>
</tr>
<tr>
<td>Cyworld</td>
<td>0.48</td>
<td>0.05</td>
<td>0.79</td>
<td>0</td>
<td>0.53</td>
<td>0.10</td>
<td>0</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>Daum</td>
<td>0.80</td>
<td>0.02</td>
<td>0.95</td>
<td>0.04</td>
<td>0.91</td>
<td>0.15</td>
<td>0</td>
<td>0.87</td>
<td>0</td>
<td>0.82</td>
</tr>
<tr>
<td>Gomtv</td>
<td>0.24</td>
<td>0.39</td>
<td>0.59</td>
<td>0.37</td>
<td>0.81</td>
<td>0.57</td>
<td>0</td>
<td>0.20</td>
<td>0.37</td>
<td>0.68</td>
</tr>
<tr>
<td>Melon</td>
<td>0.43</td>
<td>0.06</td>
<td>0.48</td>
<td>0</td>
<td>0.48</td>
<td>0.83</td>
<td>0</td>
<td>0.48</td>
<td>0</td>
<td>0.88</td>
</tr>
<tr>
<td>Naver</td>
<td>0.16</td>
<td>0.10</td>
<td>0.27</td>
<td>0.27</td>
<td>0.43</td>
<td>0.16</td>
<td>0.25</td>
<td>0.31</td>
<td>0</td>
<td>0.42</td>
</tr>
<tr>
<td>Pandoratv</td>
<td>0.61</td>
<td>0.26</td>
<td>0.71</td>
<td>0.09</td>
<td>0.60</td>
<td>0.09</td>
<td>0.03</td>
<td>0.54</td>
<td>0.05</td>
<td>0.53</td>
</tr>
<tr>
<td>Tving</td>
<td>0.46</td>
<td>0.35</td>
<td>0.68</td>
<td>0.49</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>0.75</td>
<td>0.80</td>
<td>0.29</td>
</tr>
<tr>
<td>Wemakeprice</td>
<td>0.07</td>
<td>0.04</td>
<td>0.50</td>
<td>0.04</td>
<td>0.65</td>
<td>0.20</td>
<td>0.01</td>
<td>0.22</td>
<td>0.02</td>
<td>0.79</td>
</tr>
</tbody>
</table>

III. PROBLEM STATEMENT

In this section, we claim the problem of using the raw signatures obtained from existing signature generation algorithms with our preliminary experimental results.

In the experiments, we used real network traffic and selected ten applications that have the greatest number of flows in the data set. Note that a flow is defined as a set of packets that have the identical five tuples, including source and destination IP addresses, source and destination port numbers, and protocol type. Greater detail can be found in our previous paper [14]. Table 1 shows the applications and the number of flows used in our experiments. We randomly selected 200 flows for each application from our data set to use as training flows, and a disjoint set of 100 flows also randomly selected were selected for measuring identification performance (i.e., testing set). Therefore, the intersection of the training and testing sets should be the empty set.

Initially, we have run the three signature generation algorithms (LCS, LASER, and AutoSig), and observed LCS is at least comparable to the others for identification accuracy. Hence, we chose LCS as the signature generation algorithm in this study. We employed a set of parameters for LCS as follows. Note that the default values for the parameters were chosen based on our previous study [14].

- **The number of bytes examined (B):** The number of bytes in a flow can vary from tens of bytes to hundreds megabytes or even greater. As mentioned, payload inspection is an expensive operation. At the same time, it is not enough to stress the importance of early identification [1,4]. We set B=64 by default.
- **Minimal length threshold (T):** This parameter is used to restrict the minimal length of signatures, and any signature should not be less than round(T*B). With T, thus, it is possible to eliminate signatures that are too short. We used T=0.1 by default.
- **Coverage threshold (F):** The coverage threshold F is used to determine the minimal frequency to be accepted as a signature. If a common pattern appeared less than F-%, it is rejected to be considered as a signature; otherwise, it is accepted as a signature. We set F=10% by default.

Table 2 shows our initial results for identifying the testing set with the signatures generated by LCS. Each cell in the table shows a fraction of testing flows that matched with at least one signature for each application in each row. For example, 45% of Afreeca testing flows had one or more signatures for Afreeca, 15% of Afreeca testing flows had 1+ signatures for BitTorrent, 48% of Afreeca testing flows had 1+ signatures for Cyworld, and so forth. Therefore, the diagonal cells highlighted in the table indicate the maximum bound of true positives, while the rest of the cells show maximum false positives. In this paper, we define uTP as the highest possible true positives.
that can be achieved when the signatures are used in the identification phase (i.e., the diagonal cells), while $uFP$ is defined as the highest possible false positives that can be achieved (i.e., the non-diagonal cells).  

As can be seen from Table 2, the upper bound of true positives ($uTP$) is very low from 4% to 83%, with an average of 59%. This means that identification accuracy cannot go beyond 60%, which is unacceptable by and large. We also explored with different $B$ and $T$. $B = \{64, 128\}$ and $T = \{0.1, 0.125, 0.15\}$. No significant changes were observed with different $T$ values, while $B=128$ yielded rather poor $uTP$ as shown in the table.

This will be a critical problem if the signatures were used to identify current traffic, and this gives us our motivation for this research. Our goal is to maximize the upper TP and keep the upper FP low, so as to improve identification performance. In particular, we focus on maximizing $uTP$ in this paper. In the next section, we show how we can increase $uTP$ through our refining process.

IV. THE REFINING PROCESS

As discussed in the previous section, low $uTP$ should be detrimental to overall identification performance. In this section, we introduce our refinement process to obtain a high quality of signatures that maximize $uTP$. We first describe the first refinement step that merges raw signatures to expand the signature set. Then we will introduce a set of methods to identify an essential set of signatures out of the large set of expanded signature set after merging.

4.1 Expanding signature set by merging

There can be several reasons for poor $uTP$ with the raw signature set, and one reason we observed is that a large number of signatures in that set are failed to contain truly common patterns; rather they are somewhat specific only to a small set of flows. For example, some signatures held too specific information such as dates (it is possible if a substantial number of training flows for that application were collected on the same day), and the signatures containing any specific date would not be useful for future testing flows. To help combat this problem, we merge the raw signatures by running the signature generation algorithm (LCS) against the set of signatures as illustrated in Figure 1.

In the merge process, each signature is compared with each other to generate a new signature. If the new signature meets the length threshold, we keep the signature. By doing so, it is possible to collect more common patterns for each application. If we suppose $n$ signatures before merging, there can be $n(n-1)/2$ new signatures at max after merging. After a new set of signatures are constructed, frequency is recomputed for each signature against the training data set.

```
1 function merge_sigs()
2 input: vector<Sig> sigs;
3 output: vector<Sig> refSigs;
4 begin
5   for(i = 0; i < sigs.size(); i++) {
6     for(j = i+1; j < sigs.size(); j++) {
7         newSig = LCS(sigs[i], sigs[j]);
8         if (newSig.length() < B*T) continue;
9         refSig.add(newSig);
10     }
11   }
12 end
```

Figure 1. Pseudocode for the first step of signature refinement through merging

With the new expanded signature set, we conducted the same experiment as the one for Table 2 to see how it impacts to upper TP and FP. Figure 2 shows the results of experiments using the new refined signatures. As can be seen by the graph, the average $uTP$ for each parameter setting has greatly increased. It is above 80% for all settings except for $B = 128$ and $T = 0.15$. When we use $B=64$, we can see $uTP$ goes over 93%, which are dramatic increases compared to the results with the raw signatures. This result indicates that simply running LCS algorithm with the existing set of signatures can identify more helpful signatures for identification by maintaining signatures less specific but more common. The figure also shows $uFP$, and we can also see a high degree of possible
false positives. In this study, we more focus on increasing \(uTP\) as it limits the maximum identification accuracy.

Table 3. Number of signatures before merging

<table>
<thead>
<tr>
<th>Application</th>
<th>(B = 64)</th>
<th>(B = 128)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afreeca</td>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td>BitTorrent</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>Cyworld</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>Daum</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>Gomtv</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Melon</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>Naver</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Pandoratv</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Tving</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Wemakeprice</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td>752</td>
<td>752</td>
</tr>
</tbody>
</table>

Table 4. Number of signatures after merging

<table>
<thead>
<tr>
<th>Application</th>
<th>(B = 64)</th>
<th>(B = 128)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afreeca</td>
<td>2421</td>
<td>1868</td>
</tr>
<tr>
<td>BitTorrent</td>
<td>1884</td>
<td>1321</td>
</tr>
<tr>
<td>Cyworld</td>
<td>1441</td>
<td>1136</td>
</tr>
<tr>
<td>Daum</td>
<td>500</td>
<td>464</td>
</tr>
<tr>
<td>Gomtv</td>
<td>1836</td>
<td>1402</td>
</tr>
<tr>
<td>Melon</td>
<td>1287</td>
<td>1172</td>
</tr>
<tr>
<td>Naver</td>
<td>1048</td>
<td>884</td>
</tr>
<tr>
<td>Pandoratv</td>
<td>87</td>
<td>57</td>
</tr>
<tr>
<td>Tving</td>
<td>524</td>
<td>476</td>
</tr>
<tr>
<td>Wemakeprice</td>
<td>2159</td>
<td>1605</td>
</tr>
<tr>
<td>Total</td>
<td>13187</td>
<td>10385</td>
</tr>
</tbody>
</table>

While interesting with enhanced upper TP, one challenge remains here. Table 3 and Table 4 compare the number of signatures before and after the merge process. As you can see from Table 3, most of the applications maintained less than a hundred signatures each before merging. After merging, however, the number of signatures generated for each application greatly increased. The overall totals go from below a thousand to well above 6,000 signatures. Since the number of signatures is directly correlated to identification complexity, having such a large number of signatures will result in signature-based identification impractical, with too high memory and processing complexities. However, eliminating some signatures may have a significant impact to upper TP. Thus, the next question is how can we reduce the number of signatures without a significant decrease of upper TP? Our approach for this question is to identify core signatures that should be more beneficial for improving \(uTP\).

4.2 Identification of core signatures

To identify an essential set of signatures, so as to maintain the number of signatures manageable, we explored the following four methods:

- Random-N, which takes \(N\) signatures from the expanded signature set in a random manner;
- Top-N, which selects \(N\) signatures based on their coverage information;
- Covering subset, which identifies a subset of signatures that cover the entire training flows;
- Google Refine, which selects signatures based on clustering and merging;

A. Random-N

![Figure 3. Upper TP for Random-N](image)

![Figure 4. Upper FP for Random-N](image)
This method randomly chooses \( N \) signatures in a random fashion as core signatures, which will be used in the identification phase. Thus, it is simple and it is possible to maintain a specific number of signatures for each application by configuring \( N \).

We explored a range of \( N \)'s to see the impact of the number of core signatures for Random-\( N \). Figure 3 and Figure 4 show the results for the Random-\( N \) method. Random-\( N \) starts off with low accuracies at the low \( N \) and rise as we increase \( N \). We can see that upper TP decreases with relatively small \( N \)'s \((N \leq 50)\) compared to the results in Figure 2 with the entire signature set. As in Figure 4, we can see that upper FP also decreases accordingly.

\section{B. Top-\( N \)}

The Top-\( N \) refinement takes the top \( N \) number of signatures that appeared most frequently (i.e., highest coverage factor \( F \)) in the training flows as core signatures, which will be used in the identification phase. Compared to Random-\( N \), thus, this method will choose signatures with greater coverage preferably.

We explored a range of \( N \)'s as we did with Random-\( N \) to see its impact. Figure 5 and Figure 6 show the results for upper TP and FP. As can be seen from Figure 5, Top-\( N \) gives us more stable results than Random-\( N \), showing over 90\% of upper TP over a set of \( N \)'s. We can see that upper FP is below 85\% with small \( N \)'s \((N \leq 50)\). In particular, the figure shows that it is possible to achieve a fairly high \( uTP = 95\% \) only with 30 signatures \((N = 30)\).

\section{C. Covering Subset}

Covering subset [15] is an algorithm to find a subset that covers an entire set with the smallest number of elements, and we employ this algorithm to search the minimal set of signatures that cover the entire training flows. We implemented a greedy-based covering subset for signature refinement that identifies a set of core signatures.

Figure 7 shows the experimental results of the covering subset in diverse settings with different \( B \) and \( T \). We can see that this method also achieves fairly high upper TP even with a relatively small number of signatures as shown in Figure 8. As in Figure 8, most of the number of signatures for each application is below 10. However, due to the small number of signatures, its \( uTP \) is limited and worse than top-\( N \). In addition, it is unpredictable with respect to the number of signature for each application in the refining stage, while top-\( N \) allows us to obtain the static number of signatures (as long as \( N \) is smaller than the number of signatures in the expanded set).
To optimize, it would be possible to blend this technique with Top-N to expect a static number of signatures that could be more manageable.

D. Google Refine

Another method we used for obtaining a core set of signatures is based on Google Refine [16] that provides merging based on clustering techniques. We set different parameters of the clustering algorithms until we reached our goal of reducing the total number of signatures to 40 and under for each application. We chose 40 since Top-N with 40 signatures worked well.

Figure 9. Upper TP and FP after applying Google Refine

Figure 9 shows the experimental results of using Google Refine to refine signatures. Unlike our expectation, signatures obtained after merging with Google Refine were not very accurate, and the max upper TP is less than 62%, which is generally unacceptable.

V. CONCLUSIONS

With increasing attention to network application identification for a variety of purposes, signature-based identification based on common patterns has been widely considered despite a high degree of complexities for examining packet payloads, with the expectation of high identification accuracy. However, our observations with a set of well-known signature generation algorithms indicate that simply relying on raw signatures may lead to very poor accuracy less than 60% of true positives even in an optimal case. We refer to the upper bound of true positives as *uTP*, and in this work, we focused on maximizing the metric through a signature refining process.

The first step of the refining process is to merge signatures to dilute signatures’ specificity, so as to obtain more common patterns, and we observed that a new set of signatures after merging achieve up to 99% *uTP* (that was less than 60% without refining). However, one problem posed after performing the first step is a too large number of signatures, and one application had roughly 4,000 signatures, which is unacceptable in practice. The second step in our refining process is thus to identify an essential set of signatures that can represent the entire signatures for each application. We explored four methods, Random-N, Top-N, Covering subset, and Google Refine, for identifying core signatures. According to our experimental results, we observed that Top-N with a small number of signatures (e.g., N=30) approximates yielding a fairly high *uTP* up to 95%. Covering subset worked well, but it has a potential problem of too few signatures. We are currently working on optimizing our covering subset method. For instance, fusing covering subset with Top-N may be a good option with the predictable number of signatures.

This research is the first step towards an acceptable degree of accuracy for application identification. In this work, we focused on maximizing upper TP, and the other side of minimizing upper FP remains as a future task. Ultimately we plan to apply our refining process for identifying network applications after exploring a set of optimizations.

ACKNOWLEDGEMENT

This work was supported by the IT R&D program of MOTIE/KEIT [10041548, “240Gbps real-time automatic signature generation system for application traffic classification supporting over 95% completeness and accuracy”]

REFERENCES

SOLUTION DECOMPOSITION FOR CURRICULUM DESIGN USING ENVIRONMENT-BASED DESIGN METHODOLOGY

Hong Yi
School of Foreign Languages
Hubei University for Nationalities
Enshi, Hubei, China, 445000

Xiaoguang Deng, Yong Zeng
Institute for Information Systems Engineering
Concordia University
Montreal, Quebec, Canada, H3G 1M8

ABSTRACT
The aim of curriculum design is to provide a high-level process to guide student’s learning within a specific program. Traditionally, curriculum design is considered as an educational problem, many educators put great efforts to employ social-technical approaches to design and improve the documents of education program. However, to author’s best knowledge, there is still the lack of a systematic methodology to provide holistic solutions by well refining design requirements. This paper takes initiative to propose an approach of solution decomposition for curriculum design using an engineering design methodology: environment-based design (EBD). A case study concerning English university undergraduate teaching in mid-west University of China will be illustrated.

INTRODUCTION
The purpose of curriculum design is to provide a high-level process to guide student’s learning within a specific program (Diamond, 2011). The curriculum design process aims to produce documents of course modules, such as module descriptions, validation documents, prospectus entry, and course handbook. The curriculum design process contains several perspectives, including teaching resource allocation, course marketing and promotion, students’ final outcomes and destinations, as well as learning/teaching approaches and requirements (Conole, 2010).

Curriculum design is traditionally considered as an educational problem, many educators put great efforts to employ social-technical approaches to design and improve the course documents (Pratt, 1980; Romiszowski, 1984; Brown, 1995; Van den Akker et al., 2006). Beetham (Beetham, 2012) summarized three types of institutional approaches: radical (Goodman, 1986), incremental (Everett et al., 2012), and lean methods (Emiliiani, 2004) to curriculum improvement. However, most existing curriculum design methodologies have been proposed through the observation from case studies and applications. To our best knowledge, very few systematic approaches have been reported due to the varieties among the contents of different projects.

In this background, we take initiative to propose a novel approach for curriculum design using an engineering design methodology, called environment-based design (EBD). This methodology can provide a step-by-step procedure to deliver solutions based on three activities: environment analysis, conflict identification, and solution generation. We have presented our work in problem analysis and formalization in order to identify the critical challenges (root conflicts) in curriculum design through a case study of English education in mid-west university of China (Hong et al., 2014).

In our previous work, a rough guideline of solutions has been generated according to the critical challenges. However, the guideline is rather conceptual, which is required to further decomposed and refined in order to find out an operational action plan. According to such motivation, this paper proposes an approach of solution decomposition for curriculum design. The case of English education (Hong et al., 2014) will be taken and further analyzed in order to illustrate the approach.

This paper is organized as follows. Section 2 will summarize the literature review of the methodologies for solution generation. Section 3 will give a general introduction of EBD methodology. Section 4 will present the procedure of solution decomposition based on EBD methodology. The case study is shown in Section 5, and the conclusion and the perspective will be given in Section 6.

LITERATURE REVIEW
The facing problem gives a strong driving force for people to find out the solutions to resolve it. As a matter of fact, many solution generation approaches have been
proposed in this regard, in the past 70 years. A brief introduction to some representative approaches is given as below.

**Brainstorming VS Reverse Brainstorming**

Brainstorming is an unrestricted, informal approach to gather ideas intended to solve a problem. It is a useful approach to generating radical solutions to problems, as long as it is managed well. Brainstorming is originally proposed by Alex F. Osborn in 1953 (Osborn, 1953). He claimed that group brainstorming was more effective than individuals working alone in generating ideas. This claim was proven by many experimental comparisons (Dennis and Valacich, 1993; Laughlin et al., 2002; Steiner, 2007). However, many researchers (Lamm and Trommsdorff, 1973; Diehl and Stroebe, 1991) stated that less clear evidence is available on measures of quality, uniqueness and variety of group performance of brainstorming.

Different from brainstorming, reverse brainstorming helps in problem solving by combining brainstorming and reversal techniques. Instead of asking “How to solve or prevent this problem?”, “How could I possibly achieve the opposite effect?” are asked in reverse brainstorming (Whiting, 1958; Jay and Stevenson, 1998). This technique helps you concentrate on areas that you may take for granted and highlights anything that you might have missed. By combining these, you can extend your use of brainstorming to draw out even more creative ideas. Reverse brain-storming is a good technique to try when it is difficult to identify solutions to the problem directly. It is considered a complementary to brainstorming.

Though brainstorming is commonly used, Zeng (Zeng, 2014) outlined that it has some significant drawbacks. First, brainstorming is not a goal-oriented search, not clear to evaluate whether an idea is effective or not. Besides, the process of brainstorming throw random ideas from the panel, it is quite hard to deal with the similarity from the participants and to choose the best one among many proposed ideas (Deng et al., 2010).

**TRIZ**

“TRIZ” is the acronym for Russian expression which means “Theory of Inventive Problem Solving” in English. It was proposed and developed by the researcher Genrich S. Altshuller and his colleagues in the former USSR since 1946. TRIZ research began with the hypothesis that there are universal principles of creativity that are the basis for creative innovations that advance technology (G. S. Altshuller, 1999). If these principles could be identified and codified (G. Altshuller et al., 1998), they could be taught to people to make the process of creativity more predictable (Barry et al., 2010). In other words, in most cases, the problem a person is facing, has been already faced and creatively resolved by other people. If the universal principles are available, creative solutions are more likely to be provided. TRIZ was summarized and codified by analyzing the patterns that predict breakthrough solutions to problems form millions of patents (G. S. Altshuller and Williams, 1984). The TRIZ approach standardizes the problems have been faced and solutions have been provided as general problem-solution principles. So that a personal can connect his/her specific problem to general problem(s), and find the corresponding general solution(s), finally leads to a specific solution. It provides a very practical way to help people to solve problems in a creative manner. As such, TRIZ brings repeatability, predictability, and reliability to the problem-solving process with its structured and algorithmic approach (Orloff, 2006).

Since TRIZ is summarized and extracted from patterns, there is no mathematical foundation behind. Furthermore, in order to use TRIZ, we need firstly to represent the problem in TRIZ language, that is, classify the requirements in terms of the predefined 39 parameters in engineering (Domb et al., 1998). According to Zeng’s study (Zeng, 2014), the transition from natural language description to TRIZ-based formalization still has a large research room of investigation.

**Function-Behaviour-Structure model**

Function-Behaviour-Structure (FBS) model of designing was proposed by Gero in 1990 (Gero, 1990). It has been further developed by Gero and his collaborators in the past two decades, e.g., the definitions of function, behaviour and structure, and the steps of designing. The up-to-date definitions for function, behaviour and structure (Gero and Kannengiesser, 2004), such as:

- **Function (F):** describe the teleology of the object, i.e., what it is for.
- **Behaviours (B):** describe the attributes that are derived or expected to be derived from the structure variables of the object, i.e. what it does.
- **Structure (S):** describe the components of the object and their relationships, i.e. what it is.

In FBS modeling, the functions of an object can be decomposed into its sub-functions that compose of related sub-behaviour and sub-state. In this case, a design object can be represented hierarchically. However, it is still the lack of enough research efforts to bridge the gap between linguistic design text and formal FBS model. Wang et al (Wang et al., 2013) took initiatives to propose an algorithm to transform semantic design descriptions into a FBS model.

**Environment-based Design**

Environment-based Design (EBD) methodology is a design methodology which developed over the last two decades. The fundamental concept of EBD is that design activity becomes a recursive process (Zeng and Cheng, 1991), due to the changing characteristics of the environment. EBD came into being as a response of providing step-by-step procedures to guide a designer in
the existing environment. Under the aid of EBD, the
designer is directed to collect necessary and sufficient
information for environment analysis, to decompose a
complex problem into atomic ones and to determine the
focus at each stage of design in order to identify conflicts,
and then to investigate potential solutions for each atomic
design problem. For conclusion, EBD makes design
practices more systematic and scientific.

EBD methodology includes three independent
activities: environment analysis, conflict identification,
and solution generation. In our case of curriculum design,
environment analysis refers to the holistic analysis of
education. Conflict identification means to identify the
major challenges of improvements of current system,
whereas solution generation aims to propose the solutions
to improve the education system.

Other Approaches

Apart from the aforementioned design methodologies
and models, many other researchers presented their efforts
as well, such as affordance based design (Maier and
Fadel, 2003, 2009), axiomatic design theory (Suh, 1998),
a systematic approach (Pahl and Beitz, 1988; Pahl, 2007).
Those methods help in formalizing a problem that tells a
designer what to do to deliver a good solution.

SOLUTION GENERATION IN EBD

In EBD, the solutions are generated based on the
identified critical conflict(s) from previous activity
(conflict identification). Thus, the critical conflicts are
taken as inputs, and the solutions are taken as outputs of
solution generation process. For any critical conflict, the
re-sources/actions needed to trigger or accommodate an
action have been provided. Those compose an “abstract
solution” in conflict identification activity.

In solution generation process, we start from this
abstract solution, to find the knowledge which maps the
abstract solution to the desire performance. Then, we
evaluate the solution. If it resolves the primitive
conflicts(s), then the conflict has been resolved, we move
to the next conflict. If the solution resolves non-primitive
conflict(s) only, it might be one potential conflict or raise
other pential conflicts, needed to be refined in later stage.
The concepts of primitive and non-primitive conflict(s)
will be presented as below.

As a matter of fact, at the conceptual design level,
solutions can be proposed. The recursive process of
solution generation is shown in Error! Reference source
not found..

Fig. 1 Recursive logic in solution generation activity

Primivtie VS Non-primitive Conflict

According to the structural/functional complexity of
component(s) involved in the design environment, a
conflict can be identified as a primitive conflict or non-
primitive conflict. In EBD, a primitive design is where
knowledge is available or can be systematically
discovered to resolve a conflict. Hence, it depends on the
designer’s experience and background. This is the major
distinction between the concept of “atomic” in the theory
of atomic design (Kuhlman et al., 2003). An “atomic”
object is the fundamental unit which cannot be physically
decomposed. It is independent with desinger’s knowledge
and experience, whereas an experienced designer can
have a more complex primitive design than a novice one
may have.

In EBD, recursive resolution of a complex design
problem can be conducted through environment
decomposition (Zeng, 2014), which will identify the key
conflict to start with. If it is non-primitive, we always try
to transform the non-primitive environment component to
primitive, and then find a solution from there.

Rules to Resolve a Conflict

EBD proposes two rules to resolve the primitive and
non-primitive critical conflicts respectively. For a given
critical conflict:

Rule 1: If it is a non-primitive conflict to the
designer, regardless of its other characteristics, the
designer needs to transform it to promise conflict(s) that is
resolvable to the designer. There are two ways to make
this transformation: acquiring knowledge through
experimentations or from experts, or finding knowledge through environment decomposition. By acquiring knowledge through experimentation or from experts, the conflict has been transformed into a primitive conflict. By finding knowledge through environment decomposition, the conflict can be decomposed into primitive and non-primitive one(s).

This is also named “recursive resolution strategy” in (Zeng and Cheng, 1991). The core of the recursive resolution strategy consists of a requirement decomposition operator and structure recomposition operator. The recomposition operator embodies the approaches to dealing with the conflicts between the newly generated solution and the existing solutions and problems, which leads to the reformulation of the original design problem. For the environment decomposition process underlined in Fig. 2, it has been expanded to a very straightforward process as

Fig. 3 shows. This decomposition process can be recursively applied until no more non-primitive conflict(s) exists.

In EBD, environment decomposition follows a structure operation, denoted by $\oplus$. Eq. (1) shows a general structure operation, where $\bigcup$ is a union of environment components, and $\otimes$ represents interactions between environment components $E_1$ and $E_2$.

$$\oplus E = \oplus E_1 \bigcup \oplus E_2 \bigcup (E_1 \otimes E_2) \bigcup (E_2 \otimes E_1)$$ (1)

**Rule 2**: For a primitive conflict, more sub-rules are defined according to its other characteristics:

**Rule 2.1**: If it is an active conflict, no matter it is determined or potential, it is relatively easy to provide a solution: defined the knowledge, then we can provide an answer.

**Rule 2.2**: If it is a reactive but potential determined conflict, we need to find the corresponding active conflict (missing an action to trigger such action), and then treat it as an active conflict. For determined conflict, more options are available. Since the input for the primitive design process is a design conflict, which is equivalent of physical contradiction in TRIZ. Therefore, similar actions to TRIZ can be applied: change properties in time/space, separate action in time/space, combine actions, and remove actions. New components could be added as an approach to create resources. If those do not help much, we can find its parent active conflicts to redefine new knowledge, and therefore provide new solution in a new perspective.

The process of Rule 2, including Rule 2.1 and Rule 2.2, is illustrated in Fig. 4.

**Solution Generation Strategy**

The overall solution generation strategy, combining Rule 1 and Rule 2, is shown in Error! Reference source not found.. It will result in one of the two situations:
1) adjustment of existing design parameters need to be optimized; and
2) new component needed to connect unsolved and connected physical effects in the ROM diagram.

It must be noted, however, the methods developed so far are all heuristic. It is the author’s belief that this is one of the critical stages for design creativity to happen. One advantage of the EBD is to guide designers to rapidly identify this focus so that creative solutions may be generated.

CASE STUDY

In this section, we will present how to employ EBD’s solution generation strategy for curriculum design through illustrating a case study. This case study concerns the design of undergraduate program in English study in mid-west university of China.

Due to the impact of globalization, more and more universities in China dedicate to improve the quality of undergraduate education in English studies. In this case, we emphasize on such curriculum design for the universities located in mid-west China, where the economics and transportation facility are relatively less developed comparing to other regions.

In our previous study (Hong et al., 2014), through a systematic curriculum environment analysis using EBD methodology, we have identified three major solutions faceting the critical conflicts for the improvement of current education system, such as:

Solution 1: cultural immersion to foster students’ knowledge of English culture,

Solution 2: language immersion to foster students’ basic language skill, and

Solution 3: link into practices to foster students’ basic abilities.

In this case, we can notice that both three critical conflicts are non-primitive. According the recursive process in Error! Reference source not found., we only obtained very abstract solutions in our previous study. The knowledge from experts is not immediately available. Therefore, the aim of this paper is to further decompose the non-primitive conflicts and then find out the design knowledge to propose quantitative solutions.

Decomposition to refine solution 1
In EBD, mathematically, the abstract solution \( S_1 \) can be represented as follows.

\[
S_1 = \text{language immersion},
\]

\[
E_1 = \text{students' knowledge of English culture},
\]

\[
S_1 = E_1 \otimes \text{foster} \otimes E_2,
\]

where \( \otimes \) (interaction).

According to Fig. 3, in order to find out the knowledge to the refine the solution \( E_1 \) (language immersion), we needs to decompose the structure of \( E_2 \). The sub-component of \( E_2 \) is shown as below.

\[
E_2 = E_{21} \cup E_{22} \cup E_{23} \cup E_{24} \cup E_{25},
\]

\[
E_{21} = \text{history} \land \text{politics},
\]

\[
E_{22} = \text{literature} \land \text{art},
\]

\[
E_{23} = \text{social life},
\]

\[
E_{24} = \text{geography} \land \text{environment},
\]

\[
E_{25} = \text{economics} \land \text{technology}.
\]

where \( \otimes \) (interaction), \( \land \) (conjunction), \( \lor \) (union).

According to Eq. (1), the operator of decomposition is shown as follows.

\[
(\otimes E_1) = \otimes(EE_{21} \cup EE_{22} \cup EE_{23} \cup EE_{24})
\]

\[
= \oplus EE_{21} \cup \oplus EE_{22} \cup \oplus EE_{23} \cup \oplus EE_{24}
\]

\[
\cup(EE_{21} \otimes EE_{22}) \cup (EE_{22} \otimes EE_{21})
\]

\[
\cup(EE_{21} \otimes EE_{23}) \cup (EE_{23} \otimes EE_{21})
\]

\[
\cup(EE_{21} \otimes EE_{24}) \cup (EE_{24} \otimes EE_{21})
\]

\[
\cup(EE_{22} \otimes EE_{23}) \cup (EE_{23} \otimes EE_{22})
\]

\[
\cup(EE_{22} \otimes EE_{24}) \cup (EE_{24} \otimes EE_{22})
\]

\[
\cup(EE_{23} \otimes EE_{24}) \cup (EE_{24} \otimes EE_{23})
\]

where \( \otimes \) (structure), \( \otimes \) (interaction), \( \lor \) (union).

Form Eq. (2-4), solution 1 can be refined as below:

Solution 1': cultural immersion to foster students’ knowledge of history and politics, literature and art, social life, geography and environment, as well as economics and technology, and their mutual impacts in English speaking countries.

**Decomposition to refine solution 2**

In EBD, mathematically, the abstract solution \( S_2 \) can be represented as Eq. (5).

\[
E_i = \text{language immersion},
\]

\[
E_3 = \text{students' basic language skill},
\]

\[
S_2 = E_1 \otimes \text{foster} \otimes E_3,
\]

where \( \otimes \) (interaction).

The \( E_3 \) is decomposed as follows.

\[
E_3 = E_{31} \cup E_{32} \cup E_{33} \cup E_{34},
\]

\[
E_{31} = \text{listening},
\]

\[
E_{32} = \text{speaking},
\]

\[
E_{33} = \text{reading},
\]

\[
E_{34} = \text{writing}.
\]

where \( \otimes \) (interaction), \( \land \) (conjunction), \( \lor \) (union).

Accordingly, the operator of decomposition \( E_3 \) is shown as follows.

\[
(\oplus E_3) = \oplus(EE_{31} \cup EE_{32} \cup EE_{33} \cup EE_{34})
\]

\[
= \oplus EE_{31} \cup \oplus EE_{32} \cup \oplus EE_{33} \cup \oplus EE_{34}
\]

\[
\cup(EE_{31} \otimes EE_{32}) \cup (EE_{32} \otimes EE_{31})
\]

\[
\cup(EE_{31} \otimes EE_{33}) \cup (EE_{33} \otimes EE_{31})
\]

\[
\cup(EE_{31} \otimes EE_{34}) \cup (EE_{34} \otimes EE_{31})
\]

\[
\cup(EE_{32} \otimes EE_{33}) \cup (EE_{33} \otimes EE_{32})
\]

\[
\cup(EE_{32} \otimes EE_{34}) \cup (EE_{34} \otimes EE_{32})
\]

\[
\cup(EE_{33} \otimes EE_{34}) \cup (EE_{34} \otimes EE_{33})
\]

where \( \otimes \) (structure), \( \otimes \) (interaction), \( \lor \) (union).

Form Eq. (5-7), solution 2 can be refined as below:

Solution 2': language immersion to foster students' language skill in listening, speaking, reading, and writing, and their mutual impacts.

**Decomposition to refine solution 3**

In EBD, mathematically, the abstract solution \( S_3 \) can be represented as below.

\[
E_i = \text{link into practices},
\]

\[
E_3 = \text{students' basic language skill},
\]

\[
S_3 = E_1 \otimes \text{foster} \otimes E_3,
\]

where \( \otimes \) (interaction).

The \( E_3 \) is decomposed as follows.

\[
E_3 = E_{31} \cup E_{32} \cup E_{33} \cup E_{34},
\]

\[
E_{31} = \text{listening},
\]

\[
E_{32} = \text{speaking},
\]

\[
E_{33} = \text{reading},
\]

\[
E_{34} = \text{writing}.
\]

where \( \otimes \) (interaction), \( \land \) (conjunction), \( \lor \) (union).
\[ E_4 = E_{41} \cup E_{42} \cup E_{43}, \]
\[ E_{41} = \text{teaching}, \]
\[ E_{42} = \text{research}, \]
\[ E_{43} = \text{management}. \]

where \( \otimes \) (interaction), \( \land \) (conjunction), \( \cup \) (union).

Accordingly, the operator of decomposition \( E_4 \) is shown as follows.

\[
(\otimes E_4) = \otimes (EE_{41} \cup EE_{42} \cup EE_{43}) \\
= \otimes EE_{41} \cup \otimes EE_{42} \cup \otimes EE_{43} \\
\cup (EE_{41} \otimes EE_{42}) \cup (EE_{42} \otimes EE_{41}) \\
\cup (EE_{41} \otimes EE_{43}) \cup (EE_{43} \otimes EE_{41}) \cup \ldots
\]

where \( \otimes \) (structure), \( \otimes \) (interaction), \( \cup \) (union).

Form Eq. (8-10), solution 3 can be refined as below:

Solution 3: link theory into practices to foster students’ basic abilities in teaching, research and management.

RESULTS AND DISCUSSIONS

In the analysis above, we have refined the knowledge to propose more detailed solutions: an action plan. The knowledge is presented in Table 1. The details of action plan are given as below.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Action plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>foster students’ knowledge of history and politics, literature and art, social life, geography and environment, as well as economics and technology, and their mutual impacts in English speaking countries.</td>
<td>increase culture introductory course, improve cultural immersion experience</td>
</tr>
<tr>
<td>foster students’ language skill in listening, speaking, reading, and writing, and their mutual impacts connection relation</td>
<td>increase number of qualified language personnel, improve quality of course materials, control appropriate number of students</td>
</tr>
<tr>
<td>foster students’ basic abilities in teaching, research and management</td>
<td>simplified theoretical course, improve practical assignments and projects</td>
</tr>
</tbody>
</table>

Culture introductory course

It is very challenging to determine real consensus on what role culture should play in the foreign language curriculum; there are no simple answers as to how to teach culture in undergraduate program on English studies. The first reason is that culture is a highly complex, elusive, multilayered notion. For some, it may refer to an appreciation of good literature, music, art and food. For anthropologists and other behavioral scientists, it is the full range of learned human behavior patterns. For example, as Tylor stated, culture is that complex whole which includes knowledge, belief, art, law, morals, custom, and any other capabilities and habits acquired by man as a member of society (Tylor, 1871). It is difficult to categorize and classify. The second reason is that culture is the traditional domain of anthropologists and not the field of language teachers. Even though language teachers always assert that they teach language and culture, actually they tend to focus on language teaching, especially in the beginning and intermediate English language courses, leaving culture to next period or at the periphery. Cultural perspective is difficult to manage in elaboration to most of language teachers.

As the increasing translingual and transcultural communication, culture has got an established place in foreign language curricular discussions. At first, introductory courses about English language and culture have been provided. For example, there is an optional course named “An introduction to western culture”, which is a diachronic introduction to culture from ancient Greece, ancient Rome to modern age, including language, literature and so on. As a supplement to the cultural introductory course, some internet-based courses are also provided in order to give more cultural materials, including an abundance of texts, images, and videos, to English majors to read. By doing this, English majors are presented not only the factual information but also a close understanding of how culture shapes the behaviors and interaction among peoples.

Furthermore, the culture introductory courses should be changed from optional to compulsory in order to get due attention from both students and teachers. The class size should be the same as the oral English course. In this way can the objectives of the course completed.

Assignements of cultural immersion experience

Since culture primarily relates to the way people interact with each other, it is not possible to adequately observe it in a laboratory setting. Just imagine how much more would be learned about the actual patterns of interaction of a typical American family by living in their home rather than asking one of the family members in a college or university office. Therefore, anthropologists have discovered that the best way to really get to know another society and its culture is to live in as an active
participant rather than simply an observer. Language learner can learn culture in the same way as anthropologists.

The assenonment of “Cultural immersion experience” was developed by (Canfield et al., 2009) over a 12-year period. More than 1,400 students from 14 different colleges and universities participated with the goal of development or enhancement of skills for working more effectively with persons from different cultural background. There are three formats in their “cultural immersion experience” assignment, namely, traditional on-campus courses, study-abroad program courses, and 100% Internet-based courses. While in our case, nearly all the English majors do not have chance to have study-abroad course due to the financial reasons. Therefore “cultural immersion experience” assignment will be modified in the following way. Only the other two formats can be used in our context. English majors can be required to design, propose, and participate in a personalized cultural immersion experience. Students are instructed to conceive an immersion experience, including any potential safety, legal, or ethical issues (such as privacy consideration during the interactions with other people in the course of the immersion experience). After the approval of the students’ proposal, students can begin their cultural immersion experience either personally during the weekend or on holidays or by using the Internet. As an evaluation, English majors are required to hand in their assignments in the form of paper report.

By doing so, English majors can make good use of the knowledge from the cultural introductory courses and are placed into a cultural context among a group of people with whom they have little or no prior experience or familiarity at all. They will have a better understanding of cultural relevant factors, such as race, ethnicity, religions and so on.

By immersing themselves into another culture, English majors will figure out how intercultural interaction takes place, how social identities are part of all interaction, how their own perceptions of other people and other people’s perceptions of them influence the success of communication. In this way, intercultural communication is not the transmission of information about the foreign countries.

Qualified language personnel

As we know, nearly all the foreign language departments in China have been employed foreign language teachers whose native language is English. There is no exception in the case university too. They are responsible for the compulsory courses such as oral English, English listening and speaking, English pronunciation, and so on. In one word, they are predominantly used as “oral English” practice aids of facilitating and improving the English majors’ listening and speaking skills. The mechanics and more technical aspects of English (such as grammar, sentence structure, writing and reading skills) are essentially delegated to the Chinese English teachers only (although there are a few exceptions). Therefore the qualified teachers are especially important to English majors.

Teachers first must have near-native proficiency in the oral and written forms of the English language. English majors need a solid model on which they can build their own target language skills. The only model the student will have at the beginning is the teacher. From the first day, when the student listens to the teacher and attempts to repeat the vocabulary taught, even sees the handwriting on the blackboard, he/she is developing receptive language and collecting knowledge about the language. Therefore, the level of accuracy, the accent, the vocabulary and many other aspects of the target language become very important in the language development of English majors. The teacher must be very conscious of such factors and ensure that he or she uses the target language in ways that enable the students to develop their own language skills with both speed and accuracy.

Second, teachers must be trained with updated teaching methodology and be ready to do some teaching feedbacks and some teaching research. Since the 1970s, many researchers, for example, Howsam (Howsam, 1976) and Taylor (Tylor, 1871) have demonstrated that cross-cultural experiences should be an integral part of teacher education programs. For example, Ference and Bell (Ference and Bell, 2004) once reported a cross-cultural immersion in the U.S. which aims to raise cultural awareness of preservice teachers towards Latino ESOL students. However, in the university, such kind of qualified language teachers are badly needed. Therefore, to recruit teachers from abroad or with study-abroad experience is more important to English majors. To train the young English teachers with teaching methodology and research methodology is another method to improve their qualification, such as sending them to key universities in China and even abroad to get further training or education. Since 2010, at least two English teachers have been chosen to send abroad every year and five English teachers have been successfully enrolled in other universities in China to pursue their master’s degree or doctoral degree every year.

To be a qualified personnel in Language teaching, for example, a teacher who teaches culture, he or she does not have to know everything about “the target culture”. It is impossible. In fact there are many cultures associated with a particular language. Take our China as an example, we Chinese people speak Chinese. Within our country, there are many variations on beliefs, values and behaviors from ancient time to modern age. It is too complicated. The qualified teacher’s task is to design a series of activities to enable English majors to discuss and draw conclusions from their own experience of the target culture. Factual information should be given to students and comparative
analysis between learners’ own culture and the target culture should be assigned to students. The teacher is to help learners ask questions and to interpret answers. In other words, a qualified language teacher needs to assume many different roles, such as, an effective task- assigner, material- provider, advisor, helper, and so on. 

**Effective course material**

In university English teaching, another potential obstacle is the quality course materials. As we know, nearly all the colleges and universities in China adopt the same kind of teaching materials. On one hand, it aims to provide with equal educational chances, for the students are reading the same text book, no matter you are in the capital city of Beijing or you are in the remote mountainous city. On the other hand, there are differences between students and universities. By using the same materials, some students from some university may have difficulty in understanding all of them, especially to those English majors in the case university. Compared with the English majors in key or national-government universities, the students in the case university are less proficient in their language competence. Therefore, choosing appropriate materials for English majors is very important.

At first, teachers should be able to choose from the standard text books. Then, teachers will need time to adapt or develop their own materials to supplement those obtained elsewhere. Lastly, authentic materials are also important for language learners. Sometimes, the materials may have the advantage of reflecting the educational philosophy and curricular trends which prevail in English-speaking countries, such as the United States. By using these additional materials, teachers should be very careful.

For instance, the introduction to a textbook on British Cultural Studies produced in Romania describes the expectations of the authors for the learner, which has less concerned about making you learn information by heart than with encouraging you to process the information contained here (Byram et al., 2002). Similarly, when dealing with authentic materials, teachers and students should be always critical. For authentic materials often include audio recording and a variety of written documents and visuals such as photographs, diagrams. When presenting students with texts and visual materials, learners need to acquire concepts and abilities of analyzing materials more than factual information, especially there are contrasting views. All in all, authentic materials should be presented in their context or text of situation. For example, when, where the text is produced, to whom the text is intended, what is the potential political, religious or cultural viewpoint contained in the text. Materials from different origins with different perspectives should be used together to make English majors compare and to analyze the materials critically. It is always more important that English majors can acquire skills of analysis than factual information from appropriate learning materials.

**Appropriate scale of class**

The number of students in any given class is determined by the school’s student/teacher ratio. In western countries, class sizes in public school immersion programs generally range from 20 to 35. Obviously, a small class size is desirable. But in the case university, the number of students in other departments, such as medical school, math school is 40. In foreign language school, the number is between 30 and 35. However, due to the limited number of classroom and teachers, the students from different classes are put together to attend lessons, except for oral English, listening. Sometimes, as to some optional courses, such as “western culture introduction”, the students of one grade are presented in one big classroom with one teacher talks about the culture. The number of students is over 100. Therefore, the actual class size is much bigger, which is not good for language learning.

**Simplification of theoretical courses**

During the learning process at the university, even at the very beginning, English majors should develop the ability of understanding, reflecting and critical thinking. For example, in the basic subjects, such as conversation classes, writing, grammar, students start with comparing, reflection, evaluation, problem-solving exercises. When it comes to the courses, such as linguistics, stylistics, translation and literature, English majors can also be influenced by the research methodologies in addition to development of language skills and relevant specialized knowledge.

Generally speaking, the research skills for English majors to master including the following abilities: to understand the need and essence of research skills in teaching profession, the use of various research methods for academic and classroom background; to find and select literature; to analyze various sources on a topic and interpret various views and theories; to evaluate peer presentations and outcomes, and so on. Some of the above abilities can be gained from basic language courses and some subject learning, while others are needed to be taught in special courses such as academic paper writing course.

**Practical assignment and courses**

In our case, there is a teaching practice for English majors in the fourth year. It lasts about two months. Generally speaking, all the English majors will be sent to different junior and senior middle schools and vocational schools to be student teachers. They will be assigned to learn how to do class management and how to teaching English with the help of the mentors.

As to research skills, practice assignments are also necessary to be the supplement to the theoretical courses.
English majors should be assigned different tasks to make sure that they have mastered all the basic skills on researching, such as searching for relevant and necessary materials; selecting and organizing literature, interpreting theories, viewpoints and data, obeying research ethics and so on. At the same time, all the English majors should be encouraged to be involved in different kinds of conferences and the various sections of the conference: they present the results of their research during their teaching practice, they participate in the discussions, ask questions, solve problems, demonstrate their abilities. Students also acquire organizational skills. If the undergraduate students can be involved in the international conferences, they will have many opportunities to work starting with invitation and planning, welcoming and catering, being together in the conference and afterwards activities. To put it into another, they will learn to be responsible from beginning to end. What is more, the additional value of participating in the conference is the experience of being in the international setting, development of intercultural communication, learning through doing in authentic situations.

CONCLUSION

In this paper, in contrast of traditional approach, we have proposed an approach to improve curriculum design in language university teaching through an engineering design methodology: environment-based design (EBD). Following our previous work, three obtained abstract solutions have been decomposed to find out the knowledge on primitive components. An action plan, including six refined solutions, has been developed according to the decomposition procedure. A case study, concerning English undergraduate education in mid-west university in China, was presented.

In our future work, we attempt to develop an EBD-based framework of curriculum design. Such framework contains current methodologies and future trends of related areas, more real examples will be studied to validate the effectiveness of the proposed framework.

AKNOWLEDGEMENT

Hong Yi’s research is supported partly by the “Young Teachers Overseas Exchange Program” which is granted by education department of Hubei province, China, in 2012, and partly by Institute of Ethnic Studies, Hubei University for Nationalities, Enshi, China. It is also one of outcomes of a research program numbered 2012JY007.

REFERENCES


Everett, J., Macgregor, G., Mohamed, R., and White, B., 2012, "An incremental approach to technology-supported curriculum design and approval." Paper
presented at the Proceedings of the IADIS International Conference WWW/Internet 2012.


EXPERT SYSTEM FOR MAINTAINING A HEALTHY LIFE-STYLE

Ayako Fukunaga, Yujiro Ogata, Tad Gonsalves
Department of Information & Communication Sciences
Faculty of Science & Technology,
Sophia University,
Tokyo, Japan

ABSTRACT
Life-style related diseases are constantly increasing in modern society. We are increasingly becoming aware of the need for balanced diet and exercise to maintain a healthy and fit body. So, when it comes to eating we pick and choose dishes depending on their nutrition, cholesterol and calorie content. However, doing such a thing for each and every meal each and every day is a tedious process. In this paper, we describe an expert system that helps the user in maintaining a healthy life-style by deciding on the type and amount of food and exercise he/she needs to choose to stay fit and healthy. The system is developed using the Web Flex Expert System shell and is deployed online for easy access and consultation.

INTRODUCTION
We all need some sort of assistance and guidance on almost a daily basis in maintaining a healthy life-style. This guidance should take into consideration balanced diet, nutrition level, calories, exercise, rest, etc. Hiring professionals is costly; reading health-related books and keeping a personal diary is tedious and time consuming. In order to solve these problems, we have developed a healthy life-style maintenance system that can be easily accessed and used by anyone. It is intended to provide life-style and health related information. For example, if we consider just the cholesterol module, it is possible to easily know the cholesterol levels of any food item by selecting that item from a list of food items displayed by the system. Similarly, the user is instructed to avoid a particular food item if his/her aim is to cut back on cholesterol. The system also gives advice on the optimal level of exercise the user needs to do taking into consideration the user’s age, sex, weight, height, etc.

The major difficulty in building an expert system is the knowledge acquisition bottleneck. This is chiefly because the knowledge possessed by the domain experts and professionals is, to a large extent tacit. Abiddi et al. (2005) have proposed a knowledge creation info-structure to acquire and crystallize the tacit knowledge of healthcare experts. Some examples of expert systems in health care are found in Blacker and Oborne (2003), Hata et al. (2009) and Masri et al. (2008). Hayes et al. (2010) propose a system to diagnose and treat diabetes; our system, on the other hand, advises patients and general users on how to maintain a healthy life-style so as to avoid high cholesterol. It also concentrates on regular exercise to maintain health. Research related to the design of expert systems dealing with nutrition and exercise is found in Malik et al. (2008) and Wang et al. (2005).

The basic components of an Expert System are the knowledge base and the inference engine. It is composed of four distinct knowledge bases: Cholesterol, Exercise, Glycemic Index (GI) and Food items. The forward-chaining inference engine of the expert system connects the four distinct knowledge bases to give a suitable advice to the user. The system is made online and is easily adaptable to individual needs.

KNOWLEDGE BASE
The Expert System knowledge base is constructed in the form of machine-readable facts and rules and is classified in the following three categories:

Cholesterol
In this section, we describe the relationship between cholesterol levels and health of the individual. Cholesterol is a fat-like substance chiefly found in food from animals, like dairy products, eggs, and meat. The body also produces cholesterol since it is needed for maintaining the regular bodily functions. The cell walls, or membranes, need cholesterol in order to produce hormones and the bile acids that help to digest fat. However, the body needs only a limited amount of cholesterol to meet its needs. When there is too much cholesterol, health problems such as heart disease may develop.

However, if the cholesterol level is too low, it has other adverse effects on the body. About 100 ~ 120 gm
cholesterol is present in the body of an adult, part of which is continuously replaced by the newly formed cholesterol in the liver so as to maintain the biological functions. In order to maintain health, cholesterol level of 1.0-1.5 gm is needed on a daily basis, 70-80 percent of which is synthesized in the liver, and the rest of the 20-30 percent is obtained from the diet.

**Exercise and METS**

METS (METabolic equivalentS) is a unit to measure physical activity and exercise. The state of rest (in which the physical activity is minimum) corresponds to one METS. A typical METS represents the relative intensity (or calorie consumption) when carrying out the activity or compared to that in the state of rest. The amount of physical activity and exercise measured in METS is obtained by multiplying the time period of the physical activity. For example, when performing 30 minutes of physical activity of 6 METs, the (Mets time) is: $6 \times 1/2 \text{ hour} = 3 \text{ exercises}$.

**Glycemic Index (GI)**

Glycemic Index (GI) is the measure of the amount of carbohydrates that enters the blood in up to 2 hours after the intake of food. It refers to the degree of elevation of postprandial blood glucose level. GI basically shows the absorption degree of carbohydrates in food. The smaller the value of GI, the slower is the rise of the glucose level in the blood stream.

The classification of GI into low, medium and high values is not yet rigorously defined. However, according to a study conducted at the University of Sydney, a GI value of more than 70 is considered to be high, 55–70 is considered to be medium and less than 55 is considered to be low (http://www.glycemicindex.com/foodSearch.php).

**Food Ingredients database**

Six different categories of ingredients are classified and stored in this database: rice, bread, pasta, noodles (including beans); vegetables (including mushrooms); meat, seafood (including processed products) and dairy products.

Following are some of the ingredients in each food item that are stored in the Food Ingredients database: Energy, water, protein, fat, carbohydrate, ash, sodium, potassium, calcium, magnesium, phosphorus, iron, zinc, copper, manganese, retinol, β-carotene equivalents, retinol equivalents, vitamin D, vitamin E, vitamin K, vitamin B1, vitamin B2, niacin, vitamin B6, vitamin B12, folic acid, pantothenic acid, vitamin C, cholesterol, and GI. The data base currently holds 241 different kinds of ingredients.

**RESULTS AND DISCUSSIONS**

The Health-check Expert System is implemented using the Web Flex Expert System Development shell. Flex is a Prolog-based knowledge specific programming language. It stores data and information in the form of frames and arranges the If-then rules in rule-sets. Rule-sets are evoked and inference using the forward-chain inference. Some of the features and interfaces of the Expert System are depicted in this section.

**Cholesterol**

Fig. 1 shows the drop down menu of the Foot items the user can choose from. For instance, if the user selects the “sukiyaki” food item, then the system displays the level of cholesterol contained in each ingredient of sukiyaki dish as shown in Fig. 2.
**SUkiYAKI**

Cholesterol: 454.8 mg  
(Beef: 240 mg; Eggs: 420 mg)  
This item has a high level of cholesterol.  
Try not to exceed 300 mg per day.

---

**Fig. 2. Cholesterol levels of sukiyaki ingredients**

**Food groups database**

Different food items are grouped into food groups. The user may choose any of the food items from the pull-down menu. For example, in the food items menu shown in Fig. 3, the user has selected “sea food”.

---

**Fig 3. Food groups**

The system then displays another pull-down menu displaying the individual items from the chosen food group. For example, when the sea food menu item is selected, the system further displays the following list of sea food ingredients.

---

**Fig 4. Selecting individual ingredients**

Finally, the system displays all the constituent nutrients in the selected food item (shown only partially in Fig. 5).

---

**Fig 5. Individual nutrients in a food item**

SUKIYAKI  
Cholesterol: 454.8 mg  
(Beef: 240 mg; Eggs: 420 mg)  
This item has a high level of cholesterol.  
Try not to exceed 300 mg per day.
CONCLUSION

Life-style related diseases are found to increase globally every year. A lot of people are looking for low cost and easily available consultation on improving and maintaining a healthy life-style. This paper deals with the design of a food and exercise consultation Expert System to maintain a healthy and fit life-style. It is composed of four distinct knowledge bases: Cholesterol, Exercise, Glycemic Index (GI) and Food items. The forward-chaining inference engine of the expert system connects the four distinct knowledge bases to give a suitable advice to the user. The system is made online and is easily adaptable to individual needs.

As future research we want to extend the consultation services to include many other life-style and stress related disorders and diseases.

REFERENCES


PSYCHOLOGICAL, MATHEMATICAL, AND PEDAGOGICAL ANALYSIS
OF VIDEO STREAMS FOR MEASURES OF STUDENT ENGAGEMENT

Scott Brande¹, Eric Gampher², Arie Nakhmani³, Scott Snyder⁴
¹Department of Chemistry – CAS, UAB
²Department of Psychology – CAS, UAB
³Department of Electrical and Computer Engineering – ENG, UAB
⁴School of Education – CEA, UAB
Birmingham, Alabama 35294, USA

ABSTRACT

The increasing use of multimedia in and out of the classroom is in part due to the recognition of critical roles it plays in the learning process. Video media has the potential to capture and focus student attention on information relevant to the academic subject in ways that other media do not. The phenomenon of student engagement is a beginning point in an extended process of student learning, and without engagement, student learning will be limited. We have initiated preliminary studies to explore student engagement through video media in a freshman level earth science course. In these studies, we find, not unexpectedly, that in a large sample of video snips, there is wide variation in the frequency with which students recall video at a later time. Why are some video snips recalled with greater frequency? These findings, and the questions they raise, suggest that a complex web of variables will need to be disentangled to better understand this initial phase of student learning. If associations could be discovered between videos that are recalled with higher frequency and objective measures of information encoded in the digital video streams, then this association could potentially provide a way to identify in advance videos that are more “memorable”. The selection of videos by predictive criteria, as well as criteria for generating good educational videos, may therefore contribute to enhancing student engagement and thus academic performance.

INTRODUCTION

Student engagement is a critical variable in such diverse educational issues as attendance, graduation rates, and overall academic success (Thatcher, 2014). As with any other human-related behavior, student engagement is a complex phenomenon that spans many axes of understanding, from academic to social environments. The National Survey of Student Engagement (NSSE, 2014) has surveyed over 1,000,000 students since 2000 for information at the macro level about student activities and participation, and provides institutional and faculty guidance on best practices. Studies of student engagement vary widely in their focus, from attitudes about the importance of faculty behavior and attitude towards positive undergraduate outcomes (Umbach and Wawrzynski, 2005), to content mastery in specific courses (Redish and McDermott, 1999). In a review of the literature, Handlesman et al. (2005) found “...general agreement that engaged students are good learners and that effective teaching stimulates and sustains student engagement...”.

What factors play significant roles in engaging students in an educational setting? Particular components of student engagement have been generally recognized - attendance, attention, student interactions with peers and the teacher, and, in general, the participation in “active learning” (Hacker and Niederhauser, 2000; Thatcher, 2014). “Active learning” means in general the participation of the student in a focused activity. Many educators advocate for replacing a “traditional” lecture dominated by the instructor in a one-way delivery channel by multiple channel delivery with active learning (Center for Research on Learning and Teaching, 2014).

Among the diversity of channels through which students receive and interact with information is digital video. Digital video use in the academic environment has significantly increased with the growth of online video archives (e.g., YouTube) and the build-out of internet and technology infrastructure required for classroom video viewing (Kaufman and Mohan, 2009). As a consequence, questions about how best to select and use digital video in the classroom have proliferated (e.g., Brande and Arslan, 2013), and some general guidelines have been developed (e.g., Clark and Mayer, 2011; Derry, 2005; Mayer and Moreno, 2003). Of particular relevance to studies of video multimedia use is the dual-channel theory of multimedia learning proposed by Mayer and Moreno (2003). They propose that because visual and auditory sensory input arrives through two distinct “channels”, and that channel capacity is ultimately limited, under some conditions, cognitive processing may be reduced, thus impeding learning.

Results of research in the cognition and psychology of meaningful learning indicate the importance of three
Because of the increasing recognition of the potential of digital video for enhancing student engagement and learning, we have recently begun to explore the availability and applicability of what we consider to be pedagogically relevant video to an introductory (freshman level) earth science course. We have found that among the vast YouTube archive, an increasing number of channels are accumulating geoscience video content, much of which is applicable to common topics covered in our course. Might short digital video snips viewed throughout a course lead to greater student engagement and performance? And if so, how might one more effectively and efficiently select video for the potential gains we would like our students to achieve? We are certainly unable to answer this broader question at the present time. Rather, the studies we report herein are of a preliminary and much more restricted nature.

a) Do students recall within a specified period of time instructional video snips with different frequencies? That is, are some videos, “more memorable”, recalled to a greater degree than others?

b) If some videos are more memorable, can we design future studies to generate objective properties of the video stream data that correlate with student retention and recall?

Our focus is on a student’s ability to recall video snips previously viewed during classroom instruction. Therefore, our definition of student engagement is the number of videos a student recalls and records on an examination within a specified period of time, of all those previously viewed.

MATERIALS AND METHODS

Discovery and Selection of Video

Discovery: A domain expert (Brande) searched YouTube for geoscience content relevant to topics discussed in his introductory earth science course. Video discovery on YouTube was an extended process, as the vast majority of videos retrieved from relevant searches were judged unsuitable for classroom presentation (significant factors that excluded video included, among others, unreliability of source channel, poor quality and insignificant factual content). Videos for potential inclusion were viewed and judged from reliable source channels (e.g., government agencies, scientific organizations, educational and non-profit institutions, individual researcher laboratories, major news organizations). Other significant criteria for potential inclusion were: a) short length (<3 minutes), b) factual content that aligns with predetermined lecture material, c) higher quality video display (>=640x480 pixel resolution), and d) overall attractiveness to foster student engagement (arbitrarily determined by the instructor - Brande).

Thirty one (31) videos were selected for viewing during lectures throughout a period of approximately six weeks. Video titles and their URLs are given in Table 1.

Preparation of video for classroom viewing: Links to most videos selected for classroom viewing were processed through www.ezsnips.com, an online software service (Brande and Arslan, 2012) designed to play a YouTube video via a specially constructed, web browser-compatible URL (sniplink). The sniplink was attached as a hyperlink to an object on a PowerPoint slide in the lecture presentation.

Playing the video in the classroom: At the appropriate time during the lecture, the instructor simply clicks the slide object to which the sniplink was attached, and within a few seconds, the YouTube video begins playing in a new browser tab, automatically opened by the PowerPoint hyperlink call. Some YouTube channels display banner advertisements at one or more points during the playing of a video. Due to terms of conditions imposed by YouTube, these banner ads are not to be suppressed. The instructor at the podium is able to quickly (<3 seconds) click off a banner ad that appears. At this time, we are unable to determine if there is any significant impact of these interruptions by partial ad display on student engagement.

Metric of student engagement based upon memory recall: A simple metric of student engagement was defined as the number of videos recalled from the previous period of lecture and study, approximately every 2 to 2.5 weeks. Three closed book examinations were administered. At the end of each examination, a section was provided in which students could enter a list of videos they recalled watching during the previous lectures. This section was incentivized with a bonus point for the list of recalled videos previously viewed. The requested list did not require exact titles or precise details, only enough description to enable the instructor to identify the specific video being referenced. Each video in the list written by each student was matched by the instructor to the video title played during a previous lecture. The number of times each video was identified was recorded (Figure 1, Table 2).

Correlation of student engagement with examination score average:

An hypothesis was advanced about whether or not student recall frequency (Figure 2) is related to the average of a student’s examination score (Figure 3). Because of visual deviations of the data (Figure 3) from an expected bivariate normal distribution, both non-parametric (Spearman’s coefficient of rank correlation), and parametric (product-moment correlation coefficient) tests were performed. We note that studies have found the parametric statistical test of the product-moment
correlation coefficient robust under a wide range of non-normality (Fowler, 1987).

**Image Processing**

Two video snips were selected for exploring the nature of mathematical functions of the digital video data (Figure 4; “Dust Devil in Desert”, hereinafter “Dust Devil” and “China Wall for sport”, hereinafter “China Wall”). The two snips record similar environments (desert southwestern United States), and similar inclusion of people in sport-like circumstances and activity (bicycling, all-terrain vehicle riding). Furthermore, we selected these two video snips in anticipation of determining the difference in degree, if any, to which these videos were recalled by students (results and discussion, below), and the correlation, if any, with objective, mathematical features.

As noted, our second, and longer term, goal is to discover correlations, if any, between cognitive signatures of student engagement (e.g., degree of recall from memory) while viewing instructional video and objective (computational) properties extracted from the data file of the video stream. It is clear that the number of possible features for extraction from a digital video file is infinite, both in the spatial and temporal domains.

Therefore, we must make choices that limit our investigation to particular metrics that we believe to represent features of cognitive significance in psychology. In this exploratory study, we have chosen a small set of spatial and temporal functions simply to test whether or not these functions vary to some degree among video data streams from what we subjectively judge to be somewhat similar videos. Among spatial functions, we computed average frame brightness, deviation of intensity, contrast, and other texture features (Haralick et al., 1973). In the temporal domain, we evaluated the amount of motion based on optical flow (Horn and Schunck, 1981) and the percent of non-stationary pixels in the image. Some of those features are also used in video quality evaluation and compression (Sonka et al., 2014). Spatial and temporal functions used for feature extraction are listed and described below. All video frames were first converted from RGB to gray levels before functions were computed, and results displayed on normalized scales.

**Selected functions:**

Median intensity is based upon a conversion of the pixel-level Red-Green-Blue (RGB) data to gray level via the following function - 0.2989 * R + 0.5870 * G + 0.1140 * B (Shapiro and Stockman, 2001).

Interquartile intensity is the range from the 75th percentile to the 25th percentile of the gray level intensity distribution.

Edginess is our term for describing the frequency of image edges in each frame. The value of edginess is determined from dividing the total number of edge pixels by the frame size (number of pixels in a single frame of the video). Edge pixels were determined from the algorithm in Canny (1986).

**Entropy** is a measure of information, calculated here from the frequency distribution of the pixel matrix \( p \) (Gonzalez, et al., 2009) by \( E = - \sum p_i \log_2 p_i \).

Colorfulness was computed as the sum of the standard deviations of the frequency distributions from each of the Red, Green, and Blue channels.

The next four texture features are based on the computation of the gray co-occurrence matrix \( g(i,j) \) (Haralick et al., 1973).

- **Contrast** was computed as \( \sum j (i - j)^2 g(i,j) \).
- **Correlation** was computed as \( \sum j \frac{-(\mu_i)(\mu_j)g(i,j)}{\sigma_i \sigma_j} \), where \( \mu_i, \mu_j \) are row and column averages of \( g \), and \( \sigma_i, \sigma_j \) are standard deviations.
- **Energy** was computed as \( \sum j g(i,j)^2 \).
- **Homogeneity** was computed as \( \sum j \frac{g(i,j)}{1+|i-j|} \).

Non-stationary pixels – computed as \( \sum k \frac{1}{N} \sum i_k \) | \( I_k+1 \) - \( I_k \) \( | \), where \( N \) is a total number of pixels, \( I_k \) is the current frame, and \( I_{k+1} \) is the following frame.

**Total motion** – computed as the sum of the magnitudes of vectors that represent optical flow (Horn and Schunck, 1981).

**RESULTS**

**Subjective student recall data**

A total of 31 videos (Table 1) were played during lectures and viewed by approximately 50 students throughout this study. The instructor matched student responses to the video titles, and compiled the frequency distribution given in Figure 1. A total of 347 instances of individual video recall were recorded by students over all three examinations (Table 2). One video (“Animation of river sediment”) was not recalled by any student.

The association of the number of videos recalled by each student is plotted against the student’s average examination score over the three tests (Figure 3). The parametric product-moment correlation coefficient was found to be statistically highly significant computed, \( r = 0.49 \) \( (n=49, \ p<0.001, \ 1\text{-tailed}) \). The coefficient of determination, \( R^2 \), was found to be 0.24.

**Objective image analysis**

The YouTube video snips were downloaded using DVDVideoSoft’s YouTube Download Free Studio. The video feature extraction was carried out using an in-house developed code in MATLAB (Mathworks, Natick, MA).

Measures of spatial features (per frame) are shown in Figure 5. Measures of temporal features (per frame) are shown in Figure 6.

No statistical studies based on the selected image analysis functions were possible in this preliminary study, as only one video was selected from each group (highly remembered, least remembered) for their exploratory use.
Statistical studies will commence with the acquisition of image processing data from multiple videos in each group.

DISCUSSION

Today, there are more opportunities for students to be engaged in the academic classroom because of changes in “best practices”. Many of these more recent changes devote more time to interactive activity and less time to the traditional one-way lecture delivered by the instructor. The explosive growth and accessibility of video on the web, specifically on YouTube, has given instructors a tool for the fast and effective delivery of digital video to web-enabled classrooms. There is a general recognition that video is an important component of effective pedagogy, but certainly not the only one (Clark and Mayer, 2011).

We are especially interested in the degree of memorability of instructional video because of this medium’s ability to document phenomena and bring dynamic processes from nature and the laboratory into the lecture classroom with information that potentially enhances student learning. From our perspective, what makes some earth science videos more memorable than others? Are there any factors that would help to identify in advance the more “memorable” videos? If so, could such factors be used to predict student engagement, and more importantly, student performance? Are the students who recall and record more videos simply the better students, i.e., students who tend to score higher on the examinations?

Videos selected for the course (Table 1) obviously vary in an unlimited number of ways. We hypothesized at the beginning that students would not recall each video with equal frequency. This seemed intuitively reasonable considering the variability of a large number of factors within the videos. Among them would be psychological and cognitive factors such as salience (which would include absence or presence of people and their activities in the videos) and visual factors such as color, foreground/background separation, time of occurrence and gradients of flow (action). We did not, a priori, hazard a guess as to which videos we thought would be recalled with greater frequency.

Our data may be used to investigate in more detail student responses. The frequency with which students recalled videos they had seen in lecture during the preceding period is given in Table 2 and Figure 1.

As expected, we find that not all videos are equally memorable to all students. Could it be that students whose examinations scores are higher on average than lower scoring students remember more videos? The variation of the number of videos recalled with the average examination scores for each student is given in Figure 3. It is obvious visually that the correlation is quite weak - in fact, \( R^2 = 0.24 \). In this case, about 75% of the variance is not explained by the correlation. We conclude that the ability of the student to achieve a higher score on average is not a dominant factor in explaining the number of videos recalled.

We have not collected, nor are we able to analyze, the innumerable other human factors that may control the number of videos a student recalls under these experimental conditions. However, with a sample size of about 50, it is also reasonable to assume that many of the factors (especially those that are subjective) are random with respect to their interactions, and therefore combine to cancel directional effects.

Thus, we have begun to explore the potential for objective measures to differentiate highly memorable videos from others. As previously noted, there are an infinite number of ways in which digital video streams from highly memorable videos might differ from those less memorable. To begin the exploration of objective methods and metrics, we have chosen a set of 11 mathematical functions and applied them to two videos from among our sample that have been judged to be similar in a number of ways (but are obviously different in many other ways). “Dust Devil” and “China Wall” (Tables 1, 2, Figure 4) are both set in a desert landscape of the southwestern United States, and both show people in active motion. These videos were chosen before any student response data were collected, and it was surprising to us that these two videos differed substantially in their recall frequency. “Dust Devil” was the fourth most memorable video, recalled 23 times by students, while “China Wall” was recalled by less than half this frequency, 9 times.

Different types of information, spatial and temporal, may be computed from the matrix of digital data from each frame of the video. Spatial measures are computed from single frames, while temporal measures are computed from changes that occur between adjacent frames in the stream. Results of a spatial analysis of the first 200 frames of each video are shown in Figure 5, and for temporal analysis, in Figure 6. It is obvious that these two videos, recorded by different people, at different times, in different places, of different subjects, should produce spatial and temporal metrics that differ to some degree. For example, note that for “Dust Devil” (Figure 4, top; Figure 5, left), contrast averages about 0.1 (all measurements expressed on a normalized scale), and for “China Wall” (Figure 4, bottom; Figure 5, right), contrast varies between about 0.3. Other variables are not so different. For example, median intensity is similar, around 0.5 to 0.6 for each video. Entropy is about 0.4 for each video. Colorfulness is about 0.6 to 0.7.

Comparisons similar to that discussed for spatial variables may be made for temporal variables. Figure 6 shows patterns of temporal change in the number of non-stationary pixels, and the total motion for each pixel location, derived from an optical vector flow analysis. “Dust Devil” (Figure 6, left) exhibits a segment of greater
total motion from one frame to another between approximately frame 60 and frame 120. The pattern of total motion for “China Wall” (Figure 6, right) does not exhibit a similar segment of motion within the equivalent video segment.

The visual analysis just described is not to be interpreted as objectively statistically significant in any way, especially as students viewed during lecture a longer video segment. Furthermore, we are not proposing that any of these objective measures of the digital video stream are either psychologically significant, or otherwise a cause of the difference in memorability exhibited by these two somewhat similar videos. Our purpose here is simply to explore the possibility of applying mathematical functions to the analysis of videos that we already know to be of different memorability to students under the specified conditions.

However, this preliminary analysis does point the way to further studies that could lead to valid statistical hypothesis testing. For example, with a larger selection of videos, we will be able to accumulate multiple videos of high and low recall frequency. From such a collection we could build a sample of sufficient size for statistical hypothesis testing. For example, we may find that, as a group, highly memorable videos differ from low memorable videos in the degree of motion, or of higher contrast, or in any other of these, or other variables we explore.

If such objective and significant differences are found, it would be of considerable interest to test whether or not videos that prove to be highly memorable can be predicted prior to classroom use. And if such an analytical process could be demonstrated, then a more efficient and effective use of digital video in classroom instruction may be possible. This would be a desirable and positive step in the development of tools and techniques for enhancing student learning with video multimedia.

CONCLUSIONS

Student engagement may take many different forms, and increases in student engagement are an important component of learning and performance. The incorporation of digital video snips in a lecture centered course has been designed to test students for variation in recall frequency, a proxy for student engagement. Students were given an incentivized bonus question if they could recall videos within a period of approximately 2.5 weeks after viewing. Of a sample of 31 videos viewed by students, one video was not recalled by any student. Of the 30 videos that students recalled at least once, the number of times individual videos were recalled ranged from one to 27. Although there is a statistically significant correlation between the number of videos a student recalled and the average overall examination score, the percentage of variance unexplained in the correlation is about 75%, thus pointing to, potentially, many other variables that may account for differences, potentially, many other variables that may account for differences in video recall frequencies among students.

An analysis of spatial and temporal mathematical properties of two videos selected for higher recalled frequency and lower recalled frequency show both differences and similarities in various measures, including contrast, homogeneity, edginess, total motion, and other properties. Future statistically based studies may identify image and video properties that associate with differential recall frequency.

These preliminary results provide guidance for the design of further experiments to discover and determine potential correlations between subjective assessments by students (of the degree to which videos are memorable, and thus remembered to a greater degree), and objective and as yet undiscovered quantitative properties of the digital video streams. If such correspondences could be identified and quantified, we believe this process would provide a better way to select video for instructional content, provided the video also satisfies academic requirements for the instructor and course. The selection of more memorable digital video could then be tested for enhancing student engagement and performance.

REFERENCES


### Figures and Tables

<table>
<thead>
<tr>
<th>Video Title</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mud settling in tube</td>
<td>97</td>
</tr>
<tr>
<td>Alaskan Marine Quay</td>
<td>82</td>
</tr>
<tr>
<td>Natural Nest for Dry, Less Sand</td>
<td>77</td>
</tr>
<tr>
<td>Dust Devil in Desert</td>
<td>70</td>
</tr>
<tr>
<td>Creme Marble Quarry-Spain</td>
<td>67</td>
</tr>
<tr>
<td>Sandboarding on Namibian Dunes</td>
<td>64</td>
</tr>
<tr>
<td>Global Ocean Rides System</td>
<td>60</td>
</tr>
<tr>
<td>The Big Rock - VHRI Campus Simulation</td>
<td>55</td>
</tr>
<tr>
<td>Dark Subpolar In A Scene, North West Africa</td>
<td>50</td>
</tr>
<tr>
<td>The Hadley Cell</td>
<td>45</td>
</tr>
<tr>
<td>Living coral reef</td>
<td>40</td>
</tr>
<tr>
<td>Grand Canyon adventure in a sandbox</td>
<td>35</td>
</tr>
<tr>
<td>More sandpitas in wave tank</td>
<td>32</td>
</tr>
<tr>
<td>Sandpitas in wave tank</td>
<td>31</td>
</tr>
<tr>
<td>Regional and Contact Metamorphism in Subduction Zone</td>
<td>29</td>
</tr>
<tr>
<td>PBS News Bureau Dust Road trailer</td>
<td>28</td>
</tr>
<tr>
<td>Managing the Mighty Mississippi</td>
<td>26</td>
</tr>
<tr>
<td>Experimental Handwind Erosion</td>
<td>26</td>
</tr>
<tr>
<td>Laboratory mud soils formation</td>
<td>26</td>
</tr>
<tr>
<td>Chinese Wallfaber, China, Calif.</td>
<td>26</td>
</tr>
<tr>
<td>Geology at Stocar Pt, Scotland</td>
<td>26</td>
</tr>
<tr>
<td>Water erosion and sand</td>
<td>25</td>
</tr>
<tr>
<td>Stream Loads</td>
<td>25</td>
</tr>
<tr>
<td>BHI Relay - Where did the Moon Come From?</td>
<td>24</td>
</tr>
<tr>
<td>Hydrothermal Vent &quot;black smoker&quot;</td>
<td>24</td>
</tr>
<tr>
<td>Movement of sand in wind tunnel</td>
<td>24</td>
</tr>
<tr>
<td>Sand blowing across dunes</td>
<td>23</td>
</tr>
<tr>
<td>Asian Dust Transport Computer Model</td>
<td>22</td>
</tr>
<tr>
<td>Braided Stream</td>
<td>21</td>
</tr>
<tr>
<td>Sources of Water on Earth</td>
<td>20</td>
</tr>
<tr>
<td>Experimental Hotwind Erosion</td>
<td>20</td>
</tr>
<tr>
<td>Stream Loads</td>
<td>20</td>
</tr>
</tbody>
</table>

**Table 1.** Video titles included in this study, and their web links

**Table 2.** Counts of videos recalled by students on examinations

---

186
Figure 1. Frequency of recall for a selection of short videos shown during lecture periods.

Figure 2. Number of videos recalled by each student.

Figure 3. Variation of the number of videos recalled with test score.

Figure 4. Video sequences from two snippets shown to students. Top – “Dust Devil in Desert”, Bottom – “China Wall for sport, Glamis, Calif.”

Figure 5. Values of spatial variables (on normalized scale, per frame) for two video snippets. Left – “Dust Devil in Desert”, Right – “China Wall for sport, Glamis, Calif.”

Figure 6. Values of temporal variables (on normalized scale, per frame) for two video snippets. Left – “Dust Devil in Desert”, Right – “China Wall for sport, Glamis, Calif.”
APPLYING JAPANESE EXPERIENCE TO RURAL COMMUNITY DEGRADATION CAUSED BY RURAL-URBAN MIGRATION IN SARAWAK

ICHIKAWA Masahiro, Dr.
Faculty of Agriculture, Kochi University
200 Monobe Outu, Nangoku, Kochi, 783-8502, Japan

ABSTRACT
Rural-urban migration is progressing in Sarawak. In Japan, such migration started in the 1960s and depopulation in rural areas became a social problem. Today depopulation and aging in rural areas have become more serious and some villages are disappearing or have already disappeared. As reference for rural planning and development in Sarawak, I will present a case of a rural area of Japan—Otoyo, Kochi prefecture—describing present conditions, the process of depopulation and aging, and some countermeasures.

INTRODUCTION

In communities in the mid and upper reaches of the Baram river basin, the number of empty houses (bilek kosong) has recently increased. This occurs when all members of the house have migrated to urban areas. Today communities with 20-30% of empty houses are frequently observed, and there are some with more than 50% empty (figure 2). In communities where depopulation and aging progress, phenomenon such as inactivation of collaborative work, decline of swidden farming, and disappearance of agricultural rituals and ceremonies of funerals and weddings are observed.

Depopulation and aging in Otoyo, Kochi prefecture, Japan
After World War II, the population in Japan rapidly increased until around 2010, and then gradually started to decrease. Otoyo is a municipality of Japan suffering severely from depopulation and aging. The population has been decreasing constantly since the 1950s. The population in 2010 (less than 5 thousand) is one third of that in 1960. At the same time, the percentage of old people has increased.

The government, as measures against depopulation and aging, has improved rural infrastructure, such as roads, water supply, and communications in order to make rural life more comfortable. However, this sometimes had the opposite effect, as the rural population aggressively migrated from the area as a result of road and railway development. Local governments and communities are trying to add additional value to their local products and to seek better sales. The governments and local peoples have made efforts, but depopulation and aging, and as a result degradation of communities, have not stopped.

Micro-Electro-Mechanical Systems (MEMS)
Urban development has progressed all over the world, and especially in Southeast Asian countries where there has been progress in economic development. In Malaysia, although Sarawak is a rather poor state where rural areas are predominant, urban areas have enlarged rapidly following economic development, especially from the 1980s onwards. As a result, a large number of people have migrated from rural to urban areas.

In Japan, the population in rural areas started decreasing in 1960 and depopulation had already become a social problem in the mid-1960s. Today depopulation and aging in rural areas have become more serious and many villages are disappearing or have disappeared.

In this paper, firstly rural-urban migration and its effects on communities in Miri division, Sarawak are briefly described. Then, a case of depopulation and aging in Otoyo, Kochi prefecture and countermeasures in Japan are explained. Finally the paper discusses how to prevent the degradation of rural communities in Sarawak.

Rural-urban migration in Miri division, Sarawak
Miri started to develop after discovery of an oil field in 1910. From the 1980s especially, the urban area and surrounding residential areas rapidly enlarged. The rural and urban populations in Sarawak in 2000 were almost the same, but in Miri division, the urban population already had overtaken the rural population (figure 1).
To prevent rural degradation in Sarawak

In Japan, under policies to prevent rural degradation caused by rural-urban migration, a significant amount of national budget has been spent on infrastructure development for rural areas to improve living standards. The policy was considered to work well for the rural people who remained there. However, even when roads, electricity and water supplies were improved, young people with a longing for urban life continued to migrate away from the rural area. Since excellent brains did not remain in rural areas, attempts to improve income, such as by processing and selling agricultural products, do not go well in many cases. Although infrastructure development in rural areas is important, it is insufficient by itself to stop depopulation and aging in such areas. However, recently, a number of young people who dislike the stressful urban environment are migrating from urban to rural areas. To prevent rural degradation, it is necessary for society to recognize the “richness” of rural life.

FIGURES AND TABLES

![Figure 1. Change of rural and urban populations](image1)

![Figure 2. A longhouse with 70% empty houses in the mid reach of Baram river basin](image2)

![Figure 3 Inactivation of ritual ceremonies in rural communities](image3)

![Figure 4 Inactivation of collaborative work](image4)
DIGITAL MAMMOGRAM IMAGE ENHANCEMENT APPROACH BY CONTRAST STRECHING AND HISTOGRAM EQUALIZATION: AN IMPLEMENTATION AND RESULTS

Valliappan Raman, Putra Sumari
Multimedia Research Group Lab, School of Computer Science
University Sains Malaysia
Penang, 11800, Malaysia

ABSTRACT

The aim of the paper is to present the research methodology to solve the research questions identified in this mammogram image enhancement. First, image enhancement techniques on mammograms images are explained in detail for further stages of segmentation, enhanced feature extraction and proposed classification. This paper includes the image acquisition procedures, description of MIAS database and image enhancement technique by contrast stretching enhancement combined with histogram equalization and the algorithm description with implemented results. The removal of pectoral muscle from the breast image will be initial stage for further stages on machine learning and proposed problem solving approach for the paper.

NOMENCLATURE

\( HE \) = Histogram Equalization
\( IE \) = Image Enhancement
\( CS \) = Contrast Stretching
\( Br \) = Breast region
\( CS \) = Contrast Streching
\( ROI \) = Region of Interest

INTRODUCTION

The major global health problem for a woman is breast cancer, and it is considered as most common cancer among Western and Asian countries [Fajarado et al., 1996]. American Cancer Society predicated the number of cancer affected woman living in United States is high, statistics shows that among nine women, one of them will have breast cancer in their life time [Fajarado et al., 1996]. Similarly, Breast Cancer Society Malaysia predicts that one in nineteen have the possibility of breast cancer. Nearly 30% of the woman having breast cancer is between the ages of 40 years to 50 years

Women are highly encouraged to perform breast examination every year to check for any unusual changes. If tumour is suspected, then it is required to proceed with thorough breast examination. First stage is to proceed with image diagnostic of the breast to get more information’s, this diagnostic imaging for breast cancer is known as mammography. Mammography in particular, physician evaluates the specific area of the patient body which is highly visible for diagnosis. New England Journal of Medicine states that medical imaging is the one of the important development in medical field, where it provides the doctors with organ functions and cells inside the human bodies. Mammograms captures the X-ray with lower energy which passes through a compressed breast, from the viewpoint of X-rays, it is classified in two categories i.e. Medio Lateral Oblique View and Cranio-Caudual Lateral View. Fig 4 shows the Medio-Lateral Oblique (MLO) and Cranio-Caudual Lateral view (CC). In MLO view, whole breast is visible which includes lymph nodes, in this, one region follows to the pectoral muscle which present at the left and right corner of the image, in that image annotations and R symbol is marked to indicate the right breast. In CC view, is taken from the upper view, but sometimes it is unable to show the surrounding regions closer to the location of the chess wall. Figure 1 shows the labelled mammogram image.

The objective of the paper is to show the implementation results of mammogram image enhancement by contrast stretching and histogram equalization approach. Image enhancement (IE) is the one important stages to reduce the noise in the image by terms of pre-processing method. Section 1 explains the proposed methodology for the mammogram mass segmentation and classification, in section 2 mammogram acquisition is explained, MIAS database used for training and testing is described in section 3. In section 4, mammogram image enhancement is explained in detail with implementation results. Finally conclusion is made in section 5.

Proposed Research Methodology

Many research methods are developed in computer assisted method to identify and locate the various type of
masses and make a classification. In this thesis, we have developed an enhanced automated approach which aims to extract the abnormality of masses in the breast Region of Interest and apply new classification method to improve the accuracy of mass detection compared to existing research works. To achieve the objective of the project, overall general module of machine learning is shown in Fig 2. Research approach has four distinct steps: pre-processing, segmentation and enhanced Haralick texture feature extraction and proposed novel best case classifier are shown in Fig 3.

**Mammogram Image Acquisition**

In Medical imaging, greatest challenge is to access the real medical images for implementation and testing, it is due to privacy issues and technical problems. In mammogram image acquisition, X-ray film of the mammogram is digitized by laser scanner. The scanners will digitize the film of mammogram by measuring the optical density in the window sized region of the film, after that it will be converted into gray level intensity. The window size is computed by spatial resolution of the image. Image resolution is measured in micron per pixel where it denotes the region size in the film consists of pixel represented in the digital image. The location of pixel in the image is represented by intensity, value of the pixel is measured by ranges of optical density, and bits are used to store the grey level pixels. The accuracy of digitized images are depends on the spatial resolution and gray level occurrence in the image. In our research, original image are taken from MIAS database, for X-ray mammograms obtained from database are digitized with an image resolution of 100 × 100 μm² and 12 bits per pixel by a laser film digitizer with higher resolution. In the image small masses are usually larger than 3mm in diameter, the digitized mammograms are decimated with a resolution of 400 × 400 mm² by averaging 4 × 4 pixels into one pixel in order to save the computation time. The mammogram image in single view after digitization from MIAS database is shown in Fig 4.

**MIAS Database**

The Mammogram Image Analysis Society is a part of UK research group which is doing more research on mammogram images and the team of researchers are well capable in understanding the complex images of mammogram. In ref [4], website has the free download of sample images made it online and available for computer based clinical study. For our thesis, all the sample data’s for implementation and testing taken from the MIAS database. The X-ray film available in this database is carefully selection from National Breast Screening Programme, UK. Once the images are available, Joyce-Label scanner is used to digitize the image with resolution of 100 × 100 μm², with each pixel consists of 8 bits. The available database have 161 patients with left and right breast images with total of 322 images, which are categorized into normal, benign and malignant. Total images in the database comprised of 63 benign, 208 are normal and 51 are abnormal images (malignant). This database will used to train and test the proposed classifier for detecting the mass.

**MAMMOGRAM IMAGE ENHANCEMENT**

The very basic enhancement approach applied in the mammogram image is to improve the contrast. In the mammogram images, malignant tissue and normal tissue have the variation in contrast. As we have seen in our earlier approaches, there are many research works are done based on enhancement techniques. The reason for performing enhancement is to provide radiologist a quality image for diagnosis. There are only few enhancement methods are applicable for mammogram images; all the techniques are not suitable for these images. By traditional methods, it’s very hard to achieve good result of enhancement. From the studies, there are challenges faced by researchers in local contrast enhancement without loss information’s from the image. In the preparation stage, mammograms images need to improved by enhancing the contrast and reduce the noise in the image. Our main objective is to improve the image quality and remove the unwanted objects in the background of the image by further enhancement processing and make it ready for segmentation. The approach used in this research is of three stages; first stage is to apply Gaussian smoothening function to suppress the noise in the image. Second stage is to use histogram equalization with few modifications from the traditional methods to enhance the contrast in the image. Third stage is to apply modified contrast enhancement, which can identify the hidden information’s from the image. Three stages of enhancement approach used in our research is identify the local information about the image. These proposed approaches can provide more accurate results and will obtain full detailed information about the image for segmentation approach in the mammogram image.

**Noise Suppression Method**

The In MIAS database, mammogram images have noises in different types. Noise with higher intensity will be having more values for the optical densities which have labels and artefacts. Shadows and markings are part of the noise, which can reduce by replacing the black pixels. Suppression of noise is performed by applying Gaussian smoothing operator. Gaussian operator is 2-D convolution operator which can be used for blur images and more noisy images. Kernel is represented in the form of Gaussian shape or in form of bell-shape. Gaussian is represented in the form of:
Gaussian smoothing is considered as point spread function. Convolution is applied to reach the function in a 2D distribution. Imaging will be having collection of random pixels, need to make an approximation on the function before the convolution. It is a non zero in all part of the image, it requires a larger kernel of convolution, while in implementation it has higher values in standard deviation from mean, kernel can taken out , when we are in need of suitable kernel, Gaussian smoothing can performed. Kernel used for our research is 7x7 kernel as part of convolution filter. The images of mammogram have been classified in to three types, fatty, dense glandular and fatty glandular which have denser tissues in the breast part. Each category will have different intensity pixels and variations are observed in every type. Based on the types, standard deviation values are computed and made as choice to adjust the level of smoothening.

**Histogram Equalization**

For Histogram Equalization (HE) approach is technique applied in the images for improving the contrast, idea is to map the gray level pixels based on probability distribution, it orients the image, flattens and stretch the histogram of the image for improving the contrast.

The image of the histogram is represented with gray level value of \( K = 0,1,2, \ldots, L-1 \); the discrete function is computed by number of pixel identified with gray level values \( n_g \), it is denoted by \( D(G_k) = n_g \). Histogram of the image is normalized by each value computed is divided by total number of pixels presented in the image; therefore histogram normalization is obtained by probability density function:

\[
P(G_k) = \frac{N_g}{N}
\]

Therefore computation results provide the probability of gray level occurrence in the image. Transformation function is applied and it is represented by \( T(G_k) \), which produces the cumulative cost of adding all the pixels from the known histograms and its maps the input to the output.

\[
Gaussian(p,q) = \frac{1}{2\pi} e^{-\frac{p^2+q^2}{2\sigma^2}}
\]

Transformation function output value is a single value , and intervals are increased continuously, where value of interval ranges between 0 to 1, therefore it provides inverse transformation from the scale of output to be from black to white pixels in the image. if the function does not increase the interval values, it will invert the intensity values and produce gray levels in the image. It can guarantee by mapping the output gray level value pixels to input pixels. Inverse transformation is represented by:

\[
\text{Trans}=\text{inverstrans}\left[\frac{(\text{Trans}_{\text{min}}-\text{Trans})}{(L-1)}+0.5\right]
\]

The values are normalized and expanded to 0 to L-1, whereas \( \text{Trans}_{\text{min}} \) is the smallest value from the cumulative function vector, Trans is the transformation value of gray scale ranges from 0 to 255 in each gray level. This approach is simple and powerful in the range of change in brightness of the image. The output achieved my mammogram image improves but missing of local information occurred in this approach. In this traditional approach, there are few changes performed and applied to obtain good result. Basically when histogram distribution is uniform by mapping the cumulative distribution, during this stage, there is a simple medication performed in output of the histogram, and make it closer to the uniformly distributed histogram. The mammogram image enhancement by Histogram equalization and modified contrast enhancement method is shown in fig 5.

The The histogram equalization method does support adjustments for enhancement, but can control level of enhancement by adjusting the parameters. In this method, modified histogram \( H_1 \) and input histogram \( H_2 \) and difference computed is \( H_1 - H_2 \) and it is modified and closer to Uniform histogram, therefore it increases the contrast enhancement. Formulation for modified histogram output result is determined by:

\[
\min \| H_1 - H_2 \| + \theta \| H_1 - u \|
\]

Whereas \( H_1 \) is the modified histogram; \( H_2 \) is the input histogram; U is the Uniform histogram; \( \theta \) is the parameter value for adjustments. It varies from 0 and \( \alpha \). If the values are 0 means standard HE and \( \alpha \) means it is a original image. Modified histogram is average.
computation weight between input histogram and uniform histogram, finally based on changing the parameter values, contrast can be adjusted in the image. The resultant image will be processed for modified contrast enhancement stage. Below fig 6 shows the result of modified histogram equalization.

**Contrast Enhancement Method**

In histogram equalization, local information’s are missing, to maintain the local information details in mammogram images; modified contrast enhancement technique is performed. During this computation of global transformation, number of pixels may remove or not identified, then transformation function is applied on gray level distribution in the image. After modified histogram equalization is applied, and then contrast enhancement is performed. The contrast method has obtained statistical values from the output of histogram.

Discrete gray levels are represented by \([0, L-1]\) and it is denoted by \(d\). Normalized histogram are computed and it is denoted by \(P(d_i)\). The moment of \(D\) and its mean is calculated for measuring the average gray level occurrence in the image, average contrast is computed by standard deviation or variation.

\[
m = \sum_{i=0}^{L-1} d_i P(d_i)
\]

(6)

Moment is calculated based on first order and second order where as \(n=1, 2\). The mean and variance are used as the parameters for contrast enhancement.

The modified contrast enhancement is computed after achieving the histogram modification.

\[
H_1(x, y) = \begin{cases} 
E.f(x, y), m < K_0, K, D, K_2D \\
F(x, y) & \text{otherwise}
\end{cases}
\]

(7)

\(H_1(x, y)\) and \(F(x, y)\) are the final outputs of enhancement and modified histogram images, where the values of \(E, K_0, K_1\) are the parameters to be adjusted \(D\) is the standard deviation, if the values of \(K_0 < K_1 < ... K_n\), if the values are greater than 1 , it will enhance light areas of the pixels and lesser means for darker area of pixels. The parameters are very powerful in enhancing the images, it provides promising results on images, level of contrast enhancement is adjusted based on input of the original image contrast. When we have the contrast to be lower, images have narrow histogram and then we apply the histogram equalization to reduce the contrast and make it uniformed by further enhanced with modified contrast enhancement. In histogram modification, value of is 0.9 for the image in MIAS mammogram database. The values of \(E = 2\), \(K_0 = 2K_1 = 0.005\) are the taken for enhancement. The E value is very important for balancing the contrast in the image, and computational complexity is also balanced in the image. Therefore these methods are proposed to enhance the quality image for next stage of detecting the edges to remove the pectoral muscle and detect the boundary in the image to find out the region of interest in the image for segmentation. In fig 7, modified contrast enhancement is shown and in fig 8 shows the image enhancement by histogram equalization combined with contrast enhancement.

**Locating the Breast Boundary**

After Enhancement of images, region of breast need to detected and extracted from the background of the image. To process the images for further stages of segmentation, breast boundary need to be extracted and detected. The major objective of performing this task is to find out the outer edge lines from the breast edges in the image. First starting point for the enhanced image is scanned from the right side to locate and position the first row in the image. While processing the scan, look for the surrounding pixels and compute the pixel with high priority. Plotting list area is location of the traversed pixels can be stored and later used for drawing the breast boundary. The low priority pixels can be stored in backtrack stack, if the traverse process are in dead stage. If dead stage is reached, it will taken out from backtrack stack with less priority of pixel values and continue to proceed the traversal process , till it reach the image bottom, highlighting the breast region, therefore plotting list will keep the storage of all breast boundary pixels for smoothening and processing. Fig 9 shows the results of border of edges from the breast regions

**Removal of Pectoral Muscles**

The mammogram edges are detected by boundary based detection methods, therefore we obtain a image with breast region which have the difference from background. To extract exactly the breast region, pectoral muscle region need to be removed. When we have read the mammogram images in MIAS database, all the images which has pectoral muscles have a unique intensity values compare to dense and fatty tissues of the image. The
pectoral muscles are located in the top left corner of the region, and it moves from the downwards of top margin and then moves to the left of the mammogram image to form triangle which is inversion shape. The main aim of this proposed method is to detect the breast region which is in the form of triangle shape, and locate the edges of pectoral muscles for removal in the form of border. First right pixels are identified; vertical line is drawn from top to bottom margin of the pixel regions. Next parallel vertical line is drawn from top to bottom which reaches the end of left baseline. Therefore there is a formation of rectangle, within that pectoral muscle is confined and it is removed from the connected regions of rectangle. The results of pectoral removal from images of MIAS database is shown in Fig 10.

CONCLUSION

Finally in this paper, when we compare the past research works on enhancement, our combined histogram equalization with modified contrast enhancement have produced good results. The contrast enhancements techniques will strong enhancement functions and store all the local detailed information’s which cannot be missed and reduce the quality of original image. There are several results achieved from MIAS original images which have fatty, glandular and dense tissues are taken, the applied method have given very good result compare to the previous enhancement techniques. The measurement of images after its enhancement process is very difficult and there is no universal measure for enhancement techniques. After enhancing the images, there are two algorithms applied for edge detection, breast boundary detection is performed on first stage to identify the edges from top to bottom of the image, next is to detect and remove pectoral muscle edges from the breast region. Images of various sizes shape and margins are being applied with pectoral muscle removal algorithm. The result produced by both algorithms is promising in the pre –segmentation stage. The accuracy achieved by this approaches are evaluated by quantitative measures such as golden standard. The radiologist manually draws rectangle on the region of pectoral muscles area, and it is been processed for automatic segmentation, where the comparison is made by pathologist on verification of right pectoral muscle suppression, the ground truth is measured. There is none of the algorithm could be 100% robust, especially for mammogram images which have heterogeneous property, while digitizing, there are problems such as artefacts, more background noise, dust artefacts, scratches are occurred in the image which highly affects the reliability of the applied algorithm. In MIAS database Out of 322 mammograms, 140 were tested, and it achieved a 92% accuracy of enhancement its and its processed for next stage of mass segmentation.

REFERENCES


http://peipa.essex.ac.uk/info/mias.html


FIGURES AND TABLES

Fig. 1 illustrates labeled mammogram image.

Fig. 2 illustrates general principle module of machine learning approach.

Fig. 3 illustrates the overall methodology of proposed mass segmentation and classification.
Fig. 4 illustrates the mammogram image in single view after digitization from MIAS database.

Fig. 5 illustrates the mammogram image enhancement procedure.

Fig. 6 illustrates the original mammogram image and histogram output.

Fig. 7 illustrates the contrast enhancement after modification of parameters.

Fig. 8 illustrates the results of histogram equalization and contrast enhancement after modification.
Fig. 9 illustrates the results of border of edges from the breast regions

Fig. 10 illustrates the results of pectoral removal from images of MIAS database
EXPLORING THE POSSIBILITY OF COMPANION ROBOTS FOR INJURY PREVENTION FOR PEOPLE WITH DISABILITIES

Mark Tee Kit Tsun, Lau Bee Theng
Faculty of Engineering, Computing & Science
Swinburne University of Technology Sarawak

ABSTRACT

Robot technology has been applied beyond the scope of industrial manufacturing and service-automation, and into our homes. Today, commercially available robots are engaged in direct interaction with human users, for the purpose of entertainment, assistance and companionship. However flexible the robots are, they are still bound to the limitations of their design, which depend on the target demographics. Companion robots, in particular, have to be specifically developed for use by the elderly, children with cognitive impairments, post-injury disabled patients and so forth. This study explores currently available examples of robot companions in the service of elderly people, children with disabilities and disabled patients. These three demographic groups share a common need for monitoring and support in order to prevent accidents from occurring. Some existing injury prevention methods are explored but each has its limitations. The goal of this study is to introduce the possibility of utilizing companion robots as a means of reducing those limitations and augmenting injury prevention systems. To enable studying this possibility in application to children with cognitive impairments, a concept of a companion robot for injury prevention is presented.

INTRODUCTION

In 1941, Isaac Asimov’s Three Laws of Robotics planted the seed for a future where robots and humans coexist over dimensions that go beyond the simple servitude that we expect from conventional machines and tools (Asimov, 2004). Fictional stories and movies then took their turn in supplanting the idea of robots living and working in our homes and offices, instead of isolated controlled factory environments. While Sci-Fi did introduce humanity to the benefits and wonders of robots, it also embedded a deep seated fear of machines that may one day malfunction or rebel, causing undue harm and damage.

Today, Human-Robot Interaction (HRI) serve as a branch of science that study the social-psychological impact that robots have on human users, and develop technologies in communication, control and design that enable positive experiences in life-long human-robot relations. This focuses on human-centric design in robots help introduce companion robots into homes. There were 137000 recorded home-based entertainment and “leisure” robots in 2005, with estimated growth to 2.5 million by 2007 (Dautenhahn et al., 2005).

ROBOT COMPANIONS

Today, robot companions are built to accommodate a wide variety of forms, from home security, automation and maintenance to intelligent toys and educational aids. For instance, the Roomba series of robots were developed for performing house-cleaning actions such as vacuuming, mopping and pool cleaning (iRobot Corporation, 2013). The ‘Furby’ line of toys respond to users using animatronic gestures and sound effects to mimic an emotional creature (Hasbro, 2014). For multinational businesses, robots have been making a gradual foothold in extending the flexibility of communications beyond fixed teleconferencing boundaries. Telepresence robots such as “the Double” has enabled users to interact and communicate via a remotely controlled avatar (Double Robotics, 2014).

While robots that were designed for direct human contact has been steadily growing pervasive, the journey towards widespread acceptance has not been without difficulty. Dautenhahn, Woods, Kaouri, Walters, & Werry (2005) conducted a study on people’s attitude towards robots to determine what aspects of robot design is most well received or strongly rejected. One strong theme that repeatedly surfaced was that a majority of the test subjects prefer communicating with the robot using natural language, but adversely reacted to human-like appearances and behavior. They are more likely to react positively towards anthropomorphic robots than humanoid designs. Yet another strong theme involved matching visual appearance appropriately to expected functionality. This was evident in an experiment to gauge the effectiveness of robots in facilitating group interactions between elderly people using ‘Paro’, an anthropomorphic robot that resembles a baby seal. When observing the robot flapping its fins, the subjects exclaimed that it should perform better when put into water. After having explained that the robot was not waterproof, the observers reported notable disappointment in the test group (Kidd et al., 2006). Dautenhahn’s study also shows that a majority of the subjects prefer companion robots as assistants and
appliances rather than servants or mates. House guarding and maintenance work is preferred over babysitting children, as it seems to be socially established under ‘human domain’. In terms of age, it was found that the younger demographic are more accepting of robots while the elder population are generally more ‘frightened’ by them (Dautenhahn et al., 2005).

Companion robots are also subjected to additional acceptance factors as compared to intelligent toys, due to their life-long service. A robot is usually well accepted under the effect of novelty, functionality and visual appeal. However over time, novelty of technology wears out, grief builds up over limitations and anthropomorphism may give rise to false expectations. One possible solution is to create companion robots that learn and mold its function and behavior to the user’s preferences over time (Dautenhahn, 2004). This paves the way for the next evolution of companion robots which humans may bind to easier for life-long assistance.

The results of social studies and research conducted under HRI are reflected in the growing population of commercially available robots for assisting and accompanying us, especially at home. As companion robots become progressively affordable and pervasive, it would be inevitable that they will be a common household fixture in the near future. Some examples of current robot companions are discussed over the next few sections.

Companions for Elderly People

The core aim for care of elderly people, aside from physical maintenance, is sustainment of their emotional wellbeing. Factors such as resilience, being productive, sensory enrichment, and having social connections play key roles in creating a meaningful life for them. As such, existing robot companions are geared towards providing catalysts for social interaction, daily activity support and entertainment (Khosla et al., 2013).

Paro, an anthropomorphic robot that looks like baby seal, interacts with users via actuated fins and sound effects in response to touch and voice (Fig. 1 left). This robot was used for encouraging social interaction amongst elderly patients. According to that study, Paro was known as a ‘mental commit robot’ which is aimed at eliciting interactions by forging emotional attachment to its users (Kidd et al., 2006). Due to its outward appearance and limitations of touch and sound detection, it suffered from disappointed expectations that it should operate in water.

Matilda is an effective communication robot which can utilize voice vocalization, facial recognition and tracking as well as touch sensors and a suite of preset activity maneuvers (Fig. 1 right). In one particular study, it is used in nursing homes as substitute conversation and activity companions for the elderly residents. These robots are used as a solution for countering the limited staff at these nursing facilities. Matilda was also used for facilitating group activities and games, suggesting diets and encouraging daily activity via conversations. The results of the study shows that an average of 77.5% of the subjects respond positively to conversations with Matilda, but less than 50% are comfortable with touching the robot (Khosla et al., 2013).

In supporting the need for self-sustainence of resilience and productivity, the ASTROMOBILE platform provides an automated monitoring and companionship solution to senior citizens who are independently living (Fig. 2). The robot works in combination with a sensor network, wearable electronics and wireless connectivity. It functions similar to Matilda, but is primarily tasked to monitor the position and condition of its human master. In the event of emergency or request, it can communicate directly with healthcare staff via network (Cavallo et al., 2013).

Companion robots may also specifically function as personal assistants to elderly people who may suffer from mild cognitive impairments. The Scitos G3 (Fig. 3) is one example platform that helps the user make phone calls, provide reminders of events, remembers the location of important items, as well as function as a speech recognition device to automate the house (Schroeter et al., 2013).

Companions for Disabled Patients

People who suffer from post-injury disabilities face difficulty in the form of performing daily activities rather than those of social-psychological nature. Compared to elderly users, they require more utility-centered assistive robot companions. Some dimensions of assistance include locomotion, reaching and grasping, as well as access to hands-free communication and home control.

An ideal example of utilitarian robotics being preferred over ‘companion’ counterparts for disabled people is an actuated wheelchair system that is able to move in formation with other units or companions (Fig. 4). The robot system consists of rangefinders and an omnidirectional camera, coupled with computing power for localization and position tracking. While not equipped for robot-human conversations, it facilitates social communication with fellow human beings without locomotion being a barrier (Sato et al., 2014).

A concept of a companion robot for the blind has been proposed as a solution to the limited access to guide dogs. The prototype called Amigo (Fig. 5), is developed to feature human-robot communication via speech, GPS and Social Network powered trip planning and navigation, as well as environmental sensing. The proposed final design implements a robot body that can be detached from the handle and thus be remotely controlled. The handle may also act as a manual walking stick (Lee et al., 2013).

The flexible mechatronic assistive technology system (MATS) is a 5 degree-of-freedom (DOF) self-contained manipulator for multi-purpose assistive robot that can attach itself to installed docking stations (Fig. 6). It can
Companions for Children with Cognitive Impairments

Children with cognitive disabilities benefit from companion robots which focus on both emotional-psychological support as well as assistance of therapies and exercises. While similar to the forms and functionalities of robots for the elderly and the disabled, child-centric robots are developed according to therapeutic needs that range from encouraging active play to facilitating social interaction with caregivers or other children.

Kaspar is a minimally expressive humanoid robot used for facilitating one-on-one social interaction exercises for children with impaired social skills (Fig. 7). The upper extremity and facial features of the robot is actuated to react to the interactions of the child, both via facial recognition and direct control by the therapist. The use of mediatory robots such as Kaspar is inspired by an interesting observation which shows that autistic children respond more readily to artificial constructs (such as a robot) compared to other human companions. With regular use, robots such as Kaspar could help coax the child to getting used to conversing and interaction with others through facilitating imitation exercises, turn-taking games and discussion sessions (Wood et al., 2013).

While robots such as Kaspar require the operator to have active control at all times, there are others which function as self-contained companions that accompany and entertain children in place of caregivers. One such system is the NAO robot, developed by Aldebaran Robotics (Fig. 8). NAO is a fully articulated humanoid robot that has built-in routines for storytelling, game playing, imitation exercises, and a wide variety of programmable therapeutic scripts. The robot is also equipped with speech and face recognition capable hardware, which enables advanced applications to attract the child’s interest and instigate creative play (Shamsuddin et al., 2012). While NAO is an ideal platform of a robot companion, its main drawback is the high cost and required expertise to maintain its continued use.

As an alternative to high-functioning humanoid robots such as NAO and Honda’s ASIMO, systems such as the IROMEC seek to consolidate the key therapeutic functionalities into simpler, more robust form factors for sensible daily company. The IROMEC is a robotic toy that encourages development of communication, motor, cognitive and social skills via interactive play (Fig. 9).

The robot is designed to be modular, programmable and extendable according to the developmental needs of the child as determined by the therapists. The main goal of using the IROMEC is to be an exercise device that facilitates repetitive practice during the lapse between therapy sessions (Ferrari et al., 2009).

There are also companion robots that are specifically designed for operation in preset situations. For instance, a full speech robot was developed for engaging children in active conversation during travel in automobiles. This provides the travelling child with entertainment while reducing the chance of distraction to the parent who is driving the automobile (Hiah et al., 2013). Fig. 10 shows the prototype in-car companion robot and the testing environment in which usability experiments are carried out.

INJURY PREVENTION

Our quality of life regularly faces the threat of injuries while performing daily activities. When occurred, the severity of our injuries determines the extent of limitations that are imposed upon our otherwise unimpeded capacity for carrying out routines that lead to satisfaction and happiness. This threat appears even greater to elderly citizens, patients of post-accident disabilities and children with cognitive disabilities. Such occurrences would cause disastrous emotional impact that would further reduce their quality of life, disrupt commitment towards therapy and introduce complications into social-psychological development.

Injuries Faced by Elderly and Disabled Citizens

The effects of injuries are perhaps the most grievous when experienced by senior citizens. Due to limited reaction capacity towards the environment, they are highly susceptible to fires, scalds, poisoning and falls than their younger counterparts. In the United States alone, more than 2.3 million unintentional injuries have been reported annually, all of which occurred in and around the victims' homes. In the case of fires, 8 out of 10 elderly deaths were reported to have occurred in homes (Shields et al., 2013). In 2000, almost 750000 hospitalizations of elderly patients have been due to injuries that resulted from falls. 314006 of those cases were diagnosed with hip fractures (Boyé et al., 2012).

There are many factors that contribute to explaining why citizens above the age of 65 suffer more grievously from injuries at home. Aging causes diminishing senses of the immediate surroundings, making it difficult for them to readily perceive the occurrence of environmental hazards such as fires, floods and gas-leaks. Medical complications such as ailments and decreasing physical aptitude make the effects from burns, scalds, falls and cuts more severe. The same medical reason makes the recovery process much slower as well, as compared to younger patients (Shields et al., 2013).
In the case of elderly people as well as patients suffering from post-accident disabilities, discussions of injuries usually highlight those caused by falls. An average of 30% of home-dwelling senior citizens falls every year while half of those reported have done so repeatedly. Falls for elderly and disabled people often lead to extreme pain, severe limitations of functional capability and high expenses required for medical treatment and therapy. More than 1000 elderly people have died in Finland annually, due to falls, which is 4 times more than the number of traffic fatalities per annum in that country (Palvanen et al., 2014).

The elderly population of the world has been on the rise. There have been over 39.6 million elderly citizens over the age of 65 in the United States in 2009 alone. That number has been expected to grow to 72.1 million by the year 2030 (Shields et al., 2013). This is indicative that there will be a growing demand for medical care, nursing support and care services for elderly people. This demand is expected to include future monitoring systems that aim to prevent injuries for elderly and disabled people who are living alone at home.

**Injuries Faced by Children**

Children below the age of 18 years are also significant subjects who are injury prone, but their circumstances and susceptibility factors differ from elderly and disabled patients. Children are limited in terms of size, reach, maturity of their sense of smell, sight, hearing and balance (Gururaj, 2013). Coupled with limited experience and a sense of curiosity, children have a tendency of getting non-intentional injuries from their home environment. Unlike the case with senior citizens, a majority of child injuries result from dangerous environments which were managed by ill-informed or disadvantaged families (Mulvaney et al., 2013).

Child deaths in India are estimated to reach a million per year. 10 to 15 percent of those resulted from unintentional injuries. The same reason accounts for 20 to 30 percent of hospital registrations and 20 percent of disabilities. The leading causes of those injuries are road traffic injuries, burning, drowning, poisoning and falls, which draws much similarity to those that faced by elderly and disabled citizens. Globally, The Alliance for Safe Children (TASC) and the United Nations Children’s Fund (UNICEF) has revealed that 20 percent of children deaths are due to injuries. 60 percent of those are due to the same 5 leading causes. The survey in India shows that in 2009, reported 31 percent of injuries are road traffic originated, 17 percent by drowning, 12 percent by suicides, 16 percent by burns and 9 percent by poisoning (Gururaj, 2013).

In the case of children who suffer from cognitive impairments, the threat of injuries could be even more severe. Learning difficulties mean that they will need much more time to understand and adhere to safety habits. While the circumstances differ from causes of injuries for elderly and disabled people, it seems that most of the leading causes of children injuries could also be prevented with sufficient monitoring and intervention habits.

**Existing Injury Prevention Systems**

Most research and study concentrate on falls being the primary cause of injuries, as the treatment for surviving from falls are very critical. A study by Yu (2008) confirms that falls are the biggest threat to elderly and disabled patients, suggesting that detection of falls before it is completed is of utmost importance. The study categorized falls into 3 classes based on the initial body condition prior to a fall: from sleeping, standing or sitting (Yu, 2008). Present approaches to fall detection are generally found in one of 3 forms. They are either wearable, ambient or vision based solutions (Mubashir et al., 2013). As falls are monitored similarly to other causes of injuries, it is believed that the same approach classes can be applied to describe injury prevention systems in general.

Wearable devices are the most common approach for injury prevention. The main idea behind them was to utilize multiple sensors strapped to the human body in order to identify and monitor health signatures according to which specific cause of injury to watch out for. In the case of fall detection, sensors are used to identify the posture and motion that the human body is current in. This helps in detecting the conditions that indicate a fall has or will about to occur, and notify the healthcare crew (Yu, 2008).

The current evolution of accelerometer-based wearable sensors is used in combinations to anticipate falls in progress. One study combines the use of smartphone connectivity and a pair of shoes with pressure sensors to detect the signature of an abnormal walk gait and notify caregivers of a possible fall in progress (Majumder et al., 2014).

Alternatively, a piezoresistive insole was developed to be integrated into a shoe for constant pressure monitoring of the foot function. The shoe is created for assisting gait training for people with walking difficulties. Its function can be adapted to identify whether the user has lost footing in the event of a fall (Canavese et al., 2014).

Identifiable injuries are not exclusive to those caused by falls. Some may result in applied strain to the body beyond its physical tolerances. One way of identifying such accidents is to examine the posture of the body. A study presented a new wearable system that incorporates a vest with multi-posture monitoring systems. It presents a highly portable, cost effective solution to accompany present health care system for tracking health as well as act as a warning system in case of accidents (Lin et al., 2014).
Wearable solutions are usually the most cost effective and easy to implement alternatives to the other forms of injury detection systems. However, mounting sensors to the human body and forcing the constant use of devices are intrusive efforts that work against the will of their users. A high degree of commitment and will is needed for the patient to consistently utilize those devices.

The ambient approach seeks to eliminate the intrusiveness of wearable devices through embedding sensors into the living environment instead. One application involves installing vibration sensors into the floor of the premises. Electronically simple to implement, this system can detect impacts on the floor and pinpoint which location they occur at. However, installation of such a system would require an overhaul of the apartment’s flooring. Another example utilizes bladders installed into the bed of a patient. Through sensing the distribution of fluid within those bladders, the system can perceive how the patient shifts or whether a fall from sleeping position has possibly occurred (Yu, 2008).

One solution of consolidating multiple wearable sensors and systems embedded into the environment was presented as an implementation of Ambient Assisted Living system (AAL). The AAL describes a new dimension of smart homes and living spaces that incorporate components of telemetric health monitoring, indoor localization, emergency detection and so forth. Used in combination, an AAL environment can be used to effectively detect occurrences of falls, possible fires from an untended stove, as well as critical health conditions. This is made possible through multi-modal sensory and prediction algorithms (Spasova & Iliev, 2014).

Another implementation of intelligent ambient environments for injury prevention is proposed as an integrated autonomous system of sensors, alarms and triggers. Similar to the AAL, this system also monitors health conditions, and incorporates electrochemical actuators which perceive those conditions and trigger adjustments in the environment for relief. In the event of abnormal conditions or if a fall was detected, an emergency speed dial system helps to alert emergency personnel (Taraporewala, 2014).

Other ambient approaches to injury prevention can be easily installed. One prime example is a smoke detector mounted in the kitchen or fire extinguishers located at conveniently reached areas. The downside of these current systems is that they rely on the affected individual to personally react to them. Children, elderly and disabled people would be severely disadvantaged in these instances even with the presence of current ambient fixtures. Also, though not having to contend with approach intrusiveness or requiring user commitment, ambient systems are more expensive and difficult to install into place.

Camera-based or vision-based approach to injury detection involve the use of cameras and image processing in place of Micro-Electro-Mechanical Systems (MEMS) in order to discern a fall taking place. Instead of mounting devices on the body or the environment, a camera system which utilizes infra-red or 3D imaging to capture depth images is used. This raw data can be processes in order to detect the shape, size and orientation of humans, after which a skeleton can be applied to represent the posture and position of those subjects (Mubashir et al., 2013). Though training and calibration, input images can be compared with a database of poses to determine whether a fall has occurred.

One key implementation of this approach was carried out in 2013, using a Kinect sensor by Microsoft to perform visual monitoring. This system tracks a human subject by comparing the current pose with a set of templates that indicate over 14 different activities ranging from standing, walking and sitting to drinking and eating. This system logs those activities and captures snapshots for real time notifications over the Internet and Short Messaging System (SMS). The system also determines the vertical threshold of the skeleton position in the capture image, referred to as the Y coordinate range limit (MYREF). MYREF is used to see if the skeleton profile shows a fall has occurred (Fig. 11) (Ann et al., 2013).

Vision-based systems are simpler to install and relocate while not being intrusive to the patient. However, the detection performance is subjected to a much wider spectrum of factors compared to the other 2 approaches. Yu (2008) indicated that vision systems detect fall using inactivity detection, shape change analysis and 3D head motion analysis as components. However, due to the environmental inconsistencies in lighting, subject orientation and distance as well as false detections contribute to a degree of misinterpretations and false reporting for all 3 components. The Kinect-based activity monitoring study also reports the problem of a blind-spot, which was to be reduced by using a motorized gimbal that pans the sensor to track the subject (Fig. 12) (Ann et al., 2013).

Sensors such as the Kinect require the subjects to be within a specific optimal distance and orientation in order to be effective. While the Kinect itself is specified for 0.5 to 5 meters optimum range, the loss of resolution is exponential to the distance (Mankoff & Russo, 2013). Also, the effective viewing angle is only within 57.8 degrees (I Heart Robotics, 2010). In addition, specific image processing software such as the Kinect SDK are optimized for front-facing subjects for profiling. These limitations may be acceptable for conventional gesture-based gameplay (as the device is originally developed for) but is not ideal for tracking patients who wander around the living environment.

**RESEARCH PROBLEM**

Exploring the current available companion robots have shown that while they are mainly developed for social and activity support in place of human caregivers,
most of them do not readily support comprehensive monitoring and injury detection functionality to sufficiently achieve injury prevention. On the other hand, available injury detection systems employ intrusive wear or are too expensive to acquire and install. Even vision-based systems which are both affordable and simple to set up suffer from image processing hardware limitations. Requiring that the subject maintain position and orientation for optimal detection is not realistic and feasible, especially when the system is expected to provide monitoring while the subject is preoccupied with daily routines or have cognitive impairments. In addition, most fixed vision-based injury detection systems only extend to push notifications, alerting caregivers of occurrences. It is the responsibility of the caregivers to personally reach the subject in order to respond.

PROPOSED SOLUTION

This paper presents a concept of merging the form and functionality of companion robots with the potential of vision-based injury detection. The target demographic for this proposed concept is the children with cognitive disabilities who require constant monitoring. The self-sustaining capabilities of companion robots make it possible for them to serve as autonomous mobile platforms for vision-based injury detection systems that can constantly relocate and readjust themselves in an indoor environment in order to acquire the subject within optimal detection conditions. Detection of activities and possible poses that lead to injuries can be reported to the caregiver via network connectivity.

Another benefit of utilizing companion robots in this manner is that most of the current platforms are capable of supporting telepresence functions. Upon successful connection to the robot via network, a caregiver may remotely pilot the robot while maintain video and audio communications to the robot’s surrounds, similar to teleconferencing. When notified of a possible injury-inducing pose, the caregiver can connect to the robot to assume direct control in order to personally survey the situation and respond.

An ideal implementation of this robot will require a degree of hardware and software modularization. This is because the existence of a wide variety of companion robot platforms, remote connectivity software and vision systems with differing operational characteristics. Therefore, rather than proposing a prototype with specific components be built, this paper presents a first look at a conceptual architecture and system execution template to illustrate how a companion robot aimed at injury prevention be developed.

CONCEPTUAL DESIGN

A conceptual template was developed to illustrate how an assistive robot for children with disabilities should consist of both interchangeable hardware components and modular software scripts (Fig. 13). This basic roadmap visualizes how a robot for accompanying disabled children should be reconfigurable according to the nature of the monitoring or companion needs, as well as support reprogrammable behaviors for interchangeable activities and roles.

Beginning from the hardware, the robot system is composed of input and output components. Input components can either be sensors used by the robot to navigate and perceive the environment while the direct interface devices are used by human operators to control it. Output components range from contact to visual implementations in order to express output to the environment.

The software of the robot should consist of multiple module scripts that are interchangeable based on the role of the robot that is desired (such as a monitor, playmate, or assistant). Each script is structured to have 5 layers of modes that define each role in terms of programming. The modes determine what mode of interaction the robot is using to communicate with the environment, what communication software packages are needed to interface with devices, what type of therapy operation activity is available, which bank of recorded Interventions to use when interacting autonomously, and all necessary embedded processing needed for the robot to run.

The interaction design of the companion robot for injury prevention is shown in Fig. 14. The Use Case diagram shows basic interaction activities which include detection of danger poses, notifications towards caregiver, option to establish telepresence, sending real-time status updates and autonomous actions such as maintain orientation, distance and avoiding collision.

In applying the conceptual framework for a reconfigurable multi-purpose assistive robot to implement the desired interaction, the hardware components will include a robot platform that is fully equipped with proximity sensor arrays, actuators for mobility, wireless network interface hardware for connectivity, a Kinect vision system setup, a dedicated image processing system for visual servoing, a tablet PC for audiovisual telecommunication and control. The software platform will initially be populated by a base script that defines the robot’s monitoring role. Its interaction mode is interchangeable between autonomous and telepresence according to the caregiver’s preference. Software Interfaces are required to work with Kinect sensor, communication to additional image processing hardware for orientation and distance maintenance, network connectivity to host a media server and controlling the robot’s mobility. Since this role is purely for monitoring, there will not be any therapy Operations. Interventions need to be defined for preset situations when operating autonomously. Finally, embedded programming is required for visual servoing (autonomous navigation to acquire the target child in optimal orientation and distance
for the Kinect sensor to work) and constant collision detection.

The process strategy for each execution loop is visualized in the Activity Diagram shown in Fig. 15. During each execution loop, the robot may either be in autonomous or telepresence mode. This mode is toggled by the caregiver remotely – a condition that is checked for at the end of the execution loop. During telepresence mode, the caregiver has complete navigational control and audio-visual communication with the environment. If run autonomously, the robot will in turn avoid collision, perform visual servoing to maintain optimal injury detection, check for danger poses and update real-time feed of activity logs and pictures to the media server for remote access by the caregiver. If a danger pose is detected, it will output a prerecorded message matching the detected situation and post a notification to the caregiver. It is up to the caregiver whether to establish telepresence control of the robot or otherwise.

DISCUSSION & CONCLUSION

Planning the development of a companion robot that is modularly equipped for vision-based injury detection raises several implementation issues. One such concern is the operation of the vision system and visual servoing using the same camera hardware. Image processing for these two operations may result in performance deterioration as a result of insufficient computational resources. The planned solution for this problem was to use a separate embedded system for visual servoing. Visual servoing uses patterns mounted on the human subject to identify his/her orientation and distance from the robot. Autonomous navigation will ensure that the robot is repositioned so that optimal orientation and distance from the injury detection system is achieved.

Another issue to consider is the type of mobile platform to use according to the operating environment. Units that support stair-climbing or outdoor traversal would be too costly and may not have the same load-bearing capacity required to carry the necessary equipment for injury prevention. A basic platform for operating in a single story apartment scenario is preferred for an initial prototype build.

Being a companion robot, the proposed build should be expected to include similar social and therapy support functionalities. However, doing so may detract development efforts towards injury prevention via monitoring, while overlapping research on dedicated robot companions. Thus this paper proposes a prototype build that is dedicated for the purpose of monitoring and intervention for injury prevention. Care must also be taken to implement a visual appearance that would not attract too much attention from the subject as the vision system may not be able to function well if the subject is constantly adjacent to the robot.

Yet another concern involves the effectiveness of interventions administered in person versus through telepresence. Studies have suggested that children with disabilities have an increased affinity towards non-human robots than human companions (Colton et al., 2009). This may not apply to robots that act as avatars for human caregivers. Also, the administration of interventions through a robot may not carry the same weight of effect as it would have if done personally.

In concluding, with the wide variety of companion and telepresence robots in the consumer market, there is limited doubt that their prevalence at home will gradually grow. Its eventual presence as a daily companion opens up new opportunities for human-centric technologies, including those that address the need for preventing accidents and injuries.

This paper has suggested the utilization of companion robots for augmenting the effectiveness of vision-based injury detection systems. By integrating telepresence functionality with real-time notifications, a complete package for injury detection and intervention for children with cognitive disabilities can be implemented. A general framework for organizing the components of a modular assistive robot is presented and used to describe the key elements for this robot design. Finally, a process loop design for the robot’s behavior is proposed, illustrating the strategy of monitoring, notification, telepresence and intervention method presented in this paper.

REFERENCES


FIGURES

Fig. 1 ‘Paro’ (left) and ‘Matilda’ robots (right)

Fig. 2 Laboratory simulation of ASTRO

Fig. 3 Scitos G3 Platform

Fig. 4 Multi-wheelchair system in formation

Fig. 5 Amigo and the final concept shell

Fig. 6 MATS

Fig. 7 Kaspar

Fig. 8 NAO

Fig. 9 IROMEC

Fig. 10 The in-car companion

Fig. 11 MYREF indicating a fall

Fig. 12 Panning gimbal
Fig. 13 Conceptual framework of a reconfigurable multi-purpose assistive robot template for children with physical and cognitive disabilities

Fig. 14 Interaction Design
Fig. 15 Execution Loop Strategy
MODELLING AND VALIDATION OF A DFIG WIND TURBINE

Md Ayaz Chowdhury\textsuperscript{a}, Weixiang Shen\textsuperscript{b}, Nasser Hosseinzadeh\textsuperscript{c}, Hemanshu Roy Pota\textsuperscript{d}

\textsuperscript{a}Faculty of ECS, Swinburne University of Technology Sarawak Campus, Kuching, Malaysia
\textsuperscript{b}Faculty SET, Swinburne University of Technology, Hawthorn, VIC, Australia
\textsuperscript{c}Department of ECE, Sultan Qaboos University, Muscat, Oman
\textsuperscript{d}School of EIT, University of New South Wales, Canberra, ACT, Australia

ABSTRACT

This paper presents modelling and validation of a Doubly-fed induction generator (DFIG) wind turbine. The model is represented in terms of behavior equations of each of the subsystems, mainly the turbine rotor, the drive train, the induction generator, the power converters and associated control systems and a protection system. Simulation results obtained from the models are compared to the field measurement data. It is concluded that the model is reasonably accurate and can hence be used for representing wind turbines in power system dynamics simulations.

NOMENCLATURE

\begin{align*}
V_W & = \text{Wind speed} \\
P_m & = \text{Mechanical power} \\
\rho & = \text{Air density} \\
A & = \text{Swept area of the blades} \\
C_p & = \text{Power coefficient} \\
\lambda & = \text{Tip speed ratio} \\
\beta & = \text{Pitch angle} \\
T_m & = \text{Mechanical torque} \\
T_e & = \text{Electromagnetic torque} \\
H_t, H_g & = \text{Turbine and generator inertia constant} \\
\omega_t, \omega_g & = \text{Turbine and generator rotor angular speed} \\
\theta_t, \theta_g & = \text{Turbine and generator rotor angle} \\
K_s & = \text{Shaft stiffness} \\
D_t, D_g & = \text{Turbine and generator self-damping} \\
D_m & = \text{Mutual damping} \\
v_s, v_r & = \text{Stator and rotor voltage} \\
i_s, i_r & = \text{Stator and rotor current} \\
R_s, R_r & = \text{Stator and rotor resistance} \\
\phi_s, \phi_r & = \text{Stator and rotor flux linkage} \\
L_l & = \text{Leakage inductance} \\
L_m & = \text{Mutual inductance} \\
p & = \text{Number of poles} \\
P_e & = \text{Total output active power} \\
Q_e & = \text{Total output reactive power} \\
P_{st}, P_r & = \text{Stator and rotor output active power} \\
Q_{st}, Q_r & = \text{Stator and rotor output reactive power} \\
V_{DC} & = \text{DC link capacitor voltage} \\
C & = \text{DC link capacitance} \\
P_1 & = \text{Converter output active power} \\
Q_1 & = \text{Converter output reactive power} \\
R_{DC} & = \text{Equivalent resistance of the DC chopper} \\
K_{CP} & = \text{A parameter that depends on the geometry of the wind turbine} \\
\beta_c & = \text{Command signal for mechanical servo system} \\
\omega_{g,ref} & = \text{Upper limit (reference) value for the speed controller} \\
T_d & = \text{Time constant for the servomotor} \\
K_P & = \text{Proportional gain of the PI controller} \\
K_I & = \text{Integral gain of the PI controller} \\
\tau & = \text{Current loop time constant}
\end{align*}

INTRODUCTION

Wind power has been the fastest growing energy source since last decade due to its inherent attribute of the reproducible, resourceful and pollution-free characteristics. Higher capacity wind farms are integrated into high voltage transmission networks (Slootweg, et al., 2003). With the increasing amount of wind power penetration in power systems, wind farms begin to influence power systems. This justifies the need of adequate models for wind farms in order to represent overall power system dynamic behaviour of grid-connected wind farms during both normal operations and grid disturbances.

This paper represents the dynamic wind turbine model in terms of behavior equations of each of the subsystems, mainly the turbine rotor, the drive train, the induction generator, the power converters and associated control systems. Simulation results obtained from the models are compared to the field measurement data. A qualitative comparison of the field measurement data and the simulation results is carried out. A quantitative comparison is not possible due to rotor wake and lack of ability of a single anemometer for adequate measurement of $V_W$ acting on the large surface of the rotor.
DYNAMIC MODEL OF DFIG WIND TURBINE

A DFIG wind turbine comprises of an induction generator (IG) and a wind turbine connected through two shafts with a gearbox in between with the generator stator directly connected to the grid and the generator rotor connected to the grid through two back-to-back insulated gate bipolar transistors (IGBT) pulse width modulator (PWM) converters with an intermediate DC link capacitor (see Figure 1). It means that the generator is fed from both stator and rotor sides. In this paper, the dynamic wind turbine model is represented in terms of behavior equations of each of the subsystems, mainly the turbine rotor, the drive train, the induction generator, the power converters and associated control systems.

Turbine Rotor Aerodynamic Model

The wind turbine rotor, that extracts the kinetic energy from the wind and converts it into $P_m$, is a complex aerodynamic system. An algebraic relation between the $V_w$ and $P_m$ extracted by the wind turbine is assumed, which can be expressed as (Hansen, et al., 2008)

$$P_m = \frac{\rho AC_p V_w^3}{2}$$

(1)

The power coefficient corresponds to maximum mechanical power extraction from wind and is a function of $\lambda$ and $\beta$. $C_p(\lambda, \beta)$ characteristic of a turbine aerodynamic model is approximated by nonlinear functions from data provided by the manufacturer (Chowdhury, et al., 2013)

$$C_p(\lambda, \beta) = 0.22 \left( \frac{116}{\lambda_i} - 0.4 \beta - 5 \right) e^{-12.5 \lambda_i}$$

(2)

where

$$\frac{1}{\lambda_i} = \frac{1}{\lambda + 0.08 \beta} - \frac{0.035}{\beta^3 + 1}$$

(3)

$T_m$ applied to the shaft can be easily computed as

$$T_m = \frac{P_m}{\omega_i}$$

(4)

Drive Train Model

In power system stability studies, it is justified to include a two-mass model of the drive train (Pedersen, et al., 2003; Chowdhury, et al., 2010; Salman and Teo, 2003). Two-mass model includes only the relatively soft low-speed shaft neglecting the gearbox and the high speed shaft of the wind turbine, which is assumed to be infinitely stiff (Slootweg, 2003).

In two-mass model, the rotor is conventionally treated as two lumped masses, i.e. turbine mass and generator mass are connected together by a shaft with a certain damping and stiffness coefficient values. Turbine mass includes lumped inertia of the turbine, part of the gearbox and the low-speed shaft and generator mass includes generator rotor mass, high speed shaft along with its disk brake and the rest part of the gearbox (García-Gracia, et al., 2008).

As shown in Figure 2, the structure of the drive train consists of two inertias. Three different damping components are present in the model: $D_t$, $D_g$ and $D_m$. $D_t$ represents the aerodynamic resistance that takes place in the turbine blade, $D_g$ represents mechanical friction and windage and $D_m$ balances different speeds between the generator rotor and the turbine shaft. The mathematical equations of a two mass drive train model obtained by neglecting $D_t$ and $D_g$ are given as (Perdana, 2008)

$$2H_t \frac{d\omega_t}{dt} = T_m - K_s(\theta_g - \theta_t) - D_m(\omega_g - \omega_t)$$

(5)

$$2H_g \frac{d\omega_g}{dt} = -T_e + K_s(\theta_g - \theta_t) + D_m(\omega_g - \omega_t)$$

(6)

$$\frac{d\theta_t}{dt} = \omega_t$$

(7)

$$\frac{d\theta_g}{dt} = \omega_g$$

(8)

Generator Model

For power system dynamics simulations, stator and rotor flux dynamics are included in the induction generator model. The model is also referred to as the 5th order induction generator model, which is also known as the electromagnetic transient (EMT) model.

In modelling of the IG, the following conventions are considered:

- The generator structure is symmetrical and three-phase balanced
- Magnetic saturation is neglected
- Flux distribution is sinusoidal
- All losses are neglected except for copper losses
- The sum of the stator current equals zero
- Generation convention is considered, i.e. positive $P_e$ and $Q_e$ are fed into the grid
- All parameters are given in per unit quantities

The reference frame theory is implemented to facilitate the simulation in wind energy conversion. A synchronously rotating $dq$ reference frame with arbitrary rotating speed $\omega$ is chosen to model the IG in this paper. The IG space-vector model is first detailed, which is composed of three sets of equations: voltage equations, flux linkage equations and motion equations. Then, the $dq$-axis model of the IG is obtained from the space-vector model.

Reference Frame Transformation

Variables in the $abc$ stationary frame are transformed to the $dq$ rotating frame with an arbitrary speed $\omega$, which relates to $\theta$ by

$$\omega = \frac{d\theta}{dt}$$

(9)
This is done by deriving simple trigonometric functions from the orthogonal projection of the variables \( x_a, x_b, \) and \( x_c \) to the \( dq \)-axis variables \( x_d \) and \( x_q \). Figure 3 shows that the \( d \)-axis component is oriented along the phase \( a \) at \( t=0 \), which starts rotating at a speed of \( \omega \) at \( t=0^+ \). The \( q \)-axis component is \( 90^\circ \) ahead of the \( d \)-axis with respect to the rotation direction (García-Gracia, et al., 2008).

The transformation of \( abc \) variables in the \( dq \) frames can be expressed as (Wu, et al., 2011)

\[
x_d = \frac{2}{3}(x_a \cos \theta + x_b \cos(\theta - 2\pi/3) + x_c \cos(\theta - 4\pi/3)) \quad (10)
\]

\[
x_q = \frac{2}{3}(x_a \sin \theta + x_b \sin(\theta - 2\pi/3) + x_c \sin(\theta - 4\pi/3)) \quad (11)
\]

A coefficient of 2/3 is added to the equation so that the magnitude of the two-phase voltage is equal to that of the three-phase voltage after the transformation.

**State Space Model**

The voltage equations for the stator and rotor of the generator in the arbitrary reference frame are given by (Wu, et al., 2011)

\[
\begin{align*}
v_s &= R_s i_s + \frac{d\phi_s}{dt} + j\omega\phi_s \quad (12) \\
v_r &= R_r i_r + \frac{d\phi_r}{dt} + j(\omega - \omega_g)\phi_r \quad (13)
\end{align*}
\]

The flux linkage equations for the stator and rotor of the generator in the arbitrary reference frame are given by (Wu, et al., 2011)

\[
\phi_s = (L_s + L_m)\varphi_s + L_m i_r \quad (14)
\]

\[
\phi_r = (L_r + L_m)\varphi_r + L_m i_s \quad (15)
\]

Drive train equations along with \( T_e \) equation are termed as motion equations. \( T_e \) can be expressed as

\[
T_e = \frac{3}{2} P \text{Re} \left( \frac{\vec{\phi}_s^* \vec{\phi}_r}{} \right) \quad (16)
\]

The above equations constitute the vector-space model of the IG, whose equivalent circuit representation is given in Figure 4. The generator model is in the arbitrary reference frame, rotating in space at the arbitrary speed \( \omega \).

**Generator Model in the \( dq \) Reference Frame**

The \( dq \)-axis model of the IG can be obtained by decomposing the space-vectors into their corresponding \( d \) and \( q \)-axis components as

\[
x = x_d + jx_q \quad (17)
\]

where \( x \) can be voltage or current or flux linkage in the stator or rotor side of the generator.

The \( dq \)-axis voltage equations of the IG are thus obtained as (Chowdhury, et al., 2011)

\[
v_{ds} = R_s i_{ds} + \frac{d\phi_{ds}}{dt} - \omega\phi_{qs} \quad (18)
\]

\[
v_{dr} = R_i i_{dr} + \frac{d\phi_{dr}}{dt} - (\omega - \omega_g)\phi_{qr} \quad (20)
\]

\[
v_{qr} = R_i i_{qr} + \frac{d\phi_{qr}}{dt} + (\omega - \omega_g)\phi_{dr} \quad (21)
\]

where

\[
\phi_{ds} = (L_s + L_m) i_{ds} + L_m i_{dr} \quad (22)
\]

\[
\phi_{qs} = (L_s + L_m) i_{qs} + L_m i_{qs} \quad (23)
\]

\[
\phi_{dr} = (L_r + L_m) i_{dr} - L_m i_{ds} \quad (24)
\]

\[
\phi_{ds} = (L_r + L_m) i_{qr} - L_m i_{qs} \quad (25)
\]

\[
T_e \text{ is expressed in (Salman and Badrzadeh, 2004) as}
\]

\[
T_e = \frac{3}{2} P \left( \phi_{ds} i_{qs} - \phi_{qs} i_{ds} \right) \quad (26)
\]

\[
P_s \text{ and } Q_s \text{ can be written as (Salman and Badrzadeh, 2004)}
\]

\[
P_s = \frac{3}{2} \left( v_{ds} i_{ds} + v_{qs} i_{qs} \right) \quad (27)
\]

\[
Q_s = \frac{3}{2} \left( v_{qs} i_{ds} - v_{ds} i_{qs} \right) \quad (28)
\]

**Power Converter Model**

The power converter is made up of two back-to-back DC linked converters, namely: rotor side converter (RSC) and grid side converter (GSC), which are connected to the rotor winding and grid, respectively. This is known as ‘Scherbius scheme’. The converters are typically made of voltage-fed current regulated inverters, which enable a two-directional power flow. The inverter valves make use of IGBTs provided with freewheeling diodes (Figure 5) (Perdana, 2008).

The RSC acts on rotor current components to ensure the independent regulation of \( P_s \) and \( Q_s \). The GSC helps to keep \( V_{DC} \) constant acting on grid current components. It also regulates \( Q_s \) during voltage re-establishment after a grid fault condition. Both RSC and GSC are modelled as current controlled voltage sources. The switching dynamics of the converters are neglected, since the PWM modulation frequency is much higher than the system frequency. These converters are assumed lossless and hence the DC link capacitor dynamics can then be described according to (Perdana, 2008) as

\[
\frac{dV_{DC}}{dt} = \frac{1}{C} \left( \frac{P_s - P_t}{V_{DC}} \right) \quad (29)
\]

The DC chopper consists of a resistor and a switch connected to the DC bus, in parallel with the DC link capacitor. This is controlled by an IGBT when a DC overvoltage is detected. It dissipates the excess of energy that cannot be evacuated to the grid during a fault. When the DC chopper is activated, Equation 29 becomes
\[
\frac{dV_{DC}}{dt} = \frac{1}{C} \left( \frac{P_e - P_i}{V_{DC}} - \frac{V_{DC}}{R_{DC}} \right)
\]  
(30)

Control System Model

There are two levels of control in the wind energy conversion system. The high level control or speed control actuates \( \beta \) and gives torque reference signals to the converter. The low level control or converter control drives the converter IGBTs to meet its control objectives.

Speed Controller Model

Speed controller has two different objectives depending on the region where the machine is operating. When the DFIG wind turbine is partially loaded, the mission of the speed controller is to maximize the power extracted from the wind referencing the proper electromagnetic torque signal to the converter. The electromagnetic torque reference is calculated using (Trilla, et al., 2011)

\[ T_e^* = K_{cp} \cdot \omega_i^2 \]  
(31)

When the DFIG wind turbine is fully loaded, the task of the speed controller is to keep extracting the nominal power varying \( \beta \) to reduce \( C_p \) to maintain a constant \( T_e \). It is expected that the \( \omega_i \) will eventually be controllable before it reaches its upper limit.

Figure 6 shows a conventional simplified PI controller for the purpose of controlling \( \beta \). Pitch angle controller regulates the output in accordance with the error between \( \omega_g \) and \( \omega_{g,ref} \). The error signal is then sent to the PI controller generating \( \beta_c \).

The choice of \( \omega_{g,ref} \) is totally system dependent. This is chosen so that its minimum possible value enables generation of maximum possible power (1 p.u.) as higher \( \omega_g \) is vulnerable to power system instability. For the specific wind turbine system adopted in Reference (Chowdhury, et al., 2012), \( \omega_{g,ref} \) is chosen as 1.21 p.u.

The mechanism governing \( \beta \) is usually a hydraulic actuator or a servomotor that can be modelled using a first order delay system with \( T_d \) as (Valsera-Naranjo, et al., 2011)

\[ \beta = \frac{1}{1 + T_d \beta_c} \]  
(32)

The inclusion of the servomotor along with the rate limiter provides a realistic response from the controller.

Converter Controller Model

RSC Controller

The RSC acts on rotor current components to ensure the independent regulation of \( P_i \) and \( Q_i \). Figure 7 shows stator field oriented control of the RSC.

GSC Controller

The main objective of the GSC is to maintain \( V_{DC} \) constant regardless of the magnitude and direction of the slip power. A current controlled PWM scheme is used, where \( d \) and \( q \) axes currents are used to regulate \( V_{DC} \) and \( Q_i \). Figure 8 shows stator field oriented control of the GSC.

The main task of the control scheme in the PWM converters is to force the current to follow their reference signals prior to the error signal generated from the reference and actual signal fed into the PI controllers. PI controllers are tuned using internal model control (IMC) method as (Nunes, et al., 2004)

\[ K_p = \frac{L}{\tau} \]  
(31)

\[ K_i = \frac{R}{\tau} \]  
(34)

Parameters of the DFIG wind turbine along with its control system used in the simulation studies of this paper are shown in Table 1 (Slootweg, et al., 2003 and Chowdhury, et al., 2012).

MODEL VALIDATION

The responses of the simulated model to a particular \( V_w \) sequence is investigated and compared to actual measurements (Figure 9a). The field measurement data are obtained from a wind turbine manufacturing company under a confidential agreement.

The available field measurement data cannot be used for a quantitative validation of the model because \( V_w \) is measured with single anemometer which is unable to provide adequate measurement of \( V_w \) acting on the rotor as a whole as the rotor has a large surface. Its location on the nacelle also disturbs the \( V_w \) measurement by activating rotor wake. From the measured responses in Figure 9, it is observed that there is not a reasonable match between different variables of the DFIG wind turbine from about \( t=27 \) s to \( t=38 \) s because with an increasing trend of \( V_w \) during this period, \( \omega_g, P_e \) and \( Q_e \) decreases with no generation of \( \beta \) as expected.

This shows a clear picture of inadequacy of \( V_w \) measurement with a single anemometer. Therefore, it is not possible to feed a measured \( V_w \) sequence into the model in order to compare measured and simulated response of the DFIG wind turbine and hence only a qualitative comparison is carried out.

From the simulated responses of the DFIG wind turbine, when \( \omega_g \) exceeds 1.21 p.u. during two periods, one is between about \( t=11 \) s and \( t=16.5 \) s and the other is between about \( t=32 \) s and \( t=37 \) s, the pitch angle controller is activated to limit the mechanical power extraction (Figure 9c) and hence maintain \( \omega_{g,ref} \) to its control speed value. Therefore, \( C_p \) remains no more constant,
rather become a non-linear function of $\lambda$ and $\beta$ according to Equation 2.

This is how generated $P_e$ is saturated at its maximum rating value (1 p.u.) and $Q_e$, at about -0.0656 p.u. (negative value indicates that $Q_e$ is being absorbed by the wind turbine from the grid due to having capacitive load (Figure 9d and 9e).

In the rest of the period, $\omega_\text{r}$ never exceeds the upper limit value 1.21 p.u. In this situation, there is no generation of $\beta$ (Figure 9c) and $C_p$ remains constant. As a result, the $P_e$ flow on the grid and $Q_e$, absorbed by the wind turbine from the grid follow the $V_W$ curve proportionally (Figure 9d and 9e).

When the measured responses are compared with the simulated one, two dissimilarities have been observed. One is the behaviour of the DFIG wind turbine during the period between $t=27$ s and $t=38$ s, which is due to lack of ability of a single anemometer for adequate measurement of $V_W$ acting on the large surface of the rotor as stated before. The other is the time delay in the measured response, which is due to a finite time taken by the generation system to produce power according to the wind incident on the turbine blade, while simulation gives almost instantaneous response from the wind incident. A manipulation in inertia constant with an arbitrary coefficient may make higher degree of correspondence between measured and simulated response.

Otherwise, a high degree of similarity is observed in the rest of the period. Measured responses show that $P_e$ and $Q_e$ curve follow the $V_W$ curve proportionally, as well, when the DFIG wind turbine is partially loaded. Measured responses also show that $\omega_\text{r}$ exceeds 1.21 p.u. when the DFIG wind turbine is fully loaded and enables the activation of pitch angle controller acting in a similar way with respect to the rate of change (3 deg/s) as compared with the simulated response (Figure 9c). The generated active and $Q_e$ saturate at 1 p.u. and -0.0656 p.u. during these times, as well.

Although a quantitative validation is not possible with the available field measurement data, the qualitative comparison of the measured and simulated responses gives confidence about the accuracy and applicability of the derived model.

**CONCLUSION**

This paper accomplishes the development of a dynamic DFIG wind turbine model, basic unit of a DFIG wind farm with a goal that the model would be of simplest possible design with an optimum accuracy for proper power system dynamic behavior and stability investigation with wind energy integration within an optimum simulation run time. Mathematical description of all subsystems of the DFIG wind turbine, mainly the turbine, the drive train, the induction generator, the power converters with associated control systems and the crowbar, are elaborately explained. Since a quantitative validation of the model is not possible due to the inadequacy in measurement with a single anemometer and rotor wake, a qualitative validation of the developed model is carried out. It gives good confidence about the accuracy and applicability of the developed model.

**REFERENCES**


**FIGURES AND TABLES**

![Fig. 1 Configuration of a DFIG wind turbine](image1)

![Fig. 2 Drive train model](image2)

![Fig. 3 Reference frame for the generator equations](image3)

![Fig. 4 Equivalent circuit of an induction generator dynamic model](image4)

![Fig. 5 Power converter in DFIG wind turbine](image5)

![Fig. 6 Pitch angle controller](image6)

![Fig. 7 Stator flux oriented control of RSC](image7)
Table 1 Simulated DFIG Turbine Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wind Turbine</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inertia constant</td>
<td>2.5</td>
<td>s</td>
</tr>
<tr>
<td><strong>Generator</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominal power</td>
<td>1.5*0.9</td>
<td>MW</td>
</tr>
<tr>
<td>Nominal voltage (line to line)</td>
<td>690</td>
<td>V</td>
</tr>
<tr>
<td>Stator leakage resistance</td>
<td>0.0084</td>
<td>p.u.</td>
</tr>
<tr>
<td>Stator leakage inductance</td>
<td>0.167</td>
<td>p.u.</td>
</tr>
<tr>
<td>Rotor leakage resistance</td>
<td>0.0083</td>
<td>p.u.</td>
</tr>
<tr>
<td>Rotor leakage inductance</td>
<td>0.1323</td>
<td>p.u.</td>
</tr>
<tr>
<td>Mutual inductance</td>
<td>5.419</td>
<td>p.u.</td>
</tr>
<tr>
<td>Inertia constant</td>
<td>0.5</td>
<td>p.u.</td>
</tr>
<tr>
<td>Friction factor</td>
<td>0.01</td>
<td>p.u.</td>
</tr>
<tr>
<td>Number of poles</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td><strong>Converter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Converter maximum power</td>
<td>0.5</td>
<td>p.u.</td>
</tr>
<tr>
<td>Grid side coupling inductor inductance</td>
<td>0.15</td>
<td>p.u.</td>
</tr>
<tr>
<td>Grid side coupling inductor resistance</td>
<td>0.0015</td>
<td>p.u.</td>
</tr>
<tr>
<td>Nominal DC voltage</td>
<td>1200</td>
<td>V</td>
</tr>
<tr>
<td>DC capacitance</td>
<td>10</td>
<td>mF</td>
</tr>
<tr>
<td><strong>Controller</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grid voltage regulator gain</td>
<td>$K_p$</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>$K_I$</td>
<td>300</td>
</tr>
<tr>
<td>Droop</td>
<td>0.02</td>
<td>p.u.</td>
</tr>
<tr>
<td>Power regulator gain</td>
<td>$K_p$</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$K_I$</td>
<td>10</td>
</tr>
<tr>
<td>DC bus voltage regulator gain</td>
<td>$K_p$</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>$K_I$</td>
<td>0.05</td>
</tr>
<tr>
<td>GSC current regulator gain</td>
<td>$K_p$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$K_I$</td>
<td>100</td>
</tr>
<tr>
<td>RSC current regulator gain</td>
<td>$K_p$</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>$K_I$</td>
<td>8</td>
</tr>
<tr>
<td>Pitch angle regulator gain</td>
<td>$K_p$</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$K_I$</td>
<td>10</td>
</tr>
<tr>
<td>Maximum pitch rate of change</td>
<td>$±3$</td>
<td>°/s</td>
</tr>
<tr>
<td>System servo delay</td>
<td>0.25</td>
<td>s</td>
</tr>
</tbody>
</table>

Fig. 8 Stator flux oriented control of GSC

Fig. 9 Measured (solid lines) and simulated (dashed lines) responses from the DFIG wind turbine: (a) Wind incident, (b) Generator rotor speed, (c) Pitch angle, (d) Active power and (e) Reactive power
A NOVEL AGGREGATION TECHNIQUE FOR DFIG WIND FARM MODEL

Md Ayaz Chowdhury^a, Weixiang Shen^b, Nasser Hosseinzadeh^c, Hemanshu Roy Pota^d
^aFaculty of ECS, Swinburne University of Technology Sarawak Campus, Kuching, Malaysia
^bFaculty SET, Swinburne University of Technology, Hawthorn, VIC, Australia
^cDepartment of ECE, Sultan Qaboos University, Muscat, Oman
^dSchool of EIT, University of New South Wales, Canberra, ACT, Australia

ABSTRACT

A novel aggregated model for doubly-fed induction generator (DFIG)-equipped wind farms is proposed in this paper. In the proposed model, a mechanical torque compensating factor (MTCF) is integrated into a full aggregated wind farm model to deal with the nonlinearity when the DFIG is loaded partially and to make it behave as closely as possible to a complete model of the wind farm. The MTCF is initially constructed to approximate a Gaussian function by a fuzzy logic method and optimized on a trial and error basis. A large scale offshore wind farm comprising of 72 DFIG wind turbines is used to verify the effectiveness of the proposed aggregated model. The simulation results show that the proposed model is able to approximate collective dynamic responses at the point of common coupling with significant reduction in the simulation computation time.

NOMENCLATURE

\[
\begin{align*}
V_W &= \text{Wind speed} \\
T_m &= \text{Mechanical torque} \\
T_e &= \text{Electromagnetic torque} \\
V_{\text{avg}} &= \text{Average wind speed} \\
n &= \text{number of DFIG wind turbine} \\
\alpha_{\text{avg}} &= \text{Average turbine rotor speed} \\
\alpha &= \text{MTCF} \\
T_m^{\text{agg}} &= \text{Mechanical torque of the complete model} \\
T_m^{\prime\text{agg}} &= \text{Mechanical torque of the proposed model} \\
V_W &= \text{Wind speed deviation} \\
\sigma &= \text{Standard deviation from mid-point of the considered wind speed range} \\
l &= \text{Peak value of Gaussian function} \\
A_a, B_b &= \text{Antecedent part} \\
C_c &= \text{Consequent part} \\
\Pi &= \text{Grade for the antecedent} \\
Z_e &= \text{Equivalent impedance (EI) of the aggregated model} \\
Z_{\text{awt}} &= \text{EI of the internal electrical network of each individual DFIG wind turbine in the complete model} \\
Z_{awt} &= \text{Impedance of the DFIG wind turbine} \\
P_e &= \text{Active power} \\
Q_e &= \text{Reactive power} \\
n_{P_e} &= \text{Number of instantaneous values of active power} \\
n_{Q_e} &= \text{Number of instantaneous values of reactive power}
\end{align*}
\]

INTRODUCTION

Wind power has been the fastest growing energy source since the last decade due to its inherent attribute of the reproducible, resourceful and pollution-free characteristics. With the increasing amount of wind power penetration in power systems, wind farms begin to influence power systems. This justifies the need for adequate models for wind farms in order to represent overall power system dynamic behavior of grid-connected wind farms during both normal operations and grid disturbances. A wind farm may consist of tens to hundreds of wind turbines. This leads to model complexity and computation burden (Rodriguez-Amenedo, et al., 2002; Tapia, et al., 2001).

An aggregated wind farm model is required to simplify the complete wind farm model, provided that the aggregated model can represent the behavior \((P_e, Q_e)\) exchanged with the power system at the point of common coupling (PCC) of the wind farm during normal operation grid disturbances, such as voltage drops and frequency deviations.

Two types of wind farm aggregation techniques have been proposed: the full aggregated and the semi aggregation techniques. The full aggregated model consists of one equivalent wind turbine and one equivalent generator for a wind farm with one operating point at an average wind speed for all the wind turbines in the wind farm (Nunes, et al., 2004; Akhmatov, et al., 2002). The semi aggregated model consists of all the wind turbines in the wind farm and one equivalent generator (Fernández, et al., 2009; Chowdhury, et al., 2010). Figure 1 shows the complete and the aggregated wind farm models.
For a wind farm consisting of DFIG wind turbines, the ability of the full or semi aggregated model to approximate the complete model depends on the operating region of the DFIG wind turbines. The operating regions of the DFIG wind turbine adopted in this paper is shown in Figure 2, which can be segmented into two parts: a partial load region, where the wind speed ranges between 4.5 m/s and 14.5 m/s and a full load region, where the wind speed ranges between 14.5 m/s and 25 m/s). The DFIG wind turbine is stopped when wind speed is less than 4.5 m/s or greater than 25 m/s.

The full or semi aggregated model can represent the complete model when DFIG wind turbines in the wind farm operate in the full load region regardless of the differences in the operating points of the wind turbines in the wind farm since all generators produce the same current at its maximum rating in this region. But, the full aggregated model cannot provide an accurate approximation of a complete model when DFIG wind turbines in the wind farm operate in the partial load region. This is due to the fact that the full aggregation technique does not consider the operating points of all corresponding wind turbines in the wind farm and a nonlinear relationship between \( V_W \) and \( T_m \) (\( T_m \) is proportional to the cube of \( V_W \)) as shown in Figure 2.

The semi aggregated model, on the other hand, improves the approximation of a complete model in the partial load region by considering the operating points of all corresponding wind turbines in the wind farm. The use of an average generator rotor speed for all of the wind turbines still contributes to discrepancies in the magnitude of \( T_m \) and consequently \( T_e \).

This paper proposes a new aggregation technique with the incorporation of a mechanical torque compensation factor (MTCF) into the full aggregated wind farm model to enable approximation of collective dynamic responses at the point of common coupling with significant reduction in the simulation computation time.

FORMATION OF A COMPLETE DFIG WIND FARM MODEL

The wind farm model comprises of 72 DFIG wind turbines (Figure 3). Details of dynamic modelling and parameters of each unit of DFIG wind turbine can be found in References (Chowdhury, et al., 2012; Chowdhury, et al., 2010). Each DFIG wind turbine is connected to the cable sections through 0.67/30 kV transformer (LV/MV) and a line impedance of 0.08+j0.02 p.u. The wind farm is connected to the power grid through a 30/132 kV tertiary transformer (MV/HV) and then through a high voltage (132 kV) transmission network (HVTN) with the impedance value of 1.6+j3.5 p.u. Short circuit capacity viewed from the PCC into the HVTN is around 1500 MVA. The power grid is modeled by an infinite bus with the MVA-rating of 1000 MVA.

The parameters of the DFIG wind farm used in the simulation are shown in Table 1 (Akhmatov, et al., 2002; Fernández, et al., 2008).

PROPOSED AGGREGATED DFIG WIND FARM MODEL

Figure 4 shows the proposed aggregated DFIG wind farm model that consists of a mechanical torque compensating factor (MTCF) incorporated into a traditional full aggregated model.

The full aggregated DFIG wind farm model is driven by \( V_{agg} \)

\[
V_{agg} = \frac{1}{n} \sum_{i=1}^{n} V_{W_i}
\]  

(1)

This gives the mechanical torque as

\[
T'_{agg} = \frac{\rho MC\beta V_{agg}^3}{2\omega_{agg}}
\]  

(2)

\( \alpha \) is a multiplication factor to \( T'_{agg} \) that minimizes this inaccuracy in approximation. \( T_{agg} \) is thus calculated by

\[
T_{agg} = T_{agg} \times \alpha
\]  

(3)

Basis of MTCF Calculation

In the full load region, the MTCF takes a value equal to 1 as the full aggregated model can provide an approximation of the complete model in this region.

When the wind turbines operate at different wind speeds and thus different operating points in the partial load region, the adoption of an average operating point for the DFIG wind turbines causes the discrepancies between the complete and full aggregated models. Since \( T_m = V_W^3 \), \( T_m \) of the full aggregated model is generally lower than that of the complete model. Thus, in the partial load region, \( \alpha \) takes a value more than 1.

It means that \( \alpha \) increases from the value 1 as \( V_{agg} \) increases from 4.5 m/s or \( V_{agg} \) decreases from 14.5 m/s, which implies that \( \alpha \) may take its maximum value between 4.5 m/s and 14.5 m/s. On the other hand, \( \alpha \) maintains a proportional relation with \( V_{agg} \) and it takes a value equal to 1 when the operating points of the DFIG wind turbines in the wind farm are identical (i.e., \( V_{agg} = 0 \)). Thus, \( \alpha \) is a function of \( V_{agg} \) and \( V_{agg} \) may be ‘approximated’ by an ideal Gaussian function (see Figure 5) in the partial load region:

\[
\alpha = 1 + \frac{-(V_{agg} - V_{agg0})^2}{2\sigma^2} \]  

(4)

The empirical rule of the central limit theorem (Voelker, 2001) gives the value of \( \sigma \) and \( l \) as 1.25 and 0.32, respectively. In this paper, \( V_{agg} = 9.5 \) m/s (mid-point of the considered wind speed range).
MTCF Calculation by Fuzzy Logic System

The FLS is initially constructed by assigning overlapped triangular membership functions for the fuzzy sets and setting fuzzy rules based on the ideal Gaussian function. The design is optimized by making possible changes in membership functions for the fuzzy sets and fuzzy rules on trial and error basis to achieve less than 10% discrepancy between the proposed aggregated model and the complete model.

The FLS takes two inputs: \( V_{\text{agg}} \) and \( V_{w} \). In the design of the FLS, \( V_{\text{agg}} \) ranges between 4.5 m/s and 14.5 m/s. \( V_{w} \) ranges between 0 and its maximum possible value. The value of \( V_{w} \) is the maximum when wind speeds received by the wind turbines are equally spaced within the specified range of \( V_{\text{agg}} \). For 72 DFIG wind turbines, the maximum value of \( V_{w} \) is found by the following calculation:

\[
V_{w,\text{max}} = \frac{1}{72} \sum_{i=1}^{72} \left( \frac{14.5 - 4.5}{72 - 1} i - 9.5 \right)^2 = 5.25 \tag{5}
\]

Then, according to Eq. (4), \( \alpha \) is 2.7 when \( V_{\text{agg}} = V_{w} \) and \( V_{w}=5.25 \) (takes its maximum value), thus the range of \( \alpha \) should be between 1 and 2.7.

Figure 6 shows that triangular membership functions are assigned to each input or output variable. 7 membership functions have been selected for \( V_{\text{agg}} \), 7 for \( V_{w} \) and 8 for output \( \alpha \). Overall 49 (i.e., 7x7) rules are built by crossing the fuzzy sets, as shown in Table 2.

The 7th fuzzy rule is expressed as [10]

\[
\text{Rule 7: if } V_{\text{agg}} \text{ is } A_7 \text{ and } V_{w} \text{ is } B_7, \text{ then } \alpha(n) = C_c. \tag{6}
\]

The FLS gives the values of the MTCF (\( \alpha \)) by applying the center of gravity method (Senjyu, et al., 2006)

\[
\alpha(n) = \frac{\sum_{i=1}^{49} \Pi_i C_c / \sum_{i=1}^{49} \Pi_i}{49} \tag{7}
\]

\( \Pi \) is the product of grade for the antecedents of each rule.

Equivalent Internal Electrical Network

The aggregated wind farm must operate at an equivalent internal electrical network. Thus, the internal electrical network of each individual DFIG wind turbine in the complete model is required to be replaced by equivalent impedance in the proposed aggregated wind farm model. The short circuit impedance of the aggregated wind farm must be equal to that of the complete wind farm, which gives the calculation of \( Z_e \) (Fernández, et al., 2008)

\[
Z_e = Z_{\text{agg}} - Z_{\text{wt}} \frac{n}{n} \tag{8}
\]

SIMULATION RESULTS

MATLAB/Simulink software package is used to implement and simulate the wind farm model which consists of the nonlinear equations. Both the proposed aggregated model and the full aggregated model are simulated to obtain the dynamic responses at the PCC under the following two conditions: (1) normal operation and (2) grid disturbance. The variables considered for the comparison are \( P_e \) and \( Q_e \) exchange between the wind farm and power system.

Figure 7 shows the speeds of the wind received by the first DFIG wind turbines in each group, which have been adopted from Bureau of Meteorology, Melbourne. The time delay and wake effect are accounted for approximating wind speed for the following DFIG wind turbines in each corresponding group (Magnusson and Smedman, 1999; Chowdhury, et al., 2013; Perdana, 2008).

Normal Operation

The collective responses of the complete, full aggregated and the proposed aggregated wind farm models at the PCC during normal operation are shown in Figure 8.

The proposed aggregated model has a higher correspondence in approximating \( P_e \) (see Figure 8a). Comparing with the complete model, it has the maximum and average discrepancy of 2.94% and 2.35%, respectively. The multiplication factor MTCF, dynamically produced by a well-tuned FLS, manipulates the mechanical torque to compensate the existing nonlinearities in the wind farm in order to have a better approximation in the proposed aggregated model.

The proposed aggregated model has a higher correspondence in approximating \( Q_e \) as well (see Figure 8b). Comparing with the complete model, it has the maximum and average discrepancy of 5.45% and 4.36%, respectively. The multiplication factor MTCF for DFIG wind turbine depends on the \( P_e \) and the generation voltage. These variables differ in each DFIG wind turbine when the incoming winds are different. Therefore, it leads to different converter controller action for each DFIG wind turbine. This is not accounted for in the aggregated model resulting in lesser accuracy in the approximation of \( Q_e \) at the PCC.

Grid Disturbance

A voltage sag of 50% lasting for 0.1 s is originated at the PCC at \( t=1 \) s to evaluate the proposed aggregated...
wind farm model during grid disturbances, the collective responses of the complete, the full aggregated and the proposed aggregated wind farm models at the PCC are shown in Figure 9.

Figure 9 shows that the $P_e$ produced by the wind farm reduces and it goes to negative values for a short time (i.e. the grid supplies $P_e$ to the DFIG to keep it spinning) during grid disturbances. On the other hand, the $Q_e$ which is normally negative (which means the wind farm takes $Q_e$ from the grid) changes sign and increases during the disturbance. This means the wind farm supplies $Q_e$ to the grid during the disturbance caused by the voltage sag.

It also shows a high correspondence among the collective responses at the PCC of the complete, the full aggregated and the proposed aggregated wind farm models with negligible discrepancies on the active and reactive power. However, $P_e$ slightly mismatches in both aggregated models right after clearing the fault when different parameters start retaining their normal values. This high level of correspondence is partly due to the fact that the grid disturbances are much faster than the wind speed variations (Fernández, et al., 2008) and, therefore, the discrepancies during normal operations are unimportant during grid disturbances.

**EVALUATION OF THE PROPOSED TECHNIQUE**

In the following, the proposed aggregation technique is evaluated in terms of the accuracy in the approximation of the collective responses at the PCC, such as $P_e$ and $Q_e$, and simulation computation time.

**Accuracy is approximation**

The discrepancy between any instantaneous output power of the proposed aggregated model and that of the complete model can be calculated by (Trilla, et al., 2011)

$$\Delta x = \frac{x_{\text{comp}} - x_{\text{agg}}}{x_{\text{comp}}}$$

where $x$ can be either $P_e$ or $Q_e$. Suffix $\text{comp}$ denotes the complete model and $\text{agg}$ denotes the aggregated model.

The results of the accuracy in approximating the collective responses are shown in Table 3. It shows that less than 10% discrepancy has been achieved between the proposed aggregated model and the complete model. In normal operation, the proposed aggregated model approximates $P_e$ and $Q_e$ more accurately than the full aggregated model by 8.7% and 12.5%, respectively. However, in grid disturbance both models show the same level of accuracy.

**Simulation Computation Time**

The comparison of computation time for the complete and both aggregated wind farm models are made and the results are shown in Table 4 (In Table 4, CM means complete model, FAM means full aggregated model and PM means proposed model). The simulations are carried out on a personal computer with the following specifications: Intel (R) Pentium (R) Dual CPU E2200, 2.20 GHz, 1.96 GB of RAM.

It can be seen that the proposed aggregated wind farm model has higher simulation computation time than the full aggregated wind farm model by 2.38% and 3% during normal operation and grid disturbance, respectively. A slight increase in the computation time is caused by the additional computing block with the FLS to generate the MTCF. However, it has significantly reduced the simulation computation time by 90.3% and 87%, respectively, comparing with the complete model during normal operation and grid disturbance.

**CONCLUSION**

This paper describes the development of a novel aggregation technique with the incorporation of a mechanical torque compensation factor (MTCF) into the full aggregated wind farm model to obtain dynamic responses of a wind farm at the point of common coupling. The aim is to simulate the dynamic responses of the wind farm with an acceptable level of accuracy while reducing the simulation time considerably by using the aggregation technique. The MTCF is a multiplication factor to the mechanical torque of the full aggregated wind farm model that is initially constructed to approximate a Gaussian function by a fuzzy logic method and optimized on a trial and error basis to achieve less than 10% discrepancy between the proposed aggregated model and the complete model.

The proposed aggregated model is then applied to a bigger 120 MVA offshore wind farm comprising of 72 DFIG wind turbines. Simulation results show that the proposed aggregated wind farm model has the average discrepancy in approximating $P_e$ and $Q_e$ of 2.35% and 4.36%, respectively, during normal operation as compared to the complete model. But it has 8.7% and 12.5% more approximation capability of $P_e$ and $Q_e$, respectively, than the full aggregated model. However, the proposed aggregated model can mimic $P_e$ and $Q_e$ with negligible discrepancy during grid disturbance. Computational time of the proposed aggregated model is slightly higher than that of the full aggregated model but much faster than the complete model by 90.3% during normal operation and 87% during grid disturbance.

**REFERENCES**


Chowdhury M. A., Hosseinzadeh N. and Shen W. X., 2012, "Smoothing wind power fluctuations by fuzzy logic


Table 1 DFIG wind farm parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal network</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base power</td>
<td>( S_{WTG} )</td>
<td>1.5/0.9</td>
<td>MVA</td>
</tr>
<tr>
<td>Base voltage</td>
<td>( V_{WTG} )</td>
<td>575</td>
<td>V</td>
</tr>
<tr>
<td>LV/MV transformer</td>
<td>-</td>
<td>0.69/30</td>
<td>KV</td>
</tr>
<tr>
<td>( S_T )</td>
<td>2</td>
<td></td>
<td>MVA</td>
</tr>
<tr>
<td>( \varepsilon_{cc} )</td>
<td>6</td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>Line impedance</td>
<td>( Z_L )</td>
<td>0.08+j0.02</td>
<td>p.u.</td>
</tr>
<tr>
<td><strong>External network</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MV/HV transformer</td>
<td>-</td>
<td>30/132</td>
<td>KV</td>
</tr>
<tr>
<td>( S_T )</td>
<td>150</td>
<td></td>
<td>MVA</td>
</tr>
<tr>
<td>( \varepsilon_{cc} )</td>
<td>8</td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>HVTN impedance</td>
<td>( Z_T )</td>
<td>1.6+j3.5</td>
<td>p.u.</td>
</tr>
<tr>
<td>PCC short circuit capacity</td>
<td>( S_{PCC} )</td>
<td>1500</td>
<td>MVA</td>
</tr>
<tr>
<td>( (X/R)_{PCC} )</td>
<td>20</td>
<td></td>
<td>p.u.</td>
</tr>
<tr>
<td>Grid short circuit capacity</td>
<td>( S_G )</td>
<td>1000</td>
<td>MVA</td>
</tr>
</tbody>
</table>

Fig. 3 A 120 MVA offshore DFIG wind farm model

Fig. 5 Gaussian distribution of \( \alpha \) with respect to \( V_{Wagg} \)

Fig. 6 Membership functions: (a) \( V_{Wagg} \), (b) \( V_{Wo} \) and (c) \( \alpha \)
Table 2 Rules of the FLS

<table>
<thead>
<tr>
<th>α</th>
<th>$V_{W_a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>$V_{agg}$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 2 3 3 4</td>
</tr>
<tr>
<td>2</td>
<td>1 1 2 3 3 4 4</td>
</tr>
<tr>
<td>3</td>
<td>1 2 3 3 5 6 7</td>
</tr>
<tr>
<td>4</td>
<td>2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>5</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>6</td>
<td>1 1 2 3 4 5 6</td>
</tr>
<tr>
<td>7</td>
<td>1 1 2 3 3 4 5</td>
</tr>
</tbody>
</table>

Fig. 7 Wind speed received by the first DFIG wind turbine in each group

Fig. 8 Evaluation of the proposed aggregated wind farm model during normal operation at the PCC: (a) $P_e$ and (b) $Q_e$

Fig. 9 Evaluation of the proposed aggregated wind farm model during grid disturbance at the PCC: (a) $P_e$ and (b) $Q_e$

Table 3 Accuracy in approximation

<table>
<thead>
<tr>
<th>Operation type</th>
<th>Full aggregated model</th>
<th>Proposed model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal operation</td>
<td>$n_{P_e}$ (%) 91.3</td>
<td>$n_{P_e}$ (%) 100</td>
</tr>
<tr>
<td>Grid disturbance</td>
<td>$n_{Q_e}$ (%) 87.5</td>
<td>$n_{Q_e}$ (%) 100</td>
</tr>
</tbody>
</table>

Table 4 Comparison of simulation time

<table>
<thead>
<tr>
<th>Operation type</th>
<th>Simulation computation time (s)</th>
<th>Reduction in simulation time (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM FAM PM</td>
<td>FAM PM</td>
<td>FAM PM</td>
</tr>
<tr>
<td>Normal operation</td>
<td>1476 110 142</td>
<td>92.5 90.3</td>
</tr>
<tr>
<td>Grid disturbance</td>
<td>2283 235 298</td>
<td>89.7 87</td>
</tr>
</tbody>
</table>
IMPACT OF DFIG WIND FARM ON POWER SYSTEM STABILITY DURING TRANSIENT EVENTS

Md Ayaz Chowdhury\textsuperscript{a}, Weixiang Shen\textsuperscript{b}, Nasser Hosseinzadeh\textsuperscript{c}, Hemanshu Roy Pota\textsuperscript{d}

\textsuperscript{a}Faculty of ECS, Swinburne University of Technology Sarawak Campus, Kuching, Malaysia
\textsuperscript{b}Faculty SET, Swinburne University of Technology, Hawthorn, VIC, Australia
\textsuperscript{c}Department of ECE, Sultan Qaboos University, Muscat, Oman
\textsuperscript{d}School of EIT, University of New South Wales, Canberra, ACT, Australia

ABSTRACT

This paper carries out a quantitative impact analysis of wind farm equipped with doubly-fed induction generator on transient stability of power system. This is done by evaluating the transient energy margin through the formulation of a modified transient energy function. Simulation results show that power systems integrated with wind farms are highly sensitive to transient events with higher voltage sag, fault clearing time and wind penetration and lower load demand is below a threshold value. It is also found that fault response becomes more diverse for individual synchronous generators at different locations with larger wind energy penetration.

NOMENCLATURE

\begin{itemize}
\item $E_{cl}$ = Transient energy
\item $E_{cr}$ = Critical energy
\item $T$ = Transient energy margin
\item $E_{KE}$ = Transient kinetic energy
\item $E_{PE}$ = Transient potential energy
\item $\theta_{cr}, \theta_{sys}$ = Angle between rotor flux and resultant magnetic flux in the air gap of the critical and the remaining machines
\item $\omega_{cr}, \omega_{sys}$ = Rotor speeds of the critical and the remaining machines
\item $P_m$ = Mechanical input power
\item $P_e$ = Electrical output power
\item $J_{cr}, J_{sys}$ = Moment of inertia of the critical and the remaining machines, respectively
\item $J_{eq}$ = Equivalent moment of inertia
\item $n_{cr}, n_{sys}$ = Number of the critical and the remaining machines, respectively
\item $V$ = Internal voltage of the generator
\item $B$ = Susceptance of the generator
\item $G$ = Conductance of the generator
\item $\theta_u$ = Unstable equilibrium point
\item $\theta_c$ = Critical clearing angle
\item $\theta_{SEP}$ = Stable equilibrium point
\end{itemize}

INTRODUCTION

Higher installation capacity of wind farms and higher interconnection voltage level bring about wider influence of wind farms on the transient stability of power systems during normal operations and transient faults. Transient stability is dominantly influenced by generator technology and hence, it is imperative to study the impact of wind turbine generators on power system dynamic behaviours elaborately and systematically with high penetration of wind farms into power systems. Due to a number of advantages compared to other wind turbine generators, doubly-fed induction generator (DFIG) has become more popular worldwide in recent times. Thus, this paper focuses on investigating the transient stability of power systems integrated with DFIG wind farms.

Early research advocates that if some traditional synchronous generators (SGs) are replaced with DFIGs of the same capacity, transient stability would be enhanced (Nunes, et al., 2004; Li, et al., 2009). Further research reveals that transient stability can be either improved or reduced when some traditional SGs are replaced with DFIGs of the same capacity (Samarasinghe and Ancell, 2008). This finding is affirmed with identification of electromechanical modes of oscillation using eigenvalue analysis that influences transient stability beneficially or detrimentally with increasing DFIG integration (Gautam, et al., 2009). However, these qualitative studies are unable to point to any definite transient state of power system.

An advanced quantitative study is carried out by means of the critical clearing time (CCT) (Yanhua, et al., 2011), the transient rotor angle stability index (TRASI) (Meegahapola, et al., 2010) and the transient stability index (TSI) (Libao, et al., 2009). The study reveals that transient stability increases first and then decreases with the increasing capacity of DFIGs and may have an adverse impact in response to large faults initiated near the DFIG wind farms. The results in the above-mentioned articles are still limited and preliminary. Again, these parameters, in general, can only measure the sensitivity of transient stability, i.e., the status of the stability of the
system followed by a fault.

This paper replaces one of the SGs in the IEEE New England power system with a DFIG wind farm of the same capacity and investigates the impacts of the DFIG wind farm on transient stability with the variation of different factors, like the voltage sag, the fault clearing time, the load and the wind power penetration level (termed as wind penetration in this paper). For quantitative assessment of transient stability, the transient energy margin (TEM) is used, which is calculated through the formulation of a modified transient energy function (TEF). The TEM determines not only the status but also the degree of system stability. Thus, the TEM can provide deeper insight into the impact of the DFIG wind farms on the transient stability of the power system for improving the accuracy of the operational decision making process to maintain the system stability.

**TEM CALCULATION**

At the occurrence of a fault, the generators accelerate and the power system gains kinetic and potential energy and moves away from the equilibrium point. After fault clearing, the kinetic energy is converted into the potential energy. The system must be capable of absorbing the kinetic energy before the generators can operate in a new equilibrium point. Otherwise, the system goes unstable in case of its failure of absorbing the kinetic energy. The post-fault energy injected into the system is termed as transient energy and the maximum energy absorbed by the post-fault system is termed as critical energy. The ratio of the difference between the critical energy and the transient energy to the critical energy gives the TEM

\[ T = \frac{E_{cr} - E_{cl}}{E_{cr}} \times 100\% \]  

(1)

If \( T \) is positive, it indicates that the post-fault system is stable with the system’s capacity for further absorbing \( T \) percent of the critical energy; if \( T \) is negative, it indicates that the post-fault system is unstable and the system should be capable of absorbing an extra \( T \) percent of the critical energy for switching into a stable state. \( E_{cl} \) and the critical energy \( E_{cr} \) can be calculated through the formulation of the TEF.

**TEF FORMULATION**

According to the law of energy conservation, if the TEF is conservative (remains unchanged) during the post-fault period, i.e., transient kinetic energy \( (E_{KE}) \) and transient potential energy \( (E_{PG}) \) are equally exchanged after fault is removed (Athay, et al., 1979). The TEF conservation is violated, due to the separation of a number of critical machines (machines that are likely to lose synchronism from the rest of the system) from the remaining machines immediately after a fault (Da-Zhong, et al., 2000). The system simulation result in References (Da-Zhong, et al., 2000; Rahimi, et al., 1993) proves that the total TKE never reaches zero, even though the system transient is stable. It means not all \( E_{PG} \) participate in system’s first swing separation. It also shows that not all \( E_{PG} \) are responsible for absorbing TKEs during a first swing transient; a part of \( E_{PG} \) balances that portion of \( E_{KE} \) which does not contribute to the first swing separation (Fouad and Vittal, 2000).

The phenomenon of the separation of the critical machines from the remaining machines is taken into consideration in formulation of the TEF for accurate calculation of transient stability assessment. Critical machines are identified through their non-coherence with remaining machines in a particular group using support vector clustering (Agrawal and Thukaram, 2011).

Swing equations for the multi-machine power system are

\[ \frac{d\theta}{dt} = \omega \]  

(2)

\[ J_{eq} \frac{d\omega}{dt} = J_{eq} \sum_{i=1}^{n} (P_{mi} - P_{ei}) - J_{sys} \sum_{i=1}^{n} (P_{mi} - P_{ei}) = f_i(\theta) \]  

(3)

\[ \omega = \omega_{cr} - \omega_{sys} \]  

(4)

\[ \theta = \theta_{cr} - \theta_{sys} \]  

(5)

\[ J_{eq} = \frac{J_{cr} + J_{sys}}{J_{cr}} \]  

(6)

\[ P_e = \sum_{j=1, j \neq i}^{n} \left( V_j V_j B_{ij} \sin \theta_j + V_j G_{ij} \cos \theta_j \right) \]  

(7)

Replacing Equation (2) in Equation (3) results in

\[ J_{eq} \omega d\omega = f_i(\theta) d\theta \]  

(8)

Integrating Equation (8) with the appropriate upper and lower limits provides the final expression of the TEF for \( (n_{cr} + n_{sys}) \) generators is

\[ E(\alpha, \theta) = \frac{J_{eq} \omega^2}{2} - \frac{J_{eq}}{J_{cr}} \sum_{i=1}^{n} P_{mi}(\theta_i - \theta_{SEP_i}) + \frac{J_{eq}}{J_{cr}} \sum_{i=1}^{n_{cr}} \sum_{j=1}^{n_{cr}} V_i V_j B_{ij} (\cos \theta_j - \cos \theta_{SEP_i}) \]  

\[ - \frac{J_{eq}}{J_{cr}} \sum_{i=1}^{n_{cr}} \sum_{j=1}^{n_{cr}} \int_{\theta_{SEP_i} + 0}^{\theta_{i} + 0} V_i V_j G_{ij} \cos \theta_j d(\theta_i + \theta_j) \]  

\[ + \frac{J_{eq}}{J_{sys}} \sum_{i=1}^{n_{sys}} P_{mi}(\theta_i - \theta_{SEP_i}) + \frac{J_{eq}}{J_{sys}} \sum_{i=1}^{n_{sys}} \sum_{j=1}^{n_{sys}} V_i V_j B_{ij} (\cos \theta_j - \cos \theta_{SEP_i}) \]  

\[ - \frac{J_{eq}}{J_{sys}} \sum_{i=1}^{n_{sys}} \sum_{j=1}^{n_{sys}} \int_{\theta_{SEP_i} + 0}^{\theta_{i} + 0} V_i V_j G_{ij} \cos \theta_j d(\theta_i + \theta_j) \]
with external power electronic devices that decouple control of active and reactive power and restoration of terminal voltage. Thus, the system possesses more favorable transient response during the wind operation than the base operation. On the other hand, the DFIG has a softer and more flexible shaft system than the SG, it can accumulate a high amount of energy in the rotating mass of the DFIG wind turbine. This large amount of transient energy is released to the system followed by the fault, which is difficult to be absorbed by the system due to having limited energy absorbing capability. Thus, the wind operation becomes more vulnerable to transient instability than the base operation as voltage sag increases.

The crossover points for G2, G4 and G9 are 42%, 61.5% and 82.5%, respectively. This indicates that DFIG wind farm has less impact on SGs that are at farther distance because those SGs generate a small amount of transient energy that can be easily absorbed by the system.

The standard deviations of the TEM ($\sigma_{TEM}$) of the SGs for both base and wind operations under different voltage sags are shown in Figure 3. It is observed that $\sigma_{TEM}$ is higher among SGs at different locations during wind operation when voltage sag is above 20%. It means DFIG wind farm integration into power systems results in diverse fault response for individual SGs at different locations when voltage sag is above the mentioned threshold.

**Impact of Fault Clearing Time**

Simulation result for different fault clearing time ($t_c$) values is shown for G10, for example, in Figure 4. The fault clearing time crossover point of the TEM for base operation is about 0.048 s. It means the wind operation provides more positive impact on transient stability if the fault clearing time is less than 0.048 s. Otherwise; the wind operation is more vulnerable to power system instability as compared to the base operation. This is because the system encounters with higher modes of oscillations if the fault clearing time is longer. In such a situation, the natural damping ability of the SGs due to having higher inertial constant as compared to the DFIG is more helpful than the decoupled control ability of the DFIG wind turbine.

The crossover points for G2, G4 and G9 are 0.057 s, 0.074 s and 0.082 s, respectively.

The $\sigma_{TEM}$ curve in Figure 5 shows that DFIG wind farm integration into power systems results in diverse fault response for individual SGs at different locations when fault clearing time is above 0.03.

**Impact of Load Demand**

Simulation result for different load demand ($P_L$) values is shown for G10, for example, in Figure 6. It shows that the SGs possess better transient stability with
higher load demand at bus 39 for both operations. At lower load demand, the power system suffers from power imbalances prior to the fault and due to this fact the system cannot absorb the transient energy generated effectively before the CCT. It results in transient instability of the system.

The $\sigma_{TEM}$ curve in Figure 7 shows that DFIG wind farm integration into power systems results in diverse fault response for individual SGs at different locations when load demand is below 88%.

Impact of Wind Penetration

Simulation results in Figure 8 show that the SGs possess poorer transient stability with higher wind penetration since higher wind penetration refers a larger number of the soft and flexible shaft of the DFIG wind turbines. This causes the power system suffer from transient instability as discussed earlier.

The DFIG wind farm plays a positive role on transient stability if the wind penetration is below 15.8%. The critical wind penetration (that value of penetration when the TEM of a particular SG reaches zero) for G10, G2, G4 and G9 during the wind operation are about 21.3%, 22.5%, 25.2% and 26.2%, respectively. The reason behind increasing critical value with increasing distance of the SGs from the fault location is that the transient energy generated by those SGs is less, which can easily be absorbed by the system to sustain transient stability.

The $\sigma_{TEM}$ curve in Figure 9 demonstrates that SGs at different locations possess more diverse fault response for individual with larger DFIG wind farm penetration into power systems.

CONCLUSIONS AND RECOMMENDATIONS

The impact of DFIG wind farm on power system transient stability under a variety of conditions defined by the voltage sag, the fault clearing time, the load and the wind power penetration level has been studied quantitatively by using the transient energy margin, which is calculated through the formulation of a modified transient energy function. Following conclusions are reached through the research works in this paper:

1. Power systems integrated with the DFIG wind farms must be equipped with advanced switchgear and faster isolators to ensure its reliable operation during transient events.
2. Each individual machine should have advanced protection systems with sensitivity of a wide range of stability (i.e., transient energy margin).
3. An efficient power reserve systems or advanced reactive power compensating device should be used to overcome the power imbalances due to the variability and intermittency of wind power along with dynamically varying load demands.

REFERENCES


FIGURES AND TABLES

Fig. 1 Single line diagram of IEEE New England 10 machine 39 bus power system

Fig. 2 TEM for different voltage sags

Fig. 3 $\sigma_{TEM}$ for different voltage sags

Fig. 4 TEM for different fault clearing times

Fig. 5 $\sigma_{TEM}$ for different fault clearing times
Fig. 6 TEM for different load demands

Fig. 7 $\sigma_{TEM}$ for different load demands

Fig. 9 $\sigma_{TEM}$ for different wind penetrations

Fig. 8 TEM for different wind penetrations: (a) G10, (b) G2, (c) G4 and (d) G9
EFFECT OF FLOCCULATED AND DISPERSED MICROFABRICS ON MECHANICAL PROPERTIES OF RECONSTITUTED KAOLIN SOILS

Steven Tung-Yew Wong¹, Dominic Ek-Leong Ong¹, Arul Arulrajah²
¹Swinburne University of Technology, Sarawak Campus, Malaysia
²Swinburne University of Technology, Melbourne, Australia

ABSTRACT

Mechanical properties of kaolin clay and silt with different geometric arrangements of microfabric, namely being flocculated and dispersed, are investigated. The samples were prepared and consolidated to 100 kPa by reconstituting commercially available kaolin powder with different pore fluids. A series of K₀ consolidated undrained triaxial tests with different effective vertical stresses and axial strain rates were performed on the kaolin specimens to study its behaviour in shear and subsequently its shear strength parameters. Compressibility of kaolin was explored through tests carried out using the one dimensional consolidation test. The results show that the arrangement of microfabric will affect the mechanical properties of kaolin. For the normally consolidated dispersed samples at 300 kPa effective vertical stress, when ‘peak’ deviatoric stress occurs, the shear resistance of kaolin has shown a continuously reduction (termed strain softening). However, flocculated and dispersed microfabrics under normal consolidation showed no distinct peak strength up to axial strain of 20.0%. Such observation is useful to relate to the clayey soils most likely found at the transition regions between river and sea, where they tend to behave more like a dispersed sample studied herein due to appropriate concentration of salt in the ambient pore fluids and the presence of cations sodium. The (dispersed) clayey soils at those areas also tend to experience strain softening than those (flocculated) clayey soils along the riverbank with fresh water.

NOMENCLATURE

\( p'_c \quad = \quad \) Preconsolidation ratio
\( q \quad = \quad \) Deviatoric stress (kPa)
\( q_f \quad = \quad \) Deviatoric stress at failure (kPa)
\( w \quad = \quad \) Sample moisture content (%)
\( \varepsilon_a \quad = \quad \) Axial strain during shear deformation (%)
\( \varepsilon_{af} \quad = \quad \) Axial strain at failure (%)
\( \varepsilon_v \quad = \quad \) Volumetric strain (%)
\( \phi' \quad = \quad \) Effective angle of internal friction (°)
\( \sigma'_1 \quad = \quad \) Major principal stress (kPa)
\( \sigma'_3 \quad = \quad \) Minor principal stress (kPa)
\( \sigma'/\sigma'_3 \quad = \quad \) Effective stress ratio

INTRODUCTION

Background

Riverine infrastructures namely jetties, ferry ramps, wharves and bridges are commonly constructed in riverbanks to facilitate movement of people and goods along the many rivers in Sarawak. Many damages and failures to riverine infrastructures have been reported due to detrimental riverbank soil movements. Prime suspect is the lack of in-depth technical knowledge of complex soil-structure interaction caused by the ever-changing river geomorphology as a result of large tidal fluctuations and soft riverbank deposits. The possible sustainable solution is to increase the life cycle of riverine infrastructures by developing proper design guidelines.

Few researchers have performed the triaxial tests on the kaolin specimens with different microfabrics and analysed the relationship of microfabric of the specimen with its mechanical properties (Ajanta and Dayakar, 2007). In order to simulate a more precise riverbank soil-pile interaction finite element model (not in the scope of this paper), it is proposed that the mechanical properties of clay and silt with particle sizes of smaller than 2 \( \mu m \) and 2 \( \mu m \) - 75 \( \mu m \) respectively shall be studied first. This is because the soil particle sizes along the downstream riverbanks, especially at river mouths, tend to be finer than the soil particle sizes at upstream riverbanks as the finer soil particles are usually transported further downstream than the larger soil particles due to the slower river water flow at the river mouth.
Through understanding mechanical properties of kaolin such as shear strength and friction angle, compressibility and permeability, the measured mechanical properties will form a baseline or benchmark data for the future collected real riverbank soils of Sarawak which are dominated by clays and silts. As the tested flocculated or dispersed samples are pure kaolin, it will for now, greatly reduce the complexity of analyzing the real clayey soil and silty soil, which may be heterogeneous in nature i.e. contains a combination of gravel, sand and/or decayed organic matter.

Objective

The objective of this paper is to investigate the effect of flocculated and dispersed microfabrics on mechanical properties of kaolin. The investigated mechanical properties shall include compressibility, deviatoric stress at failure, effective angle of internal friction and stress-strain behavior.

SAMPLE PREPARATION

Specimen

Commercially available KM20 kaolin powder was used to prepare the flocculated and dispersed kaolin specimens. KM20 was mined from Perak, Malaysia at 4.5 m – 6.0 m below the ground and the chemical composition was tabulated in Table 1. The chemical composition of kaolin KM20 was predominantly aluminum silicate. The flocculated microfabric kaolin specimens were obtained by mixing the kaolin powder with distilled water as the pore fluid at a weight ratio of 1 kg of kaolin powder to 1.2 kg of distilled water. For the dispersed microfabric kaolin specimen, 67 g of Sodium Oxalate was dissolved in every litre of the distilled water. A similar procedure was adopted previously by Rakesh et al. (2011). The compositions of flocculated and dispersed kaolin soil are summarized in Table 2.

The abovementioned materials were thoroughly mixed in the kaolin mixer unit (Figure 1) for about 30 minutes while drainages were closed and a suction pump (Figure 2) was used continuously throughout the mixing process to eliminate possible air voids within the slurry. Then, the kaolin slurry (Figure 3) was consolidated using dead weight of 100 kg for about 2 days before the kaolin slurry was loaded on hydraulic press unit up to a vertical consolidation pressure of 100 kPa as per Rakesh et al. (2011). Once the consolidation stage was over, the samples were obtained using a thin wall sampling tube (Figure 4). The universal extruder (Figure 5) was used to extrude the sample from the thin wall sampling tube. After that, the samples were trimmed to specimen size of 38 mm diameter and 76 mm height (Figure 6).

Ko Consolidated Undrained Test

Ko condition was imposed by regulating the volume change of the specimen so that the volume change is equivalent to the change in the axial strain (Wanatowski and Chu, 2012). In other words, there is no lateral movement in Ko condition. Ko consolidated undrained triaxial tests (CKoU) were conducted by using Geocomp TRIAXIAL (see Figure 7). Saturation of samples was ensured by applying back pressure until the B value obtained was greater than 0.95. Then, both dispersed and flocculated kaolin specimens were subjected to different effective confining pressures of 50, 100, 200, and 300 kPa respectively. In order to equalize the excess pore water pressure in undrained condition, the dispersed samples were sheared at an axial strain rate of 0.008%/min while flocculated samples were sheared at an axial strain rate of 0.05 %/min. This is because the time of 100% primary consolidation of dispersed microfabric was found to be larger than the flocculated microfabric (Ajanta and Dayakar, 2007).

RESULTS AND DISCUSSIONS

Soil Classification

The Atterberg limits are based on the moisture content of the soil. The plastic limit is the moisture content that defines where the soil changes from a semi-solid to a plastic state. The liquid limit is the moisture content that defines where the soil changes from a liquid to a viscous fluid state. From Atterberg limit tests, the liquid limit and plastic limit of dispersed kaolin samples are 48.80 % and 35.98 % respectively which are relatively lower than the liquid limit and plastic limit of flocculated kaolin samples which are 51.75 % and 44.10 % respectively. The dispersed kaolin sample has also a lower linear shrinkage than flocculated kaolin sample as summarized in Table 3. Based on the ASTM D2487-00 (ASTM Standard D2487 2000), the dispersed kaolin was categorized as low plasticity CLAY (CL) while flocculated kaolin was categorized as high plasticity SILT (MH).

Even though the dispersed and flocculated samples are reconstituted from the same source of kaolin powder, the dispersed kaolin has shown reduction in both of the liquid limit and plastic limit. As a result, the dispersed kaolin has also shown a relatively higher plasticity index (12.82%) than the plasticity index of flocculated kaolin (7.66%). The implication is that the presence of dispersing agent or sodium cations will have effect on the classification of the soils as it will alter the plasticity index of the soils.
Scanning Electron Microscope (SEM)

SEM was used to magnify and investigate the dispersed and flocculated microfabrics of kaolin, after consolidation of vertical pressure of 100 kPa. The SEM micrographs are shown in Figure 9 and Figure 10 for dispersed and flocculated kaolin specimens, respectively. The photographs show flaky kaolinite particles. For the flocculated samples (see Figure 10), the particles contacted, stacked and adhered one another, forming clusters or flocks of a larger size, and the flocculated samples would then seem to be more permeable due to drainage paths in between the kaolinite particles. The drainage paths seem more stable with less movement and less blockage. On the other hand, dispersed samples (see Figure 9) were dispersed in the material mix and the kaolinite particles seem to be relatively more mobile and tended to clog soil pores and formed layers of low permeability (Rakesh et al., 2011).

1D consolidation test

The void ratio versus vertical effective stress relationship were observed during the 1D consolidation test on dispersed and flocculated kaolin specimens. Figure 11a shows that the preconsolidation pressures (p'), of flocculated and dispersed samples were 100 kPa which were coincident with the applied vertical pressures of 100 kPa during the consolidation of kaolin slurry. The figure also shows that the compressibility of flocculated and dispersed microfabrics was almost identical. The compression index (Cc) and the recompression index (Cr) of flocculated kaolin were 0.407 and 0.085 respectively while the Cc and Cr of dispersed kaolin were slightly higher than the flocculated kaolin samples, at 0.403 and 0.085 respectively.

The back-calculated vertical permeabilities (k v) at different applied vertical stresses of dispersed and flocculated samples from the 1D consolidation test are shown in Figure 11b. Both vertical permeabilities of the dispersed and flocculated samples range between 1.00 × 10⁻⁸ and 5.86×10⁻⁸ m/s for applied vertical stresses between 100 and 800 kPa. Moreover, the dispersed samples have relatively smaller vertical permeabilities at the same applied vertical stresses than the flocculated samples. The dispersed samples have shown lower vertical permeabilities most likely due to the arrangement of kaolinite particles being relatively more mobile and tend to clog or impede the flow of pore fluids than the kaolinite particles of the flocculated samples.

The mobility of kaolinite particles of dispersed samples is higher as the attractive forces between the kaolinite particles or Van der Waals’ forces are reduced by the cations of the dispersing agent sodium oxalate (Na₂C₂O₄). The cations of sodium oxalate (Na⁺) tend to form a shell surrounding the kaolinite particles and reduce the attractive forces between the kaolinite particles. In other words, the kaolinite particles will transform from flocculated to dispersive behavior with the right concentration of salty solutions when the cations sodium present in the solution. This is also why relatively higher liquid limit and plastic limit have been recorded from the flocculated kaolin than the dispersed kaolin due to dispersed kaolin particles are more mobile at the same moisture content. These transformation phenomena can be expected when we go closer to the river mouth where the salty sea water mixes with the river fresh water and forms a right concentration of salty solutions. The clayey soils near the transition riverbank regions from river to sea may tend to behave more like dispersed samples. However, the soils tend toward salt-type flocculation when the electrolyte concentration or the salt concentration increases and exceeds certain limit at the regions closer to the sea where the sea water has higher salt concentration.

Kₜ Consolidated Undrained Test

From the CKₜ,U tests, only the dispersed kaolin sample that was tested under 300 kPa effective vertical stress normally-consolidated, has shown strength softening after reaching the peak deviatoric stress as shown in Figure 12 and a clear Roscoe surface has also been recorded in q-p’ stress path plots as shown in Figure 13. Other than 300 kPa dispersed tests, both dispersed and flocculated kaolin samples did not show clear peak deviatoric stresses. The dispersed kaolin samples have shown slightly higher effective friction angle (ϕ') than the flocculated kaolin samples which are 35.71° and 34.91° respectively as shown in Figure 14.

The criteria for the stress condition at failure are as follows: (a) maximum deviatoric stress; (b) maximum effective stress ratio (σ'd/σ'v); (c) when shearing continues at constant pore pressure for undrained test (BS 1377-8, 1990). The axial strains at failure of flocculated and dispersed kaolin samples are tabulated in Table 4. The axial strains at failure are selected by maximum effective stress ratio for all CKₜ,U tests. Comparatively, dispersed kaolin samples are having higher axial strains at failure than the flocculated kaolin samples.

At Rest Lateral Earth Pressure Coefficient (K₀)

K₀ values or the ratio of the effective horizontal stress to the effective horizontal stress (σreh/σ₁), of the kaolin samples at different effective vertical stresses during the consolidation process have been captured by regulating or altering the effective horizontal stress by the software’s close loop control adjustment system so that the horizontal strain of the samples shall be equivalent to zero. The software will ensure that the volumetric strain (eᵥ) shall
always equal to the vertical axial strain ($e_v$). $K_o$ of both dispersed and flocculated kaolin samples at different effective vertical stresses during $K_o$ consolidation shall be lying close to 0.42 according to back calculation from Jaky’s (1948) relation of $K_o = 1 - \sin \phi'$ (see Figure 15). For the tests with 100 kPa effective vertical stress, the $K_o$ values are bigger due to the rates of applied pressure during the $K_o$ consolidation stage were fast where some lateral deformations have taken place. The appropriate steps of applied pressures shall be

$$\sigma'_i = 2 \times \sigma'_{i-1}, \text{ where } i = 1, 2, 3...n \quad (1)$$

$$\sigma'_1 = \frac{\sigma''_0}{2^n}, \text{ where } \sigma'_1 \leq 10 \text{ kPa} \quad (2)$$

For example, in order to achieve effective vertical stress of 400 kPa. From equation (2), the initial effective vertical pressure ($\sigma'_1$) shall be 6.25 kPa and followed by using the equation (1) the steps of effective vertical pressure shall be 12.5 kPa, 25 kPa, 50 kPa, 100 kPa, 200 kPa and 400 kPa. The rates of reaching target pressures at each step shall be

$$\frac{\delta \sigma'_i}{\delta t} = \frac{\sigma'_i - \sigma'_{i-1}}{\Delta t} \quad \text{ for } i = 1, 2, 3...n \text{ and } \Delta t = 8-12 \text{ hrs} \quad (3)$$

From equation (3), the rate for 100 kPa effective vertical stress shall be 0.069 ~ 0.104 kPa/min. After achieving the target pressures, the effective vertical stresses shall be maintained for at least 8 hours to ensure 90% consolidation and final consolidation pressure is allowed to remain for 24hrs.

For the consolidated undrained (CU) triaxial test, the $K_o$ is normally kept to 1.0 as the triaxial apparatus has no close loop adjustment system. The implication of the CU test is that the effective vertical and effective horizontal stresses are maintained from start till the end of the tests. In other words, the CU test may not able to produce the behavior of the samples that is closer to the in-situ at rest condition due to the $K_o$ not correctly assigned.

CONCLUSIONS

From the preceding results and discussions, different microfabrics will have effects on the mechanical properties of dispersed and flocculated kaolin samples. The dispersed kaolin samples, generally, have slightly higher effective friction angle and compression index, lower at rest lateral earth pressure coefficients and lower vertical permeability than flocculated kaolin samples.

For the normally consolidated dispersed samples at 300 kPa effective vertical stress, when ‘peak’ deviatoric stress occurs, the shear resistance of kaolin has shown a strain softening. However, other flocculated and dispersed microfabrics under normal consolidation showed no distinct peak strength up to axial strain of 20.0%.

The $C_KU$ test is able to simulate the behavior of the soil samples that is closer to the in-situ at rest condition as the no horizontal strain is allowed by the control of close loop control adjustment system. CU test may not able to produce the behavior of the samples that is closer to the in-situ at rest condition due to the $K_o$ is not correctly assigned.

The clayey soils near to the transition regions from river to sea or downstream of river may tend to behave as dispersed samples due to appropriate concentration of salt in the ambient pore fluids and the presence of cations sodium. As such, the (dispersed) clayey soils at those areas may also tend to experience strain softening in contrast to those (flocculated) clayey soils along the riverbank with fresh water.

REFERENCES


FIGURES AND TABLES

Table. 1 Specification of minerals of Kaolin KM20

<table>
<thead>
<tr>
<th>Chemical Composition (XRF Test Method)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminium Oxide (Al₂O₃)</td>
<td>17.0-27.0%</td>
</tr>
<tr>
<td>Silicon Dioxide (SiO₂)</td>
<td>60.0-73.0%</td>
</tr>
<tr>
<td>Iron Oxide (Fe₂O₃)</td>
<td>Below 2.0%</td>
</tr>
<tr>
<td>Potash (K₂O)</td>
<td>Below 4.0%</td>
</tr>
<tr>
<td>Magnesia (MgO)</td>
<td>Below 1.0%</td>
</tr>
<tr>
<td>Loss on Ignition @ 1025 °C</td>
<td>6.0-9.0%</td>
</tr>
</tbody>
</table>

Table. 2 Compositions of flocculated and dispersed kaolin soil specimen

<table>
<thead>
<tr>
<th></th>
<th>Dispersed</th>
<th>Flocculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaolin Powder KM 20</td>
<td>25 kg</td>
<td>25 kg</td>
</tr>
<tr>
<td>Distilled Water</td>
<td>30 kg</td>
<td>30 kg</td>
</tr>
<tr>
<td>Sodium Oxalate</td>
<td>2 kg</td>
<td>0 kg</td>
</tr>
</tbody>
</table>

Fig. 1  Kaolin mixer in operation

Fig. 2  Suction Pump

Fig. 3  Kaolin slurry

Fig. 4  38 mm internal diameter thin wall sampling tube with kaolin sample

Fig. 5  Universal extruder
Table 3. Summary of Atterberg limit test, linear shrinkage of flocculated and dispersed kaolin samples

<table>
<thead>
<tr>
<th>Properties</th>
<th>Dispersed kaolin (1)</th>
<th>Flocculated kaolin (2)</th>
<th>Ratio Dispersed/Flocculated = (1)/(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Moisture Content, w (%)</td>
<td>50.88</td>
<td>64.63</td>
<td>0.79</td>
</tr>
<tr>
<td>Liquid Limit, LL (%)</td>
<td>48.8</td>
<td>51.75</td>
<td>0.94</td>
</tr>
<tr>
<td>Plastic Limit, PL (%)</td>
<td>35.98</td>
<td>44.10</td>
<td>0.82</td>
</tr>
<tr>
<td>Plasticity Index, PI = LL - PL (%)</td>
<td>12.82</td>
<td>7.66</td>
<td>1.67</td>
</tr>
<tr>
<td>ASTM D2487-00 Soil Classification &amp; Plasticity Chart (ASTM Standard D2487, 2000)</td>
<td>CL</td>
<td>MH</td>
<td>-</td>
</tr>
<tr>
<td>Liquidity Index, LI = (w - PL)/(LL - PL) (%)</td>
<td>1.16</td>
<td>2.68</td>
<td>0.43</td>
</tr>
<tr>
<td>Linear Shrinkage, LS (%)</td>
<td>0.36</td>
<td>2.14</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Fig. 9 SEM micrograph of consolidated dispersed kaolin slurry specimens

Fig. 10 SEM micrograph of consolidated flocculated kaolin slurry specimens

Fig. 11a Stress-void ratio relationship from 1D consolidation test

Fig. 11b Back-calculated vertical permeability of dispersed and flocculated samples from 1D consolidation test

Fig. 12 Deviatoric stress vs axial strain plot of flocculated and dispersed kaolin samples

Table 4. Axial strains at failure of flocculated and dispersed kaolin samples

<table>
<thead>
<tr>
<th>Microfabric</th>
<th>$p'_{f}$ (kPa)</th>
<th>$q_{f}$ (kPa)</th>
<th>$q_{f}/p'_{c}$</th>
<th>$\sigma'<em>{f}/\sigma'</em>{c}$</th>
<th>$\varepsilon_{af}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flocculated</td>
<td>50 111.13 1.09 3.83 10.56</td>
<td>100 185.31 1.85 3.80 11.59</td>
<td>200 341.91 3.42 3.60 18.28</td>
<td>300 553.48 5.53 3.60 20.00</td>
<td></td>
</tr>
<tr>
<td>Dispersed</td>
<td>50 118.45 1.18 4.44 18.66</td>
<td>100 216.14 2.16 3.60 19.33</td>
<td>200 308.95 3.09 3.96 19.99</td>
<td>300 237.51 2.38 2.10 20.00</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 13 q-p' stress paths of flocculated and dispersed kaolin samples

Fig. 14 Critical state lines of flocculated and dispersed kaolin samples

Fig. 15 K_o of flocculated and dispersed kaolin at different applied effective vertical stresses
ABSTRACT

Wireless Sensor Network (WSN) consists of spatially distributed autonomous sensors that monitor and sense physical and environmental conditions, such as temperature, humidity, rain, pressure, motion and vibration, and send the captured information to the Base Station (BS). WSN used unlicensed 2.4 GHz Industrial, Scientific and Medical (ISM) radio frequency band for transmitting the signals and communication among self. The radio signals could be affected when applied in outdoor environment due to signal interference and multipath fading, which could reduce the overall wireless sensor network performance. In this era, WSN is being used in many industries such as medical, military, inventory management, structural and environment monitoring.

In this paper, we have evaluated other researchers work on WSN interference and observed the real time interference effects on WSN via implementing of “MEMSIC eKo Pro Series” wireless sensor network in an outdoor field. These experiments were performed to analyze the data traffic and WSN signals interference in the presence of different aspects such as rain, wifi and bluetooth signal interference. We have used “AirMagnet Spectrum XT” spectrum analyzer tool to analyze the RF interference on eKo nodes signal quality that operates at 2.4 GHz and use the Zigbee protocol to communicate.

INTRODUCTION

In the 21st century, the wireless communication and networking industries achieved tremendous development by providing an elastic, supportive and spacious platform to human being. Currently, the scientist and companies are looking for new technologies to increase the production quality with less human force involvement, this idea elevated the cosmic horizon to the Wireless Sensor Network (WSN) technology. In this era of development, the WSN industry has expanded immensely and exploitation of wireless sensor networks in ubiquitous computing environment is increasing vastly. WSN applications have been used in many industries and civilian areas, for various purposes such as environmental monitoring, habitat monitoring, traffic monitoring, precision agriculture monitoring, security monitoring, facility automation and traceability systems (Azman & Azwan, 2009).

WSN has become prominent due to its low cost, reliable and highly productive benefits. WSN provides maximum freedom for users to deploy a large number of sensors at remote, hazardous, steep environment places and rural areas to collect required results. It works diversely in comparison to traditional network due to flexibility, tiny size and wireless infrastructure. These devices allow sensors to monitor physical and environmental conditions, such as temperature, motion, air pressure, light intensity, soil moisture and other factors. The Wireless Sensor Network requires a strong data acquisition, distribution network and a management center to monitor and control data acquisition and data reporting operations (Azman, et al., 2009).

WSN used 2.4 GHz (ISM) industrial, scientific and medical radio bands that is an unlicensed and free spectrum. Therefore, most of the personal area network related technologies operate in this spectrum, such as Bluetooth and ZigBee, WiFi, Cordless phones, Microwave ovens, Wireless USB and others. As a result, the interference levels on this spectrum are very high and some studies have shown that this spectrum is also susceptible to rain due to absorption of parts of this spectrum and due to moisture. Radio frequency waves have also been seen to maneuver really well in an area covered with trees (Natalie Anak, et al., 2014).

LITERATURE REVIEW

In the past few years, Wireless Sensor network technology has become very prominent due to its usability. The advancement of WSN has led to the development of many new protocols and hardware architecture that addresses the potential collaboration of sensor in data gathering, processing, coordinating and managing in order maximize benefits of wireless sensor network technology. Due to the advancement of the
technology, industries are able to manufacture small and low cost sensor economically. These sensors check different conditions and transmit the information via a radio signal. These radio signals contain information about node location and transpire event. A large number of mobile and static sensors can be deployed in a large geographical region for information gathering that require unattended operation.

The main objective of WSN is data sensing and processing them over a communication medium. Therefore, every sensor node consist of sensing unit, power unit, processing unit, memory unit and communication unit that have the capabilities of sensing, processing and data transmitting. These sensors sense the physical and environmental changes, convert this information into radio frequency and transmit over a radio medium. Initially analogue data is converted into digital data and then forwarded to microcontroller for further processing. The program code and data packets are stored in the memory unit. The data packets are transmitted between nodes and stored in Random Only Memory (ROM). The Flash Memory is used to store program code. The power unit provides an energy source in the form of battery for transmission and operations computation. The microcontroller performs processing task, and responsible for processing information, converting data, data acquisition, routing and transmitting the information to Base station or nodes. One WSN can have one or multiple base stations, which acts as stationary or dynamical. Even though sensor nodes are identical devices, but have diverse characteristics in the network structures. In sensor network the sensor deployments, coverage, transmission power, computation, reporting, addressing and communication pattern purely depends on the routing protocol operation and support (Ullah & Waqas Ahmad, 2009).

Layer Model

The WSN layer model is similar to OSI model; however it only has five layers such as application layer, transport layer, network layer, data link layer and physical layer, the model also contains three additional cross layers that are responsible for power, mobility and task management functions (Ahmad, et al., 2012).

The WSN uses to collect the information via coordination among sensors, process them and transmit over wireless medium within limited energy resources and environmental constraints. To perform the following tasks, wireless sensor network requires a task prioritization method to optimize the processing and transmission capabilities. Therefore, the layered model offers a classification of the protocol’s functionality at different layers to optimize and satisfy the network performance. Each layer is responsible for different services (Charfi, et al., 2009).

The last layer is known as physical layer which perform signal processing services. It is further divided into three sub-layers that perform individual task to complete respective responsibilities. The physical layer is mainly responsible to ensure the transmission over the wireless medium, with a choice of appropriate frequency, carrier frequency generation, signal detection, power amplification, modulation, channel switching, link quality, estimation, energy detection measurement and clear channel functionalities. It used IEEE 802.15.4 / Zigbee protocol (low data rate personal area) to perform the stated services. IEEE 802.15.4 / Zigbee is known as the global standard for WSN that offers a reliable, low cost, low power and low data rate services. IEEE 802.15.4 used two kinds of network topologies such as star topologies and peer-to-peer topology. These sensor devices can be classified as Full Function Devices (FFDs) and Reduced Function Devices (RFDs) (Charfi, et al., 2009).

IEEE 802.15.4 / Zigbee standard operates in the ISM band that support three frequency bands (868 Mhz, 915 Mhz, and 2.4 GHz) with 27 radio channels. Channel 0 can be operated in the frequency band 868.0 MHz – 868.6 MHz and supports only 20 kbps data rate. Whereas, Channel 1 to 10 can be operated in the frequency band 902.0 MHz – 928 MHz and its offers 40 kbps data rate. Channels 0 to 10 used Binary Phase Shift Keying (BPSK) modulation scheme. The highest data rate of 250kbps can be achieved at 2.4 GHz – 2.4835 GHz frequency band for Channel 11 – 26, it used the Offset Quadrature Phase Shift Keying (OQPSK) modulation scheme. IEEE 802.15.4 standard used the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol to reduce the interference in signal degradation at the receiver end. Similar to this Zigbee used Direct Sequence Spread Spectrum (DSSS) as a spreading technique that helps in increasing the signal frequency in order to increase the power and reduce the noise interference from nearby network. IEEE 802.14.5 required at least -3 dBm transmitting power for 10 – 75 meters transmission radius. The Zigbee standard specifies a receiver sensitivity of -92 dBm in the 868/915 MHz bands and -85 dBm in the 2.4 GHz band (Pan, et al., 2006).

Modes of Operation

Wireless networks have distinct modes of operation such as the ad-hoc networks and the infrastructure networks. The Infrastructure wireless networks usually have kind of base station, which acts as a central node to connect with wireless terminals. The base station has a fixed location and usually provides Internet access to an intranet or other wireless networks, therefore; it causes the problem of central point of failure, meaning if central point stops working none of the wireless terminals can communicate with each other. Whereas ad-hoc networks are formed without helping of a base station, in ad-hoc networks the wireless terminals may communicate...
directly with each other, while in terminal infrastructure networks have to use the base station to relay with their messages. These different standards have different capabilities when it comes to modes of operation (Suhail et al., 2014).

Medium Access

In order to conduct a two-way wireless communication, a detailed protocol on how the wireless terminals should access medium needs to be established. The 802.11 standards specifies a method called Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) to solve above issues. The CSMA/CA technique allows wireless transmissions to be completed without interference from other terminals. If the communication was one-way only the communication channels could be better utilized because the source terminal would not have to take into consideration transmissions being sent from other terminals. In order to deal with the hidden terminal problem 802.11 can also use a request to send; clear to send (RTS/CTS) scheme in order to reserve the channel for communication (Suhail et al., 2014).

ZIGBEE ARCHITECTURE

802.15.4 (Zigbee) standard was created by the Institute of Electrical and Electronics Engineers (IEEE) organization and mainly used to specify the two layers such as the media access control (MAC) and the physical (PHY), of a wireless personal area network (WPAN). It is famous in lower cost, lower range, low power consumption and lower data rate networks. Zigbee also used 2.4 GHz band for actuators and sensors device communication, similar to WLAN ad Bluetooth technologies. (Gislason 2008).

The Zigbee architecture consists of several layers such as Physical, Media Access Control, Network, Security and Application. However, Zigbee Alliance is only accountable for the Network, Security and Application Layers. The Physical and Media Access Control layers are not the responsibility of Zigbee Alliance and are based on the IEEE 802.15.4 standards (Rahmani 2005). The Figure 5 shows the architecture of Zigbee.

The physical layer is in-charge of translation of packets into bit streams and vice versa. It uses different modulation schemes for this process depending on the different ISM bands. For 868MHz and 915MHz, it uses DSSS to spread the bits into bit streams of 15 bits. It then uses Binary Phase Shift Keying (BPSK) to modulate the bit streams. As for the 2.4GHz band, an advanced version of DSSS called the 16-ary orthogonal modulation is used to spread the bits to bit streams of 32 bits. These bit streams will then be modulated using either the OQPSK or MSK modulation technique.

The media access control layer is used to control the flow of traffic in the network. CSMA-CA technique is used to ensure that the data sent is uncorrupted, while the acknowledgement and retransmission technique is to ensure that the receiver have received the data. It also provides the security for the overall network by using AES-128 encryption as mentioned previously.

The network layer is in control of anything that is involved in forming the network. This includes configuration of network and discovery of devices (Le 2005). Zigbee networks are able to support 3 types of network topology, namely, star, mesh and cluster tree that can be seen in Fig 1.

There can be only one coordinator node in each network. The coordinator’s role is to act as a router to the other networks. All the information about the network is stored in the coordinator and serves as the root of a network tree. Full Function Devices (FFD) function as intermediary routers that are used for transmission of data from other devices. It uses less memory when compared to the coordinator node. These devices can also act as a coordinator and is able to operate in all the topologies. As for the Reduced Function Devices (RFD), they require the least memory due to their function in the network. All of these devices do talk in the network, particularly to the coordinator only and does not relay any sort of data from other devices. The use of RFD can be found in a star topology. Besides the being accountable for the formation of the Zigbee network, the network layer also consists of some commands that are used to ensure the security of the network. These commands are usually used for devices to join or rejoin the network securely. Since the overall payload of the network frame encrypted, it can be said that the whole network layer in a Zigbee network is secured (Gislason 2008).

All running applications in the network are the responsibility of the application layer. It acts as a filter to remove duplicated messages that are sent by the network layer (Gislason 2008). It also obtains a local binding table which is used to keep a record of whichever group or nodes in the network that the node wants to communicate with (Natalie Anak, et al., 2014).

INTERFERENCE REVIEWS

WSN consists of many tiny sensor nodes that consistently communicate and transmit data via radio signal. Some of the WSN topologies use multiple path communication patterns in order to achieve higher throughput and reduce the end-to-end latency. However, under some condition it reduces the overall network performance due to the interference between multipath paths, rain, collision, which raise a big problem in heavy traffic networks such as surveillance monitoring that use to deal with live data of video traffic. Various solutions have been proposed in the market promising the best approach. A Novel Interference-Aware Multipath routing
for Video Delivery (IAMVD) was one of approach that is proposed by (Nikseresht, Ilnaz, et al., 2012) for wireless multimedia sensor networks. It established multiple paths, while considering the different QoS parameter to transmit multi-priority packets that helps in improving the network throughput. They used (frame delivery ratio, throughput, energy consumption, frame delivery latency, and frame loss ratio) metrics to evaluate the developed algorithm stability. This method helps in finding two nodes-disjoint interference minimized path for a single pair of source and destination.

IEEE 802.15.4 (Zigbee) standard has gained more attention since last decade due to its ultra-low cost and ultra-low power consumption characteristics. It used 2.4 GHz ISM unlicensed band for communication purposes. Whereas, Radio Frequency Identification (RFID) is another low cost wireless non-contact technology, which use a radio-frequency electromagnetic fields to transfer the data, in order to automatically identify and track tags objects. RFID can operate in various frequency bands from KHz to GHz. Both of these technologies are designed for different purposes, so when RFID is operated in the ISM band, it causes significant interference to other ISM band devices, which could face the degradation in their performance such as WSN. In the following paper (Chen, Junjie, et. al., 2009) the authors analyzed Bit Error Rate (BER) and Packet Error Rate (PER) of Zigbee (IEEE 802.15.4) under the interference of RFID.

(Shuaib et al. 2007) also performed some simulation based experiments to analyze the Zigbee performance under interference. They evaluate Zigbee performance with the existence of Bluetooth and WLAN in Matlab / Simulink software. Zigbee standard operates at three bands, the 2.4 GHz band with a maximum rate of 250kbps, the 915 MHz band with a data rate of 40 kbps, and the 868 MHz band with a data rate of 20 kbps. The simulations were performed based on different modulation schemes and frequency bands. The simulation results show how BER versus the Signal to Noise Ratio (SNR) effected. It also shows how interference effects of Zigbee devices on the throughput performance of IEEE 802.11g and vice versa. Zigbee interference most disturbing uplink performance of 802.11g rather than downlink (Natalie Anak, et al., 2014).

Interference Effects on Zigbee

Most of the new technologies are running on 2.4GHz ISM band due to its license-free features, which cause interference between signals in this band. Interference of rain to the signals on the wireless communication is known as rain fade. Rain fade occurs when the wavelengths of the signal are almost similar to the separation of the rain droplets (Rouse 2005). During rain fading, the signal will face a tremendous drop in signal strength. Since rain is a part of nature, the downpour of rain is not constant. Therefore, rain fade will not last long as it is dependent on the downpour of rain.

The Wireless Local Area Network (WLAN) is based on the IEEE 802.11 standard and use the DSSS spreading sequence that also operates in the 2.4GHz ISM band. It occupies 14 channels from channel 1 – 14. Each of these 14 channels has a bandwidth of 22MHz and a 5MHz bandwidth separation in between each channel (Atmel 2013). Since Zigbee and Wifi are in the 2.4GHz band, interference between co-channels will occur. Zigbee adopts the DSSS technique which is not a type of frequency hopping system. Therefore, out of 79 times, only one time a channel will overlap.

Similar to Zigbee, wireless USB uses DSSS. It has the same number of channels as Bluetooth – 79 channels, with each channel having a bandwidth of 1MHz. Both of these technologies support frequency agility and the use of CSMA-CA and transmission duty cycle, which is small helps to maintain good transmission of data with low data rate loss. These technologies are able to coexist well with one another (Atmel 2013).

Cordless telephones functions in the 2.4GHz band. However, it does not adopt any standards like other technologies. Some of the cordless telephones use DSSS while most of them use FHSS. By using DSSS, users can change its channel manually. As for FHSS, the channel cannot be changed manually as the frequencies of the telephone keeps changing. Cordless telephones based on FHSS causes a great interference to the Zigbee personal area network (PAN). It is able to block out the whole Zigbee communication system and it transmits a large amount of power that can affect the transmission of data. To avoid interference from occurring, Zigbee channels must be placed very far apart from the cordless telephones based on FHSS. As for DSSS cordless telephones, Zigbee channels must not overlap the channels of these

can be considered negligible (Atmel, 2013).

Atmospheric attenuation occurs when transmitted signal encounters with air molecules and other suspended particles in the air that cause of the significant power drop in signal. The wind itself does not effect on wireless signal, but when it’s put force on antenna and moves the antenna from line of sight the signal can be lost. In wireless communication all signals suffer with minimum Path Loss while travelling from transmitter to receiver but it can be controlled by an efficient antenna design. However, moisture, such as fog, rain and snow also shows significant attenuation in the signal’s path. The attenuation due to rain is depending on the amount of rainfall and size of raindrops, the heavier raindrop will have higher velocity of rainfall that will cause higher attenuation. The
rain causes the scattering of wireless signal, whereas water particles and carbon dioxide cause the absorption in wireless signals (Alma, et al., 2008).

The Atmospheric attenuation can be express as (Olsen, et al., 1978)
\[ \alpha = e^{-\sigma l} \]

Where \( l \) is measurement distance and \( \sigma \) is the specific attenuation coefficient per unit of length. We can calculate the value of \( \sigma \) by using Kruse and Kim relations equation (M. Al Nabousi, et al. 2003).
\[ \sigma = 3.912 / V (\lambda/550)^4 \]

In the above equation, \( V \) is visibility (km), \( \lambda \) is wavelength (nm) and \( q \) is the size distribution of diffusing particles. Different values of \( q \) are given in (M. Al Nabousi, et al. 2003).

Countries like Malaysia (rain rate = 120mm/hr) where the rain appears throughout the year, we can calculate an empirical relationship between attenuation \( A \) and rainfall rate \( R \) by using following equation (Markham, et al., 2010).
\[ A = \alpha R^\beta \]

In above equation \( \alpha \) and \( \beta \) are depended parameter on frequency and temperature. The path loss at 2.5GHz under extreme rain conditions of 250mm/hr, using worst case thunderstorm distribution \( I \) is calculated as 0.124 dB/km \( (\alpha = 1.63 \times 10^{-4}, \beta = 1.2) \). The following paper (Markham, et al., 2010) shows that rain really effect on wireless link quality, but with the very minimum path loss of 0.01 dB for 100m distance.

**IMPLEMENTATION**

In our research we have used "MEMSIC eKo Pro Series" wireless sensor system, they are mainly made for agricultural and crop outdoor monitoring activities. We have implemented 3-4 eKo node attached with different sensors in order to collect the data and to monitor interference on wireless sensor communication medium. Each eKo sensor node was placed with 60 - 100 meter distance.

In order to monitor and analyze the interference on eKo node communication network, we have used "AirMagnet Spectrum XT" (Fluke Networks.). It is spectrum analyzer for 802.11 networks that proactively identifies and finds sources of RF interference.

**eKo Technology**

eKo nodes used a new generation of integrated wireless sensor technology, which help users easily and efficiently deploy a wireless monitoring system in multiple locations to gather the required data. Each eKo node can support up to four sensors such as soil moisture, ambient temperature and humidity, leaf wetness, soil water content, and soil water. These nodes come with pre-programmed installed and MEMSIC’s XMesh low-power networking protocol configuration, which provides plug-and-play network scalability. Once a user plugged in sensors to the eKo node, it will auto scan and identify the connected sensors on the ports (MEMSIC Inc.).

**eKo Node**

The eKo Nodes come with a fully integrated, rugged outdoor sensor package. It has an energy-efficient radio and sensors that is used for extended battery-life and performance. These nodes are integrated with MEMSIC’s IRIS radio board processor and rechargeable batteries and a solar cell and have the ability to communicate over 2 miles of outdoor range (MEMSIC Inc.).

**eKo Base Radio – eB2110**

The eKo Base Radio (eB2110) is used to communicate with eKo nodes and act as a Gateway, which also come with fully integrated package. The base radio also integrated with a MEMSIC IRIS processor/radio board, antenna and USB interface board, which is preprogrammed with MEMSIC’s XMesh low-power networking protocol for communication with eKo Nodes. The USB interface is used for data transfer between the base radio and the eKoView application running inside the eKo Gateway (MEMSIC Inc.).

**eKo Sensors**

We have used three sensors such as ES1100, ES1201, and ES2000 (weather station suite).

The “ES1100” sensor is a soil moisture and soil temperature sensor. It basically used to monitor the sensor measurements between irrigations.

The “ES1201” sensor is a temperature / humidity sensor, which is used to measure the ambient relative humidity and air temperature.

The ES2000 is a weather station sensor suite that can directly plug into the eKo node. It has integrated solar powered and use to measure rain, a temperature and humidity sensor with radiation shield, solar radiation sensor, barometric pressure sensor, and an anemometer in one package. Rainfall and precipitation data are transmitted with every 15 minutes (MEMSIC Inc.).

**eKoView Web Interface**

eKoView offers a familiar and intuitive web browser based (i.e. Internet Explorer, Firefox) interface for sensor network data visualization. The eKoView web application makes easy for users to start monitoring and access their data from anywhere in the world via a laptop or smart phone. eKoView comes pre-installed on the eKo Gateway, a plug-and-play web server (MEMSIC Inc.).

**Wireless Mesh Technology**

eKo wireless mesh network is based on MEMSIC’s proprietary XMesh technology, which is a self-healing, self-configuring network that allows nodes to extend their
radio range by hopping messages. XMesh is a full featured multi-hop, ad-hoc, mesh networking protocol developed by MEMSIC for wireless sensor networks. All eKo Nodes can originate sensor data and also forward data from other eKo Nodes. Each node runs MEMSIC’s XMesh low-power mesh networking protocol. eKo Nodes can act as a repeater if placed without sensors. Each eKo Node monitors the radio traffic in its neighbourhood and keeps track of possible alternate radio paths; if one path is blocked or degraded it will switch to an alternate path. The hopping effectively extends radio communication range and reduces the power required to transmit messages. MEMSIC recommends a maximum network size of 35 eKo Nodes per gateway. Likewise, if there is a bad radio link between two nodes, that problem can be overcome by rerouting around the area of bad service. Normally the nodes run in a low-power mode, spending most of their time in a sleep state, in order to achieve multi-year battery-life (MEMSIC Inc.).

RESULTS AND DISCUSSIONS

In the first scenario, the eKo wireless sensor network was deployed in a rural area paddy field that does not have any interference of 2.4 GHz devices nearby. Our analysis shows that there was not any problem of interference due to clean environment surrounding.

In the second scenario, we have deployed the WSN in a centre of a city that has many WiFi networks interferences nearby.

Figure 8 show the scan result of available Channel and Wireless Access Points (AP) operating on 2.4 GHz band in the deployed area. The results show the available channel 1 to 13 on 2.4 GHz frequency, which are being used by different AP in the relevant area. We can see that 31 Ap’s are operating on Channel 11, Channel 1 shows 21 AP and Channel 6 show 10 AP operates on them.

Figure 9 show the scan result of SSID operating at 2.4 GHz frequency.

Figure 10 show the Real Time Fast Fourier Transform (FFT) graph that provides a real-time view into the RF energy in the environment with current, max, max-hold and average RF signal levels. We can see overlay the channel duty-cycle of the Real-time FFT graph to streamline their efforts in detecting and focusing on RF interference sources that have the maximum impact on the performance of the network.

The AirMagnet Spectrum XT is used to detect and identifies other non-WLAN sources that interfere and lower the performance of Wireless devices. The Figure 11 shows the interference of non-WiFi devices such as Cordless Phone, Bluetooth and Microwave Oven operating at 2.4 GHz frequency.

Figure 12 show the Interference Power graph displays the average power readings of interfering devices on the different channels. We can see the interference power and overlapping of WiFi, Bluetooth, Cordless Phone and Microwave signal for each channel. In the Figure 9, we can see that 31 AP’s are running on Channel 11 and Cordless Phone does create interference to WiFi Channel, which can be observed in Figure 12.

Eko Pro Series Analysis

If we look into the Figures 8, 10, 11 and 12 captured by AirMagnet Spectrum XT, we are unable to see eKo pro series operating channel that Eko nodes are used for communication, whereas it also operates on 2.4 GHz. It is because eKo base radio and eKo nodes are factory configured to ZigBee channel 25, which doesn’t overlap with Wi-Fi band and thus minimizes the chances of interference. Furthermore, these eKo nodes are integrated with IRIS family processor/radio module that a direct sequence spread spectrum radio (DSSS) supporting the 2.4 GHz global ISM band and compliant with IEEE 802.15.4 standard.

We can also observe in the Figure 13 that all 4 nodes are working perfectly under many WiFi AP existent. The Figure 11 also show the route results of Node 4, where it is connected to BS via node 5 and node 3.

Figure 15 show the continue data record of Ambient Temperature via node 2 and Wind via node 3. In the first hour that data is reported every 31.25 seconds. After the first hour, the data reporting interval changes to every 15 min and 37.5 seconds. In the current eKo Pro Series, the user cannot change the data update rate.

The Figures 14 and 16 show the wind speed and rain measurement respectively at deploying area. We can observe that under heavier wind and rain measurement values, WSN signal quality does not effected and neither it reduces the network throughput. It is because eko Nodes use internal antenna and really prepared for rugged outdoor environment.

In our finding, we have observed that the current eKo Pro Series does not support to detect any failure or spoiled sensor node. If the Base Station server (eK2110) stopped working due to electricity failure or other reason, the sensor node will also stop until user fixes the Base Station manually. So once BS is down no data will be recorded. The eKo Pro Series does not have alternative system or a way to record down the data reading from Eko sensors. The routing protocols are unable to detect physical hardware failure problem over networks which cause loss of data and degrade the network performance.

We can see in Figure 17 that BS was down due to electricity problem and there were no data recording for one week until we have fixed it back after one week. The red circle in Figure 17 showing the communication collapse of one week.

Currently eKo Pro Series only allows manually uploading of graphical location maps and setting the sensor node coordinates for localization. It is because to avoid heavy battery consumption; which does not provide
accurate and reliable node locations. So later on if any node location has been changed, the coordinates need to reset again, this does not offer the accuracy and required a lot of time.

Eko Pro Series used a mesh topology and each node maintains two paths to the Base Station for reliable communication and more suitable handle 35 nodes. This can raise a problem in the wider network. To decide the best path is not trivial when dealing with mesh networks. Usually, the best path decision will rely on a distributed and adaptive logic, where every node in the network is responsible for maintaining its routing table updated with information coming from two sources: its own database (containing information of the network topology - static), and the information resulting from the protocol execution.

CONCLUSION

In this paper we have focused on to understand the interference problem in 2.4 GHz ISM frequency band. The devices operating on 2.4 GHz are use frequency-hopping spread spectrum (FHSS) and direct-sequence spread spectrum (DSSS) modulation techniques. In the Wireless Sensor Network, ZigBee that also operated on 2.4 GHz ISM band that has sixteen channels; each channel occupies 3 MHz and channels are centred 5 MHz from each other, giving a 2-MHz gap between pairs of channels.

We have seen some papers on interference in WSN, most of them are using simulation studies to analyse the interference effects on WSN network. Whereas in our paper, we have implemented a “MEMSIC eKo Pro Series” base WSN into the field and analyses the interference in 2.4 GHz channels from other Wifi networks and rain. We understood that due to a lower frequency band 2.4 GHz, rain attenuation does not affect much of WSN performance, whereas other Wifi network cause interference due to running on the same frequency.

The eKo series WSN design has good capabilities to mitigate the interference effects and provide reliable platform except the sensor node should have some internal RAM capabilities, which could enable them to keep the data record in case if BS has a problem and cannot communicate.

In our future work, we are looking to introduce a storage method into sensor node to store the data as a backup, if Base Station collapse. So that user can retrieve the previous data reading once the Base Station power is back.

REFERENCES


Atmel 2013, 'Atmel AT02845: Coexistence between ZigBee and Other 2.4GHz Products', September 2013, pp. 1-12.


Olsen, R. O. G. E. R. S., David V. Rogers, and Daniel B. Hodge. 2978, "The aR b relation in the calculation of

Pan, Meng-Shiuan, and Yu-Chee Tseng, 2006, "ZigBee Wireless Sensor Networks and Their Applications," National Chiao Tung University, Hsin-Chu, Taiwan.

Rahmani, E 2005, “ZigBee/IEEE 802.15.4”, ECE Department, University of Tehran.


Shuaib, K, Boulmalf, M, Sallabi, F & Lakas, A 2006, Co-existence of ZigBee and WLAN, A Performance Study, thesis, College of Information Technology, UAE University, Al-Ain, UAE.


Tjensvold, Jan Magne., 2007, "Comparision of the IEEE 802.11, 802.15. 1, 802.15. 4 and 802.15. 6 wireless standards.” IEEE: September. Vol. 18.


FIGURES AND TABLES

Fig. 1 WSN Layer Model (Ahmad, et al., 2012)

Fig. 2 Layer Model for WSNs (Charfi, et al., 2009)

Fig. 3 Sub-layers integrated at the physical layer for WSN (Charfi, et al., 2009)

Fig. 4 The IEEE 802.15.4 Channel Structure (Callaway, et al. 2002)

Fig 5: Zigbee Architecture (Le 2005)

Fig 6: Zigbee Network Topologies (Le, 2005)
Table 1: Modes of Operation for different Wireless Standards (Tjensvold, 2007)

<table>
<thead>
<tr>
<th>Standard</th>
<th>Ad hoc</th>
<th>Infrastructured</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.11a/b/g/n</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>802.15.1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>802.15.4</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>802.15.6</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Table 2. eKo Pro Series Configuration (MEMSIC, Inc.)

<table>
<thead>
<tr>
<th>eKo Node</th>
<th>eN2100</th>
<th>eKo Node</th>
<th>eN2100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor Types</td>
<td>Each port supports either an eKo compatible simple or smart sensor (MEMSIC ESB protocol).</td>
<td>Outdoor Range Per Single Radio Hop</td>
<td>Typical 2000ft to 2 miles line-of-sight per hop.</td>
</tr>
<tr>
<td>Sensor Measurement Interval</td>
<td>One measurement every 15 minutes (default).</td>
<td>Outdoor Coverage</td>
<td>Flat with no overhead canopy: One eN2120 per 100-150 acres. Hilly but no overhead canopy: One eN2120 per 20-30 acres. Overhead canopy such as forest, orchards: 1 eN2120 per 4-5 acres.</td>
</tr>
<tr>
<td>Frequency</td>
<td>2.405 to 2.480 GHz</td>
<td>Antenna</td>
<td>Dipole, internal</td>
</tr>
<tr>
<td>Channels</td>
<td>16 channels available</td>
<td>Operating Current</td>
<td>0.5 mA average (no sensors) at 15 minute data sampling rate</td>
</tr>
<tr>
<td>Type</td>
<td>DSSS, IEEE 802.15.4</td>
<td>Solar Panel</td>
<td>Self-contained 1.3” x 2.5” solar panel to recharge batteries</td>
</tr>
<tr>
<td>Batteries</td>
<td>Standard: 3 AA low-leakage NiMH rechargeable (via internal solar panel).</td>
<td>Life Expectancy: 3 months with no solar recharging; &gt; 5 years field life</td>
<td></td>
</tr>
</tbody>
</table>
VARIABILITY MODELING IN COMPONENT ORIENTED ENGINEERING

Muhammed Çağrı Kaya¹, Selma Suloglu², Ali H. Dogru¹
¹Department of Computer Engineering
Middle East Technical University
Ankara, TURKEY
²Sosoft Information Technologies
Ankara, TURKEY

ABSTRACT
A new meta-model is proposed that establishes a variability specification and system configuration environment for Component Oriented System Engineering Modeling Language (COSEML). Variability is integrated to COSEML that can be viewed as an Architectural Description Language emphasizing the decomposition view. The textual version of this language is also presented and demonstrated with an example specification.

INTRODUCTION
Components are units of implemented software building blocks which are then composed to yield complex functionalities. Component Oriented Software Engineering (COSE) suggests development by integration rather than code writing. In a purely component oriented approach where most functions are assumed to have been implemented in components. This approach starts with a structural decomposition to match existing components, representing the system both in logical and physical levels hierarchically.

COSE Modeling Language (COSEML) is implemented as a graphical modeling tool (Dogru, et al., 2003). The tool can transform the graphical representation into a textual representation obeying the grammar rules of the language. Model building activity as a top-down approach is used to introduce the building blocks of the system. In lower levels, when the module is expected to correspond to a component, a temporary bottom-up approach can be taken and required subcomponents can be integrated into a super-component to achieve a certain goal.

Fulfilling various needs in time with little effort requires a systematic approach in which applications can be composed through components taking commonality and variability into account. In such variability centered system development, an explicit representation of variability eases specifying, tracking, understanding, and managing activities. For this purpose, there have been different models and approaches introduced to cope with component, interface, connector, and composition variability to some degree.

Variability in component oriented systems can be specified as a separate model which then mapped with components, their properties and relations. However, although there are attempts to model variability in COSE, there needs to be a systematic approach that enables variability specification in component, connector and composition explicitly. Owing to the hierarchical structure of component orientation, specification of variability should be handled in the same hierarchical manner.

The objective of this paper is to introduce a new metamodel which incorporates variability with component oriented development in a hierarchical way. The authors inspired from their previous work (Suloglu, et al., 2013) within the scope of variability specification.

A META-MODEL FOR VARIABILITY MANAGEMENT IN COSEML

To facilitate modeling and integration of variable component oriented systems, we introduce a metamodel which incorporates COSEML and variability specifications. The metamodel basically enables to define system decomposition in a top-down manner along with package and component variability. Metamodel overview is given in Figure 1 which depicts package and component main blocks.

Variability is hierarchically specified in a way that only upper levels manage and determine the configuration rules of subsequent below levels. While a package or component strictly deals with variability of its member packages or components at the lower level, its own variability is managed by its upper level. By this way variability gains a decentralized fashion which reduces complexity. Variability definition also goes from abstract to concrete spanning over sub-packages, components and subcomponents.

Relying on physical and logical level decomposition, all components introduce their provided and required methods and events in their interfaces without variation.
Variability of a package intervenes and binds variability of its member packages (sub-packages) and components. Besides, package variability can lead to configuration of its member component interface by activating or deactivating of methods and events and setting or unsetting of parameters. By this way, different packages can have different component interfaces of the same component which in turn provides component and interface reusability.

Package and components define their compositions in which the way how member packages and components interact with each other is introduced. To pinpoint changeable parts of the composition, inline references of its variability are included which provides a way to model a set of possible required behavior to satisfy different composition needs.

The metamodel given in Figure 2 is separated into three main parts for understandability purposes. The leftmost part includes both structural and dynamic views of COSEML. Structural view covers the building blocks of component oriented systems, packages for logical level, components and their interfaces for physical level. Dynamic view comprises composition of package and components which consists of a set of messages and interactions. Variability constructs and relation with COSEML represented at the mid and rightmost parts are adapted from (Suloglu, et al., 2013). Packages and components define their internal and external variation points along with numerical and logical constraints. Configuration variation point enables to define an abstraction in which details of low level variability bindings are concealed. To specify variability in a hierarchical way, variability mappings from upper levels to lower levels are defined.

In summary, the metamodel supports variability in package and component, in component interface and composition. In COSEML specification, connector is modelled as a set of messages which defines a connection between ports of method in and method out. Therefore, connector variability means message variability in our case. However, even though our current metamodel has an underlying structure to support this type of variability, no mechanism is provided for now.

**XCOSEML LANGUAGE**

A new domain specific language, XCOSEML, has been developed based on the metamodel that is described in the previous section. Xtext (Xtext, 2012) is used to implement the language which provides a development environment for domain specific languages to developers with Eclipse IDE integration. Five different XCOSEML models are created:

(i) package,
(ii) component,
(iii) component interface,
(iv) configuration interface, and
(v) composition specification.

Package and component models include references to their interfaces, possible configuration interfaces and composition specifications. A package can be realized by one component if it does not include any other member package or component. A complete specification would cover all parts of components for implementation, through directly or indirectly (through included packages) by components. Component interface model comprises properties, provided and required methods, and events. Configuration interface model includes variability specifications. Composition specification indicates variability mappings and descriptions about how member packages and components interact with each other through inline variability attachments. XCOSEML models are exemplified partially based on a real life case study, a travel itinerary system.

**Travel Itinerary Example**

The Travel Itinerary System allows online bookings for travelers. They can organize trip plans even for complex trips with multiple stops. Changes are allowed. The system provides five different options to the users for booking: hotel and flight, cruise, car rental, activities and vacation packages. Vacation packages contain at least one of the hotel and flight bundle. Also, some optional activities can be added to the package. Traveler can choose any booking type (flight and/or hotel), arrival and departure dates, destination place, and other details, e.g. traveling with pets. Travel agency checks convenient hotel and flight options and gathers information from available hotels and airlines. Then, related information for the selected booking type is presented to the traveler. The system offers additional activities, cruise options and advantageous vacation packages, also. Traveler can choose one among them and can book a hotel and/or flight. When the traveler finishes bookings and confirms them, the travel agency sends a message to the traveler that contains the trip plan and all other required information. In case of any booking problem, the travel agency sends a booking cancellation message to the traveler for rearrangement.

Structural decomposition of travel itinerary system in COSEML graphical representation is given in Figure 3. Travel itinerary package is decomposed into four sub-packages, namely traveler, airline manager, hotel manager and additional facilities. Further additional facilities package splits into three; car rental, cruise and activities. Travel Itinerary system is modelled in XCOSEML based on the decomposition depicted in Figure 3. Travel Itinerary package is represented by a composition and a configuration interface in Table 1. Note that for logical levels, packages have no declaration of interface specifications.
The configuration interface given in Table 2 specifies variability of the travel itinerary package with two internal (lines 2-12) and one configuration variation point (lines 13-27). “booking” internal variation point (line 2) has two optional variants: “hotel” (line 4) and “airline” (line 5) which bind at development time. Configuration variation point “itinerary” has alternative variants: “regular” and “vacation package” respectively. For realization of “regular” variant (line 19), (i) one among “hotel” and “airline” should be selected from “booking” variation point (line 20) and (ii) one or at most three of “cruise”, “car rental” and “activities” variants can be selected. The variation point and variant declarations are referred from the package composition as depicted in Table 3. Firstly package composition imports its configuration interface (line 2) and defines member packages with/without their configuration interfaces (lines 3-5). Then context parameters are specified that are used within the composition as global (lines 6-8). Then variability mappings between “travelitinerary” package and member packages are identified. Variants of “extras” variation point is mapped with those of “facilities” variation point of “additionalactivities” package (lines 10-12). This way, for instance when “cruise” variant is selected, “additionalfacilities” component is configured as if “cruise” variant was selected. Lastly provided functions are indicated with variable parts by variability attachments: lines 19, 20, 24-27, 31-34 and 35-37. Line 19 states that if “airline” variant of “booking” variation point is selected, interaction between “traveler” and “airline_mng”, take place within the composition.

TOOL SUPPORT

After modeling a component oriented system with XCOSEML, yielding consistent component collaboration requires analysis of variability specifications which takes constraints and variability mappings into account. By this way, the effects of higher level variability bindings to those at lower levels are inspected for inconsistent situations. After model analysis, in case of consistent component collaboration, XCOSEML models are configured with regard to variant selections. For this purpose, a tool is implemented in Python 2.7 which employs parsing, analysis, and configuration phases. As an example, after selecting “regular” configuration variant of “itinerary” configuration point and only in case of “hotel” selection from booking (line 20 in Table 2), the resulting Travel Itinerary package is depicted in Table 4.

RELATED WORK

Integration of large scale component oriented systems entails more complex handling mechanisms which enable to specify, track, and bind variability and configure the system as expected. To address variability management in component oriented systems various approaches have been introduced. A list of existing approaches is compared in Table 5 with respect to the following criteria:

- **Variability Modeling**: Defines the model the approach uses to model variability.
- **Variation Point and Variants (VP&V)**: Defines whether the approach has support for variation point and variant specification explicitly. “No” indicates that there isn’t any specification, otherwise “Yes” is stated. If the approach has a mechanism to classify variation points as external or internal, then “Ext-Int” is added.
- **Constraints**: Defines whether the approach enables to define constraints within the model.
- **Variability In**: Defines the parts where the approach has support variability, namely in “Connector”, “Component”, “Interface”, and “Composition”.
- **Tool Support**: Availability of tools.

In (Ileri, et al., 2013), a component-based variability modeling approach is proposed to manage variability in component models with newly introduced Component Relational Graphs (CRG). However, the configuration logic is not explicitly defined and the approach addresses only component variability. Saidi et al. (Saidi, et al., 2007) define a multi-view variability model for business components in which UML models are extended with stereotypes to represent variability in functional, static and dynamic views. However, this approach lacks to represent complex nature of variability in composition.

In (van der Storm, T., 2004), variability in component and composition is addressed and modelled with a newly introduced Component Description Language (CDL).

Haber et al. (Haber, et al., 2011) propose a hierarchical approach to introduce variability by extending MontiArc Architectural Description Language (ADL) with variability related statements. Not only is the variability of components specified locally, but also constraints and variant selections can only be applied on the same or adjacent hierarchical levels. It fits the hierarchical structure of component oriented systems; however there is no mechanism to handle composition variability.

A variation point model (VPM) is introduced in (Webber, et al., 2004) from the point of component users, extending UML notations with stereotypes. The approach only considers component variability specification.

Bencomo et. al. (Bencomo, et al., 2008) deal with runtime variability issues dynamically adapting to changing runtime context. OVM model is used for variability specification. Another approach that uses OVM is (Razavian, et al. 2008) in which component, interface and connector variability are specified using UML. However, the approach doesn’t address composition variability.
In (Kim, et al., 2004) a variability model is explained in detail which covers five different variability types, namely variability in attribute, logic, workflow, persistency and interface. Besides, three kinds of variability scope are mentioned; binary, selection and open scope. Although the approach gives details about workflow variability, the way how to model complex variability relations is not mentioned.

de Souza Gimenes et al. (de Souza Gimenes et al., 2004) present a component-based product line for workflow management systems using Catalysis method with UML stereotype variability extensions. Nonetheless, the approach has no mechanism to model composition variability.

Koalish proposed in (Asikainen, et al., 2004) models and configures components and interfaces with explicit variability constructs and resolves variability in compile time. Nevertheless, composition variability is not taken into account explicitly. Combining Koalish with concepts from feature modeling, Kumbang (Asikainen, et al., 2006) is introduced as a domain ontology which is developed as a profile extending the UML metamodel.

**CONCLUSION**

This research presents the enrichment of an architectural description language namely COSEML with explicit variability representation. The meta-model and a textual notation are also provided. When compared to the existing approaches in incorporating variability to Component Based specification tools, we have additionally covered different levels in a hierarchical organization, allowing binding of variability in a top-down approach. Our experimentation with example models demonstrated the usability of this approach. Yet industrial scale case studies are missing. Also, Connector variability will be modelled as our future work.

**REFERENCES**


FIGURES AND TABLES

Fig. 1 XCOSEML Metamodel Overview

Fig. 2 XCOSEML Metamodel
Table 1 Package Specification of Travel Itinerary in XCOSEML Language

<table>
<thead>
<tr>
<th>Package</th>
<th>Composition</th>
<th>Configuration Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>travelitinerary</td>
<td>travelitinerary_comp</td>
<td>travelitinerary_conf</td>
</tr>
</tbody>
</table>

Table 2 Configuration Interface of Travel Itinerary Package in XCOSEML Language

```
Configuration interface travelitinerary_conf of package travelitinerary

internalVP booking:
  optional
  variant hotel
  variant airline
  bindingTime devtime

internalVP extras:
  optional
  variant cruise
  variant carrental
  variant activities
  bindingTime devtime

configuration itinerary:
  varType externalVP
  alternative
  variant regular
  variant vacationpackage
  (min:1,max:1)
  confvariant regular mapping
  VPName booking selectedVariants(hotel airline; min:1, max:2)
  VPName extras selectedVariants(cruise carrental activities; min:0, max:3)

confvariant vacationpackage mapping
  VPName booking selectedVariants(hotel airline)
  VPName extras selectedVariants(cruise carrental activities; min:0, max:3)
```
Table 3 Travel Itinerary Composition Specification in XCOSEML Language

```plaintext
Composition travelitinerary
  use configuration travelitinerary_vconf
  has hotel_mng
  has traveler with configuration traveler_vconf
  ...
Context Parameters
  hotelbookingconfirmation "false"
  ...
Variability Mapping
  VP extras maps package additionalfacilities VP facilities
  Variant cruise maps Variant cruise
  Variant carrental maps Variant carrental
  ...
  VP extras maps package traveler VP activityselection
  Variant activities maps Variant withactivities
  ...
Function planitinerary:
  ...
#vp itinerary ifSelected(hotel)# sequence (traveler.selecthotel send{hotel_mng.select(hotelID)})
  %comp hotelbookingconfirmation= hotel_mng.selecthotel%
  ...
select (guard(hotelbookingconfirmation == "true" and flightticketconfirmation == "true") parallel (
  #vp itinerary ifSelected(airline)# sequence (traveler.bookflight send{airline_mng.book(arrival,departure)})
  traveler.getticket send{airline_mng.processticket(customerID)})
  #vp itinerary ifSelected(hotel)# sequence (traveler.bookroom send{hotel_mng.book(arrival,departure,details)})
  traveler.getvoucher send{hotel_mng.processvoucher(customerID)})
  traveler.reqplan receive from{hotel.sendconfirmedplan(customerID)})
)
```
### Table 4 Configured Travel Itinerary Composition Specification in XCOSEML Language

<table>
<thead>
<tr>
<th>Line</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Composition travelitinerary</td>
</tr>
<tr>
<td>2</td>
<td>has hotel_mng</td>
</tr>
<tr>
<td>3</td>
<td>has traveler</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>Context Elements</td>
</tr>
<tr>
<td>7</td>
<td>hotelbookingconfirmation &quot;false&quot;</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td>10</td>
<td>Function planitinerary:</td>
</tr>
<tr>
<td>11</td>
<td>parallel (</td>
</tr>
<tr>
<td>12</td>
<td>traveler.requestprice</td>
</tr>
<tr>
<td>13</td>
<td>send{hotel_mng.queryprice(startdate,enddate,details)}</td>
</tr>
<tr>
<td>14</td>
<td>...</td>
</tr>
<tr>
<td>15</td>
<td>...</td>
</tr>
<tr>
<td>16</td>
<td>sequence (</td>
</tr>
<tr>
<td>17</td>
<td>traveler.selecthotel send{hotel_mng.select(hotelID)}</td>
</tr>
<tr>
<td>18</td>
<td>%comp hotelbookingconfirmation= hotel_mng.selecthotel%</td>
</tr>
<tr>
<td>19</td>
<td>)</td>
</tr>
<tr>
<td>20</td>
<td>select (</td>
</tr>
<tr>
<td>21</td>
<td>guard(hotelbookingconfirmation == &quot;true&quot;) parallel (</td>
</tr>
<tr>
<td>22</td>
<td>sequence (</td>
</tr>
<tr>
<td>23</td>
<td>traveler.bookroom send{hotel_mng.book(arrival,departure,details)}</td>
</tr>
<tr>
<td>24</td>
<td>traveler.getvoucher send{hotel_mng.processvoucher(customerID)}</td>
</tr>
<tr>
<td>25</td>
<td>)</td>
</tr>
<tr>
<td>26</td>
<td>traveler.reqplan receive</td>
</tr>
<tr>
<td>27</td>
<td>from{hotel_mng.sendconfirmedplan(customerID)}</td>
</tr>
<tr>
<td>28</td>
<td>)</td>
</tr>
<tr>
<td>29</td>
<td>)</td>
</tr>
<tr>
<td>Name</td>
<td>Modeling</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>de Souza Gimenes et al.</td>
<td>UML extension</td>
</tr>
<tr>
<td>Webber et al.</td>
<td>UML extension</td>
</tr>
<tr>
<td>Kumbang</td>
<td>Feature Model</td>
</tr>
<tr>
<td>Kim et al.</td>
<td>new VM</td>
</tr>
<tr>
<td>Saidi et al.</td>
<td>UML extension</td>
</tr>
<tr>
<td>Razavian et al.</td>
<td>OVM</td>
</tr>
<tr>
<td>Bencomo et al.</td>
<td>OVM</td>
</tr>
<tr>
<td>Van der Storm</td>
<td>CDL</td>
</tr>
<tr>
<td>Haber et al.</td>
<td>new VM</td>
</tr>
<tr>
<td>İleri et al.</td>
<td>new VM</td>
</tr>
<tr>
<td>Koalish</td>
<td>new VM</td>
</tr>
<tr>
<td>Luca Gherardi</td>
<td>Feature Model</td>
</tr>
</tbody>
</table>
IDENTIFICATION OF LOCALLY FOUND BACTERIA FOR POTENTIAL USE IN GROUND IMPROVEMENT WORKS BY MICROBIALY INDUCED CALCITE PRECIPITATION (MICP) TECHNIQUE
Nurnajwani Senian, Armstrong Ighodalo Omorogie, Peter Morin Nissom, Lock-Hei Ngu, Dominic Ek-Leong Ong
Swinburne Sarawak Research Centre for Sustainable Technologies, Swinburne University of Technology, Sarawak Campus
Kuching, Sarawak 93350 MALAYSIA

ABSTRACT
Recent technology known as Microbially Induced Calcite Precipitation (MICP) has been an important breakthrough in the field of geotechnical engineering. It is a viable technique of ground improvement that incorporates the use of bacteria that act as a nucleation site and further initiates the formation of calcite precipitation. Over the past decade, the development of this technology has increased from small-scale laboratory work to large experiment work as well as the execution of a novel field test conducted in 2011. Most researchers have been using one common bacterium known as *Sporosarcina Pasteurii* which was also known as *Bacillus Pasteurii*. It is chosen due to its high positive urease activity that degrades urea and through hydrolysis of urea, the formation of CaCO3 can be realised through the mixture of carbonate ion (CO\(_3\)^{2-}\) and the calcium ion (Ca\(^{2+}\)) from supplied calcium chloride. This paper describes the isolation and characterization of urease producing bacteria within the vicinity of Kuching, Sarawak. The sampling location of the bacteria is located at the Kuching Wastewater Treatment Plant and was further analysed and identified in the Microbiology Lab of Swinburne University of Technology, Sarawak Campus. Results from Microbact Test\textsuperscript{TM} Kit and Gram Staining Test identified the bacteria as *Ps Stutzeri* (97.04%), *F. Odoratum* (66.31%), and *Actinobacillussp.* (89.76%). The precipitation of calcite produced through the process of MICP of these locally found bacteria were able to be achieved.

INTRODUCTION
In recent years, land development and construction industries in Malaysia have seen intensive activities. Due to rapid development, marginal ground has been earmarked for development as a result scarcity of good ground. Some of these areas are not suitable for construction at times. Engineers nowadays are faced with problems related to poor ground conditions such as land having loose sands, peat soils or soft clays. Due to their relatively low shear strength, suitable ground improvement technology is seen a viable option that can be used to treat these sites. Soil strength improvement technologies that are normally used in Malaysia are mostly via chemical grouting process which involves injection or mixing soil with additives such as Portland cement, sodium silicate, lime, fly ash and etc (Sina et al. 2010). This technique is in fact not as favourable since it is not environmentally sustainable over the long term as the use of Ordinary Portland Cement (OPC) has a very high carbon footprint on the environment. Besides that, this process may take a longer execution time and involves relatively high cost.

The Microbially Induced Calcite Precipitation (MICP) technology is relatively new to the field of geotechnical engineering but its potential has been shown to be tremendous. It is a sustainable, environmentally friendly and cost effective ground improvement technique. This technique is also known as bio-grouting when naturally harvested microbes are used to precipitate calcite in order to bind the soil particles through the process of bio cementation and bioclogging. Over the years, many researchers have successfully gained positive results in which it is proven that the strength and permeability of soils can be improved through MICP (Rebata-Landa 2007; van Paassen 2009; Whiffin 2004).

The main objective of this paper is to review the potential of cultivating bacteria that are found locally in Sarawak, to test its ability in producing urease enzyme through hydrolysis of urea and its capability in calcite precipitation to potentially bind soil particles. Furthermore, the identification of the locally found bacteria is determined through Microbact Test Kit based on the biochemical characterization and its percentage (Ahmad 2013; Matota Sudam 2013; Muhammad Zein 2013; Omoregie & Mkwata 2013).

DEVELOPMENT OF BIO-MEDIATED SOIL
MICP was mostly done at experimental stage for the past few years. It was initially implemented in a laboratory scale and mainly focused on fine sands. Currently, the largest experiment so far was done by van Paassen (2009) at Delft University of Technology, the Netherlands. The prototype was 100m\(^3\) and about 43m\(^3\) was successfully cemented biologically within 12 days (Figure 1). According to the researcher, the results were not entirely satisfactory as some of the key parameters for in-situ application still need to be further improved. This remains as a great challenge in the MICP technology. On the other hand, it can be seen from Figure 2 and Figure 3 that within five years of
laboratory experiment, it has been scaled up from a 0.01m to 1.0m column samples and also from 1.0m up to the size of 43m³. A novel field test using bio-grout technique was performed on coarser materials by (van Paassen 2011). It was implemented to reduce the problem of borehole instability in gravel through horizontal directional drilling (HDD) during the installation of pipelines (Figure 4). A large-scale in-situ application is not the main agenda of this research programme, as there are many factors that need to be considered beforehand that require thorough investigation before the bacteria injection methods can be perfected and applied successfully.

Many microorganisms have been known to be able to precipitate calcium carbonate or widely known as calcite through a variety of ways. One way is through the hydrolysis of urea (Eq. 1) in which it decomposes urea \((\text{CO} \text{(NH}_2\text{)}_2)\) as the positive urease reacts with water. Hence, the precipitation of \(\text{CaCO}_3\) (Eq. 2) was formed through the mixture of \((\text{CO}_3^{2-})\) from urea hydrolysis and the calcium ion \((\text{Ca}^{2+})\) from supplied calcium chloride (Ivanov & Chu 2008).

\[
\text{CO} \text{(NH}_2\text{)}_2 + 2\text{H}_2\text{O} \rightarrow 2\text{NH}_4^+ + \text{CO}_3^{2-} \quad (1)
\]

\[
\text{Ca}^{2+} + \text{CO}_3^{2-} \rightarrow \text{CaCO}_3 \quad (2)
\]

The formation of the calcite precipitation helps strengthen the bond between the soils, which increases the shear strength as well as reducing its permeability (De Muyrck et al. 2008). *Sporosarcina Pasteurii* formerly known as *Bacillus Pasteurii* has been used extensively by researchers for the past decade due to its high urease activity (Al Qabany, Soga & Santamarina 2011; DeJong, Fritzges & Nüsslein 2006; Ivanov & Chu 2008; van Paassen 2009; Whiffin 2004). Few types of known commercial bacteria are also used by others such as *Bacillus Megaterium*; largely found in natural tropical soil (Ng et al. 2013) and *Pseudomonas fluorescens*; exists naturally in sediments (Rebata-Landa 2007). However, the novelty of the Authors’ research is to identify locally available bacteria with MICP capabilities in view of Sarawak’s rich natural biodiversity. The successful methodology is described hereinafter.

**MATERIALS & METHODS**

As mentioned earlier on, literature review shows that most researchers only used commercially available bacterium for their MICP process. Limited researches have used indigenous bacteria or locally found bacteria in this process. Hence, this part of the paper explains the methodology developed to identify locally available bacteria to produce urease enzyme through the hydrolysis of urea. Its ability in producing this particular enzyme was tested with regards to time taken to produce the enzyme.

**Collected Samples**

Samples were collected from four different locations within the compound of the Kuching Wastewater Treatment Plant. The samples were collected at a depth of 10 cm and were then placed in sterilized polypropylene bottles (Figure 5). These samples were used for further microbiological analysis.

**Soil Inoculation in Enrichment Culture**

Nutrient broth (pH 7) containing 2% urea were prepared into 500 mL capacity schott bottles and sterilized in an autoclave machine at 121°C for 15 min at 104kPa pressure. 50mL sterile nutrient broth (+2% urea) was then distributed into 100mL capacity conical flasks. One gram of soil samples were inoculated into each of the conical flasks and incubated at 30°C for 120 hours under shaking condition (130rpm). The step was repeated except 5 mL of culture media was inoculated into conical flasks containing 45mL of nutrient broth (+2% urea) and incubated at 30°C for 120 hours under shaking condition (130rpm).

**Serial Dilution**

Initially, each soil sample was serially diluted making a six fold dilutions. One mL of enriched cultures was mixed with 9 mL of sterile distilled water into test tubes using micropipette. The first and the last three dilution factors were used for the screening of bacteria species onto nutrient agar (containing 2% urea) plates. The first three dilution factors are too concentrated and the counting of bacteria colonies would be difficult. It is easier to detect growth of microorganism in those of low dilution factors as they will be scattered apart. Hence, the final three dilution factors were chosen for screening of urea degrading bacteria onto nutrient agar (+2% urea) plates (Figure 6).

**Screening on Nutrient Agar (+2% Urea)**

Nutrient agar (+2% urea) media were prepared and sterilized in an autoclave machine at 121°C for 15 min at 104kPa pressure. They were then transferred into petri dishes inside biological safety cabinet to allow cooling and solidification. 0.1mL of \(10^4\), \(10^3\) and \(10^2\) dilution factors of the diluted enriched cultures were transferred onto the petri dishes and spread evenly using L-spread. The plates were then incubated at 30°C for 48 hours. This was done to screen for urea degrading bacteria.

**Bacterial Enumeration**

Bacteria enumeration was done by using total plate count method, this was performed on nutrient agar plates and nutrient agar (+2% urea) plates. Growth of microorganism on nutrient agar (+2% urea) plates obtained from serially diluted enriched cultures which were compared against growth of microorganism on nutrient agar plates without containing 2% of urea.
obtained by the transfer of serially diluted one gram of soil samples collected at the sampling location.

Isolation of Urea Degrading Bacteria

Nutrient agar (+2% urea) media were prepared and sterilized in an autoclave machine at 121°C for 15 min at 104kPa pressure. After cooling, the media were transferred into petri dishes inside the biological safety cabinet. Bacteria colonies spread on nutrient agar (+2% urea) plates were sub-cultured repeatedly onto newly prepared nutrient agar (+2% urea) plates and were incubated at 30°C for 48 hours. The isolated organisms were purified through repeated subculture method using a serial dilution streaking technique until pure colonies were obtained on petri dishes. The purification of the isolates was also confirmed by microscopic observation.

Identification of Ureolytic Bacterial Isolates

Biochemical test was performed using Microbact™ Kits (OXOID) and was identified using Microbact 2000 software. The bacteria will be identified using morphological input; observing the size, edge, pattern, opacity, and colour of the bacteria colonies grown on the petri dishes. Gram staining was performed and used to distinguish between gram-positive and gram-negative bacteria and also to identify the isolated microorganisms which helped in the classification of the bacteria.

Production of Urease on Urea Agar Base Slants

The urea agar base slants are performed to identify the abilities of the bacteria to produce their own urease enzyme as well as time taken to produce them. Preparation of the slants was done by adding 7.25g of urea agar base powder into a 100mL Schott bottle consisting of 25mL of distilled water. Then, 3.75g of agar powder were added into a 250mL Schott bottle containing 225mL of distilled water. Only the agar media was autoclaved at 121°C for 15 min at 104kPa pressure.

After cooling the sterilized agar to 45 – 50°C, the urea agar base was sterilized by membrane filtration method using sterile syringe filter, 0.22µm. Both media were then mixed thoroughly and transferred to sterile universal bottle in slant position to solidify. Pure isolated colonies that were obtained earlier on were streaked onto the urea agar base slant and incubated for 144 hours at 30°C. Five pure colonies were obtained and it was grouped into the sterile universal bottles namely Group 1 to Group 5 respectively. The results of this test show the ability of the bacteria to produce the urease enzyme and the time taken to produce them.

RESULTS & DISCUSSION

Through a series of test that involves different methods in identifying the locally obtained bacteria, it has shown that all of the bacteria obtained are urea degrading bacteria and were able to produce urease enzyme. In brief, the bacteria were simply found at Kuching Wastewater Treatment Plant which is within the vicinity of the city. Hence, this indeed saves cost and time, knowing that it is not a difficult task to be done.

The production of ammonia creates a pungent smell which gives rise to the pH of the media. As explained by Al-Thawadi (2011), the production of carbonate is due to the ureolytic bacteria that hydrolyses urea to generate (adenosine triphosphate) ATP by the flowing out of ammonium ions through ATP-synthase process. From then on, with the occurrence of excess calcium ions, calcium carbonate precipitates. It is shown in Figure 7 where calcite precipitate was formed at the bottom of the conical flask. It was after 120 hours incubation of the nutrient broth (+2% urea) media that was inoculated with one gram of soil sample into the conical flasks at 30°C under shaking condition (130rpm). Therefore, this conforms that ureolytic bacteria does exist in all the soil samples collected and thus justifies the precipitation of CaCO3.

As for the isolation of bacteria colonies that are able to degrade urea, nutrient broth (+2% urea) was used as it only retains those that are able to produce urease enzyme and at the same time eliminating that are non-urea degrading. This can be seen in Table 1 that shows growth of bacteria in nutrient agar only is twice as much than those grown in nutrient agar (+2% urea).

During the urea agar base slants test, the agar started changing colour in less than two hours which shows the production of urease enzyme is at a faster rate compare to others (Figure 8). As for Figure 9, Figure 10 and Figure 11, they depict the changing of their colour after 24 hours, 48 hours and 120 hours respectively. At 120 hours, all the urea agar base slants have completely changed colour from orange to pink. It denotes that the bacteria can both degrade urea and are able to produce urease enzyme though the time taken to produce the enzyme differs displaying different characteristics of bacteria.

Urea degrading bacteria isolates were identified based on cultural, microscopic and biochemical characteristics. The bacteria strains were grouped into five groups. Colony morphologies of these five groups of bacteria were observed after growing on nutrient agar (+2% urea) plates after 48 hours incubation at 30°C. Most of these bacteria isolates were seen as smooth, milky or yellowish colonies. Gram staining technique revealed that all the urea degrading bacteria isolates were gram negative pink rods as shown in Table 2. Nevertheless, all the bacteria groups showed positive motility test. Biochemical characteristics of the bacteria isolates was performed by using Microbact™ test kit and the identification of these bacteria isolates are based on percentage probability which was performed using Microbact 2000 software. As mentioned by Sanchez-Moral et al. (2003) in his study, Actinobacillus sp as seen in Table 2, promotes precipitation of CaCO3 under appropriate conditions.

It is also observed that Group 1 took the longest time to produce urease during the urea agar base slant test while Group 2 records the fastest rate of producing
urease enzyme. Assuming the Microbact Test™ is accurate, the bacteria in both group are of the same bacteria except the pH level of these bacteria are at 6.11 and 8 respectively. This proves the statement made by Ng et al. (2014) in their research in which the optimum pH level ranges from 7.5 to 8. Through these results, it can be concluded that having too low or too high of pH level restricts the formation of calcite thus pH level that is optimal for MICP should range within 7 to 9.

Results from Microbact Test™ Kit and Gram Staining Test identified the bacteria as *Ps Stutzeri* (97.04%), *F. Odoratum* (66.31%), and *Actinobacillus sp.* (89.76%). The precipitation of calcite produced through the process of MICP of these locally found bacteria were able to be achieved and thus shows the potential use of these bacteria in soil improvement works, which constitute the other part of this trans-disciplinary research.

**CONCLUSION**

In conclusion, it is proven that precipitation of calcite from locally obtained bacteria can be achieved. This marks an important lead towards the field of geotechnical engineering, specifically in ground improvement methods. The next step would be applying this to the soil sample such as sands and silts for the process of bio-cementation and bio-clogging to take place. There are several outcomes that are anticipated from this research, as pointed out by Karatas (2008). It is expected that the shear strength and stiffness of the soil will increase to improve soil bearing capacity and reduce associated settlements.

With potentially successful bio-soil cementation anticipated in the near future based on the positive outcomes of the current research, it will eventually facilitate efficient and safe excavation and tunnelling processes, whereby sand liquefaction (loss of strength) under significant hydrostatic heads can be minimised. Such construction issue is faced by engineers and contractors who are required to construct deep shafts in excess of 20m underground to enable the execution of micro-tunnelling works within Kuching City.

Thus, through several successful runs of laboratory tests in identifying locally available bacteria and with appropriate methodology currently being developed to inject it into soils, it is positively anticipated that MICP can be a viable and environmentally friendly method in improving the geotechnical properties of marginal soils.

**ACKNOWLEDGEMENT**

We thank Ahmad, Bilal Zein and Matota Sudam for their contribution in this research project. This work was undertaken as a collaborative, trans-disciplinary effort between Civil Engineering and Biotechnology research students of Swinburne University of Technology, Sarawak Campus. We also thank the support of the lab technicians of both the Science and Engineering laboratories.

**REFERENCES**

Ahmad, H 2013, 'Development of Bio-Soil', Final Year Research Project (Undergraduate), Swinburne University of Technology, Sarawak Campus, Sarawak, Malaysia.


Matota Sudam, PW 2013, 'Use of microbial technology to improve strength and permeability of soil', Final Year Research Project (Undergraduate), Swinburne University of Technology, Sarawak Campus, Sarawak, Malaysia.

Muhammad Zein, BA 2013, 'Use of microbial technology to improve strength and permeability of soil', Final Year Research Project (Undergraduate), Swinburne University of Technology, Sarawak Campus, Sarawak, Malaysia.

Ng, WS, Lee, ML, Tan, CK & Hii, SL 2013, 'Improvements in engineering properties of soils through microbial-induced calcite precipitation', *KSCE Journal of Civil Engineering*, vol. 17, no. 4, 2013/05/01, pp. 718-728.


Omorogie, AI & Mkwata, HM 2013, 'Screening for proteolytic and ureolytic enzyme producing bacteria in soil samples obtained', Final Year Research Project (Undergraduate), Swinburne University of Technology, Sarawak Campus, Sarawak, Malaysia.
Rebata-Landa, V 2007, 'Microbial activity in sediments: effects on soil behavior', Dissertations PhD, Georgia Institute of Technology, Atlanta, United States.
Whiffin, VS 2004, 'Microbial CaCO3 precipitation for the production of biocement', PhD thesis, Murdoch University, Western Australia.

FIGURES AND TABLES

**Fig. 1** Set-up for large scale bio-grout experiment (100m³), (van Paassen 2009)

**Fig. 2** Bio-mediated ground improvement from 2003 till 2007, (van Paassen 2011)

**Fig. 3** Largest scale of Bio-mediated ground improvement as of 2008, with 43m³ biologically cemented, (van Paassen 2011)

**Fig. 4** Biocemented gravel after process of drilling, and the borehole remained stable, (van Paassen 2011)

**Fig. 5** From left to right; Sample 1, 2, 3 and 4 respectively
Fig. 6 Serial dilution of soil samples; 6 fold dilutions

Fig. 7 Precipitation of calcite at the bottom of flasks as a result of hydrolysis of urea by urease enzyme producing bacteria

Fig. 8 Urea agar base slants showing colour change after 2 hours

Fig. 9 Urea agar base slants showing colour change after 24 hours

Fig. 10 Urea agar base slants showing colour change after 48 hours

Fig. 11 Urea agar base slants showing colour change after 120 hours

Table 1 Comparison of bacterial growth on nutrient agar and nutrient agar (+2% urea)

<table>
<thead>
<tr>
<th>Soil Site</th>
<th>Nutrient Agar (10^9)</th>
<th>Nutrient Agar (+2% Urea) (10^9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22 14 5 13 9 2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>41 10 7 18 8 4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16 9 16 6 4 3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3 6 9 1 2 1</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Morphological and biochemical characteristics of ureolytic bacteria

<table>
<thead>
<tr>
<th>Group No</th>
<th>CULTURAL CHARACTERISTICS</th>
<th>MICROSCOPIC CHARACTERISTICS</th>
<th>BIOCHEMICAL CHARACTERISTICS</th>
<th>Supposed Bacterial Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Colony Morphology</td>
<td>Gram Staining</td>
<td>Motility</td>
<td>Oxidase</td>
</tr>
<tr>
<td>1</td>
<td>Smooth, intermediate and rough yellowish colonies</td>
<td>Negative, pink rods</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>2</td>
<td>Smooth, intermediate and rough yellowish colonies</td>
<td>Negative, pink rods</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>3</td>
<td>Smooth, Low convex, circular edge, light yellowish colonies</td>
<td>Negative, pink rods</td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>4</td>
<td>Rough irregular edge, mucous, translucent colonies</td>
<td>Negative, pink rods</td>
<td>Negative</td>
<td>Positive</td>
</tr>
</tbody>
</table>
DEVELOPMENT OF PHOTOGRAMMETRIC MICRO UNMANNED AERIAL VEHICLE (UAV) SYSTEM FOR RESEARCH PURPOSE

Norhadija Darwin, Anuar Ahmad and Othman Zainon

Department of Geoinformation
Universiti Teknologi Malaysia
81310, Skudai, Johor MALAYSIA

ABSTRACT
Unmanned Aerial Vehicle (UAV) System is one of the fast-developing technologies that have potential to be used by many researchers and engineers from the private or government agencies. For this technology to continue expanding, issues regarding the time availability, coverage area, reliability, and cost effective have to be addressed especially in research purpose. Hence, this study aims to investigate the potential of photogrammetric micro UAV such as rotary-wing and fixed-wing UAV for research purpose. This study used an autonomous micro rotary and fixed wing UAV that equipped with higher resolution digital camera for image acquisition. From the aerial images, photogrammetric image processing was executed to produce photogrammetric output such as orthophoto and Digital Surface Model. The accuracy of output was assessed based on Root Mean Square Error (RMSE). Finally, the micro UAV system has potential to be used in research purpose and produces good accuracy.

INTRODUCTION
Photogrammetry Micro Unmanned Aerial Vehicle (UAV) System

Photogrammetry is an art; science and technology to obtain information about the features on the ground and surrounding environment. Photogrammetry becomes an art due to the obtaining of reliable measurements which requires certain skills, techniques and decisions. Other than that, Photogrammetry is a science and also a technology because it obtains an image and transforms it via technology into significant results (Fryer, 1996).

Previously, the first photogrammetry is a remote sensing technology developed for the purpose of topographic mapping. Photogrammetry started in 1849 when Aime’ Laussedat, a French army officer make his attempt on the use of aerial photograph for topographic mapping applications.

Principally, there are two (2) category of photogrammetry which is the far range photogrammetry (with camera distance setting to indefinite) and closes range photogrammetry (with camera distance setting to finite values). These both depend on the lens-setting. Another grouping is the aerial photogrammetry (which is mostly far range photogrammetry) and terrestrial photogrammetry (mostly close range photogrammetry).

Digital photogrammetric involves the usage of computer software that enables data processing be done much faster and storing data in digital form. In addition, the digital method is much cheaper compared with the conventional method. Thus, this method is suitable for those with limited budget on their surveying projects. Other advantages associated with digital photogrammetric are increases the speed in processing images, it automatically acquires X, Y and Z coordinates, efficiently processes aerial photograph and finally, able to analyze more effectively.

Recently, digital photogrammetry has embraced the Unmanned Aerial Vehicle or simplified as “UAV” system technology. It then becomes known as the UAV photogrammetry. Subsequently, UAV can be understood as photogrammetry tools (Eisenbeiss, 2009). The term UAV has been used commonly by the community of computer science and artificial intelligence which defined UAV as flying in the air with no person onboard with capability controlling the aircraft. There has been tremendous emphasis on micro unmanned aerial vehicles such as rotary wing and fixed wing types over the past ten years. In addition, UAVs have also been equipped with Global Positioning System (GPS)/ Inertial Navigation System (INS) data onboard (Sannidhi, et al., 2013).
Micro UAV for research purpose

In the few past decades, UAV was used only for military and civilian applications in the first development (Eisenbeiss, 2009). Then, some developing countries began searching for rapid data acquisition with lower budget in mapping an area, video recording and several monitoring purpose. There have many applications of UAV in the photogrammetry fields such as maintenance, remote sensing, security tasks and other. Currently, the demand for UAV is popularly used in mapping applications such as on landslide, road, vegetation, surveillance and others. Mainly, the GPS/INS system is a function in UAV system used for navigating the UAV to the predicted acquisition points known as waypoints. Therefore, the applications of UAV-systems in photogrammetry became more common. In Laliberte et al., 2007 study, he used images taken from two UAVs that differed in size, payload capacity, flight duration, GPS guidance capability and cost for determining the utility of UAVs on rangeland mapping. There were micro fixed wing low-altitude aircraft have been used for agriculture mapping using Cropcam and multipurpose monitoring using Aerosonde (Lin, 2008). Figure 1 show (a) Cropcam UAV and (b) Aerosonde UAV.

Based on previous researches, UAV has many potential for application in large scale mapping and environmental features such as cultural heritage, landslide and other hazards. Subsequently, image based damage assessments can fulfill the requirement in term of response and information recovery and were collected soon after the event, at the correct location, contain sufficient geographic detailed, and evaluated by skilled interpreters.

Objective

The objectives of this paper are to investigate the potential photogrammetric micro rotary and fixed-wing UAV system for mapping purposes, assess these two UAV systems accuracy and capability and also identify the capabilities of the UAV and digital camera.

ELEMENTS OF PHOTOGRAMMETRY

Flight Planning

The process photogrammetric consists of project planning, image acquisition, image processing, and control data for image orientation, data compilation and presentation of photogrammetric product. Generally, the flight plan contains flight map and specifications which will plan the flight lines and the photo control layout for all precision photography. The flight lines were usually marked on a topographic maps where it shown the starting and ending points of each line. This is important for pilot navigation and also for the photographer in taking the pictures. This flight line is used for many reason such as to discern how many images will be (can) captured in one strip, mapping resolution, overlap of images (i.e 60% overlapped and 30% sidelp) and others is depending on the flight mission. Figure 2(a) shows an example of flight line and Figure 2(b) shows the image diagram. The specifications include the scale of photograph, the camera used for capturing the data/image, flying heights, end lap, side lap, tilt and crab tolerance (Lin, 2008).

Scale of Map

In aerial photogrammetry, the map scale is one of the important part which provides the average scale of photograph and flying height. It also can determine the number of flight lines and calculated the assigned number of photographs with sideslape and endslape. Besides, photographic scale can only be defined at a point. Figure 3 represents the aerial photographic scale.

The scale at a point on a truly vertical photograph is given by:

\[ S = \frac{f}{H} \]

where;
S= photographic scale at a point
f= camera focal length
H= flying height above datum
h= elevation above datum of the point

For a digital camera, additional calculations such as the pixel GSD (Ground Sample Distance) and the flying height need to be determined. Camera distance D; focal length of the lens f; actual size of the image on the film S_i, and the orientation of the camera axis relative to the ground are categorized as the basic geometric fickles that influence the size of the footprint covered by a photograph, d_e.

Photogrammetric Camera

Previously, it is very important to keep the metric camera lens assembly fixed with respect to the focal plane in analog photogrammetry. This camera consists of lens assembly, inner cone, focal plane, outer cone, drive mechanism, and magazine. Figure 4 shows the metric camera and its schematic diagram of aerial camera.

According to Ahmad, 2006, the large format metric camera (i.e 230mm x 230mm or 9 inches x 9 inches) was used to produce aerial photograph and cover large area but hinge on the scale of photography. However, there were only several private companies and government agencies could afford to use metric camera for aerial photography purposes due to its high operational cost.
Nowadays, there are types of digital camera being produced with different sizes and resolutions and its advantages in various research applications. The non-metric camera which is known as the digital camera offers several advantages such as cheap, easy to use, handy, and images are in digital form compared to the metric camera. Moreover, the digital camera can be placed in a light aircraft such as the microlight, and also in other platform (i.e based on their application). There are many researchers such as Fraser and Shortis, 1995; Miyatsuka, 1996; Ahmad, 2000; Ahmad and Tapsir, 2001; Ariff, et al., 2003 have used the digital camera in their studies but for different applications. Figure 5 shows two examples of high resolution digital camera been mainly used in capturing aerial photograph for research purpose.

**MICRO ROTARY WING-UAV**

In this study, Canon Powershot XS230 HS was used in acquiring the simulation model images. This Canon Powershot XS230 HS has 14.0x optical zoom lens and other specification as shown in Table 1. In addition, the Micro UAV or Hexacopter which have six blades where three blades rotate clockwise direction and three blades rotate counter-clockwise has been used in acquiring images. Hexacopter is more stable device and safe to be used for outdoor activity, especially in urban area and capable to capture images from certain altitude. These data collections have been divided into two parts comprise of Ground Control Station (GCS) and Pilot Station. Ground control station (GCS) needs to be complete with the flight planning for the flight path through the waypoints using of Mikrokopter software which covers a dimension 20 m x 10 m and subsequently complete installation for the USB port and has connection with the internet/wireless. Meanwhile, the pilot station needs to clear with the radio controller and digital camera must be attached to the Hexacopter. Because of the Hexacopter transmitter can receive signal from the radio control in the five (5) kilometer radius, so the flight planning must be limited within the specified radius. Figure 6 shows the Hexacopter rotary wing UAV and Table 1 shows the specification of Hexacopter UAV.

**Results of rotary-wing UAV**

In this study, two main results were produced. First is the Digital Terrain Model (DTM) and second is the Orthophoto. DTM is an essential data set which useful for the generation of 3D renderings at any location in the simulation model. DTM consists of X, Y and height information. It also can be used for generating contours automatically for volume computation, multi engineering design work, geodesy and surveying, geophysics, and geography. In digital photogrammetric, digital orthophoto is identified as one of the outputs. An orthophoto is a product that has pictorial quality of a photograph and correct planimetric characteristics. Orthophoto is produced through the process of differential rectification whereby photo tilt, lens distortion, and relief displacement have been eliminated and adjusted. Apart from that, the scale is uniform where the orthophoto has the same characteristics. Hence, it can be used on measuring true distances, coordinates and angles because of its accuracy on earth’s surface representation. Figure 8 (a) shows the result of Digital TM and 8(b) shows the orthophoto result of this study.

**MICRO FIXED WING UAV**

This study also uses micro fixed wing UAV which requires a pilot on the ground and ground control station (GCS) or ground crew. The Canon Powershot XS230 HS was used in acquiring the images. In addition, the Micro fixed-wing UAV which have wing span 1680mm and length 1180mm. Ground control station (GCS) needs to be complete with the flight planning for the flight path through the waypoints. This study uses of the Mission Planner v1.0 software for flight planning based on Google Earth which covers a dimension 3.6 km x 1 km and subsequently complete installation for the USB port and has connection with the internet/wireless. Meanwhile, the pilot station needs to clear with the radio controller and digital camera must be attached to the fixed-wing UAV. Figure 9 shows the fixed-wing UAV and Table 2 shows the specification of fixed-wing UAV.

**Results of fixed-wing UAV**

After images acquisition stage, a total of 504 individual digital aerial images were processed using high performance image processing software. All the digital aerial images were processed successfully with good resolution and many overlap between the images. The digital images successfully produced orthophoto and Digital Surface Model (DEM) of the coastal area based on digital image processing after the step of build geometry and texture. Figure 10 (a) and (b) shows the Digital Elevation Model (DEM) and Orthophoto that covers dimension of 3.6 km x 1 km.

**ACCURACY ASSESSMENT AND DISCUSSIONS**

The root mean square error (RMSE) was used to assess the accuracy of the outputs based on orthophoto from rotary and fixed wing UAV. Therefore, the RMSE formulae to compute RMSE for check points of the orthophoto in equation 2a and 2b. Table 3 shows the comparison of coordinates between ground survey (total station) and image processing software using hexacopter rotary-wing UAV.
Where:
\( x_i, y_i, z_i = \) measured value
\( x_0, y_0, z_0 = \) true value
\( n = \) number of dataset

\[ \text{RMSE}_{(x,y)} = \pm \sqrt{\frac{\sum_{i=1}^{n} (x_i - x_0)^2}{n}} \]  
(2a)

\[ \text{RMSE}_{(z)} = \pm \sqrt{\frac{\sum_{i=1}^{n} (z_i - z_0)^2}{n}} \]  
(2b)

Table 3 shows the comparison of check points between coordinates from ground survey (i.e total station) and coordinates obtained from image processing software, where the calculated RMSE is \( \pm 0.004, \pm 0.006 \) and \( \pm 0.002 \) meter (<1 meter) for coordinate \( x, y \) and \( z \) respectively. It can be seen that the accuracy can be achieved using rotary-wing UAV system based on the one strip of digital aerial photograph for simulation model. Meanwhile, Table 4 shows the comparison coordinate of GPS observation and image processing software using fixed-wing UAV. The smaller the RMSE calculated, the higher the accuracy of orthophoto produced. The smaller the RMSE, the better orthophoto could be produced. It can be concluded that the higher the GCPs was, the better the RMSE. Hence, the accuracy of orthophoto is influenced by the RMSE value.

Table 5 shows the differentiation between photogrammetric micro rotary-wing and fixed-wing UAV in terms of field work, processing, man power and number of digital images acquired. Based on that, both of micro UAV has a potential for rapid data acquisition and processing, less man power and good overlapped digital images for research purpose.

CONCLUSION

Nowadays with the development of digital camera, analysis can be carried out for the small format digital camera attached to UAV. A small format photograph from digital camera has the potential to be used in aerial photogrammetry and analysis can be carried out for the product of aerial photogrammetry such as orthophoto, DTM, contour line and digital map.

Based on the data collection, it is clear that photogrammetric micro UAV technology has the potential to be used for coastal studies in terms of data collection and data analysis. In general, micro UAV system and photogrammetric software is easy to use and need more experience in order to understand how the micro UAV work especially in research purpose. The micro UAV provides more advantages compare to conventional method due to the use of less manpower, limited budget and time constraint in order to produce map in sub meter accuracy.

This study proves that photogrammetric micro UAV has a potential to be used for mapping and monitoring coastal erosion. With this technology, many problems could be solved for various applications especially project with limited budget and small coverage area. As a conclusion, micro UAV platform is very helpful and economical for large scale mapping.

REFERENCES


FIGURES AND TABLES

Fig. 1 (a) Cropcam UAV; (b) Aerosonde UAV
(Source: Lin, 2008)

Fig. 2 (a) Flight line; (b) Image Diagram (Source: Lin, 2008)

Fig. 3 Aerial Photographic Scale (Source: http://www.fao.org)

Fig. 4 Aerial camera Aviophot RC20 from Leica and its diagram of aerial camera (Schenk T., 2005)

Fig. 5 Compact high resolution digital camera (Darwin, et al., 2014)

Fig. 6 Hexacopter rotary-wing UAV
Table 1 Hexacopter rotary-wing UAV Specification

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.2kg</td>
</tr>
<tr>
<td>Rotor</td>
<td>6 rotor</td>
</tr>
<tr>
<td>Flight Endurance</td>
<td>15 minutes with a 5000 mAh Li-po battery and camera</td>
</tr>
<tr>
<td>Payload</td>
<td>1kg include Lithium-polymer battery</td>
</tr>
<tr>
<td>GPS on board</td>
<td>Yes</td>
</tr>
<tr>
<td>Special function</td>
<td>Automatically return to home location (1st point)</td>
</tr>
<tr>
<td>Stabilizer</td>
<td>Inbuilt stabilizer to deal with wind correction</td>
</tr>
<tr>
<td>Capture data</td>
<td>Using software to reached waypoints</td>
</tr>
<tr>
<td>Flight control</td>
<td>Manual and autonomous</td>
</tr>
<tr>
<td>Camera stand</td>
<td>Flexible camera holder</td>
</tr>
<tr>
<td>Flight altitude</td>
<td>100m above the ground</td>
</tr>
</tbody>
</table>

Table 2 Fixed-wing Specification

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>2-3 kg with Lithium-polymer battery</td>
</tr>
<tr>
<td>Dimensions</td>
<td>Wing span 1680mm, length 1180mm</td>
</tr>
<tr>
<td>Flight</td>
<td>45 min -1h with a 5000 mAh Li-po battery and camera</td>
</tr>
<tr>
<td>Endurance</td>
<td></td>
</tr>
<tr>
<td>Payload</td>
<td>130g – 500g for one to three cameras</td>
</tr>
<tr>
<td>GPS on board</td>
<td>Yes</td>
</tr>
<tr>
<td>Special function</td>
<td>Automatically return to home location (1st point)</td>
</tr>
<tr>
<td>Stabilizer</td>
<td>Inbuilt stabilizer to deal with wind correction</td>
</tr>
<tr>
<td>Capture data</td>
<td>Using software to reached waypoints</td>
</tr>
<tr>
<td>Flight control</td>
<td>Manual and autonomous</td>
</tr>
<tr>
<td>Camera stand</td>
<td>No flexible camera holder</td>
</tr>
<tr>
<td>Flight altitude</td>
<td>&lt;450m</td>
</tr>
</tbody>
</table>
### Table 3 Comparison of coordinates based on rotary-wing UAV

<table>
<thead>
<tr>
<th></th>
<th>Total Station</th>
<th></th>
<th>Image Processing Software</th>
<th></th>
<th>Differences</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td></td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>01</td>
<td>10013.262</td>
<td>9993.894</td>
<td>20.155</td>
<td>10013.284</td>
<td>9993.909</td>
<td>20.130</td>
</tr>
<tr>
<td>02</td>
<td>10013.017</td>
<td>9995.136</td>
<td>20.108</td>
<td>10012.995</td>
<td>9995.121</td>
<td>20.133</td>
</tr>
<tr>
<td>03</td>
<td>10011.993</td>
<td>9994.318</td>
<td>20.183</td>
<td>10012.016</td>
<td>9994.338</td>
<td>20.169</td>
</tr>
<tr>
<td>04</td>
<td>10011.835</td>
<td>9994.314</td>
<td>20.157</td>
<td>10011.847</td>
<td>9994.329</td>
<td>20.137</td>
</tr>
<tr>
<td>05</td>
<td>10011.188</td>
<td>9995.444</td>
<td>19.972</td>
<td>10011.174</td>
<td>9995.453</td>
<td>20.001</td>
</tr>
<tr>
<td>06</td>
<td>10010.346</td>
<td>9994.063</td>
<td>20.086</td>
<td>10010.330</td>
<td>9994.044</td>
<td>20.061</td>
</tr>
<tr>
<td>07</td>
<td>10009.942</td>
<td>9996.762</td>
<td>19.847</td>
<td>10009.965</td>
<td>9996.777</td>
<td>19.862</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>±0.004</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4 Comparison of coordinates based on orthophoto from fixed-wing UAV

<table>
<thead>
<tr>
<th>Check Point</th>
<th>GPS Coordinate</th>
<th>Image Processing Coordinate (Orthphoto)</th>
<th>Coordinate Different</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Northing (m)</td>
<td>Easting (m)</td>
<td>Elevation Height (m)</td>
</tr>
<tr>
<td>1</td>
<td>239943.465</td>
<td>478689.248</td>
<td>5.182</td>
</tr>
<tr>
<td>2</td>
<td>239926.316</td>
<td>478856.084</td>
<td>4.877</td>
</tr>
<tr>
<td>3</td>
<td>240017.424</td>
<td>478732.621</td>
<td>5.182</td>
</tr>
<tr>
<td>4</td>
<td>240164.666</td>
<td>478859.538</td>
<td>4.877</td>
</tr>
<tr>
<td>5</td>
<td>239827.757</td>
<td>478831.213</td>
<td>4.572</td>
</tr>
<tr>
<td>6</td>
<td>239900.227</td>
<td>478642.831</td>
<td>3.658</td>
</tr>
<tr>
<td>7</td>
<td>239890.097</td>
<td>479242.283</td>
<td>4.267</td>
</tr>
<tr>
<td>8</td>
<td>239658.698</td>
<td>479112.149</td>
<td>3.962</td>
</tr>
<tr>
<td>9</td>
<td>239787.509</td>
<td>478837.331</td>
<td>2.438</td>
</tr>
<tr>
<td>10</td>
<td>239724.009</td>
<td>478985.556</td>
<td>3.658</td>
</tr>
<tr>
<td>11</td>
<td>239944.471</td>
<td>478633.628</td>
<td>5.486</td>
</tr>
<tr>
<td>12</td>
<td>239942.893</td>
<td>478460.584</td>
<td>5.791</td>
</tr>
<tr>
<td>13</td>
<td>239960.789</td>
<td>478408.08</td>
<td>5.486</td>
</tr>
<tr>
<td>14</td>
<td>240015.397</td>
<td>479053.982</td>
<td>3.962</td>
</tr>
<tr>
<td>15</td>
<td>240039.701</td>
<td>478225.58</td>
<td>5.486</td>
</tr>
<tr>
<td>Mean Square Error</td>
<td>0.006</td>
<td>0.008</td>
<td>0.022</td>
</tr>
<tr>
<td>Total of Root Mean Square Error (RMSE)</td>
<td>0.010</td>
<td>0.022</td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT

Addressing the practical issue encountered in resource constrained schools where access to computer technology is limited, the exploration of single display groupware (SDG) provides an alternate solution to the problem. SDG is a technology which allows multiple users to work concurrently on a single display. The aim of this study was to investigate the effectiveness and the efficacy of SDG as educational technology in order to improve engagement and promote collaboration. A prototype SDG system, RimbaIlmu, is designed and developed. Quantitative evaluations were carried out and the usability results showed the feasibility of RimbaIlmu towards community engagement while overcoming technical constraints.

INTRODUCTION

Computer aided learning (CAL) system is used to deliver instructions, enabling interaction between learner and tutor, as well as enabling interaction between users and resources through computer technology (Perraton and Creed, 2000). Computer aided learning supports students’ engagement, interaction, individuality and collaboration on teaching and learning. In Malaysia, computer aided learning system has been introduced to motivate students to learn (Abdollah and Ahmad, 2010), promote interaction and collaboration on teaching and learning (Saad R. M. et al., 2007), enhance students’ engagement and further improve the effectiveness of teaching and learning in classrooms.

To date, two common systems have been adopted and developed as computer aided learning systems in Malaysia. They are the multimedia courseware and game-based learning (GBL). The multimedia courseware presents the course content in various media forms such as audio, video, animation, graphics, text, voiceovers and narration (Mz and Sy, 2008). On the other hand, the digital game based learning is used to create a profound impact in specific learning goals and experiences. It is a paradigm that uses games with underpinning pedagogies to deliver the learning content in a more interactive and creative manner (Mz and Sy, 2008).

Not until recently, mobile games (WaiShiang C., et al., 2013) have been introduced to engage the students and 3D-virtual learning environment has been introduced to provide a social space and a shared platform in which students can discuss, learn, re-learn synchronously and asynchronously, unbounded by distance (Safaei and Shafeieiyoun, 2013)(Normala R. et al., 2013)(Mohd-Hishamuddin A. R. et al., 2013)(Erhel and Jamet, 2013).

As aforementioned, computer aided learning systems have been embraced as an educational tool to enhance teaching and learning in schools as well as to reinforce concepts. This has led to the exploration of the feasibility of deploying CAL and its learning effect in complementing the teaching and learning for indigenous communities within the rural areas. It is imperative to highlight the challenges faced in introducing education technology to the rural communities. Unlike the urban communities, the populace in the rural areas have minimum or no exposure to technology. Their cultural background is another hurdle to overcome which affects their behaviour and acceptance of Information technology, IT. Apart from that, limited computer facilities in those areas have discouraged the use of this system among the community members as they would have to wait for their turn to use the computer system. These are the barriers in the attempts to usher in digital based learning to these people.

To resolve these issues, we are exploring the use of single display groupware. The single display groupware allows the students to work together in a multi-user single computer and single screen. We argue that the deployment of computer aided learning system for community of development consists of the following criteria.

Community engagement: the project introduced must be tight to the community needs. Failing to do so will lead to the failure of any application for development.
Engaging the community in the project will cultivate the sense of belonging, broaden their perspectives and giving them a sense of satisfaction. It is important to accept the application pervasively as part of the lifestyle.

Fun and engaging: It is important to deliver a system that is simple, user-friendly yet engaging. A complex system will demotivate the community as they are not technologically savvy, causing them to shy away from using the system. Fun and entertaining are important qualities to include to attract them to continuously using the system.

Limited funding and low maintenance: Limited financial resources are a major drawback causing an insufficient providence of computer facilities. The students will have to share the computers. It is impossible to provide regular technical support and maintenance as the area is difficult to access.

Sustainability: The viability and sustainability of this project in the long run is a major issue to consider especially when the funds are depleted. Alternate solution will have to be derived to deal with this issue.

This paper presents the design, development and preliminary evaluation (from the aspect of community engagement) of single display groupware, RimbaIlmu, to resource constrained school in Sarawak. RimbaIlmu is designed as an after school programme to allow the students to learn Bahasa Malaysia during free time. Bahasa Malaysia is taking into account in this project due to the important of it in Malaysia education system. In RimbaIlmu, four players or students are able to involve in after school programme within a shared single display. Each student will allocate with one mouse and they can perform exercises within a single desktop and single display environment.

Section 2 presents the current works on single display groupware. Section 3 presents the design principle of RimbaIlmu. It consists of overall design of the RimbaIlmu, interaction design of RimbaIlmu. The prototype of RimbaIlmu is presented in Section 4. Usability study has been conducted in two education technology related workshops. The results of RimbaIlmu are presented in Section 5. This paper is concluded in Section 6.

REVIEW ON SHARED SINGLE DISPLAY TECHNOLOGY

Background

A single display groupware, also known as shared single display application supports face-to-face collaboration among members over a single display, shared between multiple users with each user having his own input device. The idea of single display groupware is to enable a group of users to share the same display concurrently, which can be a large display or just a monitor and each of the members will have his or her own input device to interact simultaneously with the system. This allows collaborators to benefit from the advantages of face-to-face interactions as they are co-located, as well as able to observe the non-verbal communication, such as the excitement and the enthusiasm of the other party.

Shared single display technology enables the use of multi-pointing devices on the same computer, each with its own respective cursor, coordinating the interactions between a set of cursors within a single display like a single monitor, a projector, a computerized whiteboard (Gupta A. et al., 2010). With that, it is able to reduce the computer ratio to students in schools and thus reducing computer cost per student. It will also prevent the issue where several students are required to share one single PC and only one of them has control over the mouse (Schakenbos, 2008).

Based on the research finding, it is gathered that students can learn better in groups, and at the same time increase their participations and interests (Moravejin et al., 2008). It empowers them to improve their critical thinking skills and facilitates the transfer of knowledge among each other through cooperative learning. Also, it enhance the social awareness among co-located students and to support interactions within a classroom under the supervision of an instructor (Moravejin et al., 2008).

Gupta A. et al., (Gupta A. et al., 2010) had developed an SDG application known as the “Disease and Health Awareness” (DISHA). Using similar technology, DISHA utilizes the MultiMice technology, giving every child his or her own mouse to go along with the application. In the study, this technique is able to achieve the highest retention of learning concepts as compared to single user per computer. It shows significant successful learning outcome as compared to conventional learning method.

To further support the claim, Alcoholado et al. (Alcoholado et al., 2012) study the use of shared display in the learning of Mathematics. They had created an SDG known as the “One Mouse per Child for Basic Math” and it achieves the same outcome as Mischief an DISHA.

Several MultiMice projects have been reviewed and reported in this section. However, to this day, there is no study conducted on the potential of this technology applied in CAL system within the rural communities in Malaysia. Set within this context, the next section will attempt to present the exploration of single display groupware, known as RimbaIlmu.

DESIGNING AND PROTOTYPING SHARED SINGLE DISPLAY GROUPWARE, RIMBAILMU

Desinging of RimbaIlmu

The RimbaIlmu is designed as a shared single educational groupware, which involves a group of students to accomplish certain tasks or activities. It serves as an after school programme, in which the students can explore during their free time. It does not require teacher’s supervision. The students will enjoy the latest activities...
once the teacher has populated it with new questions from time to time. In RimbaIlmu, four players or students are able to involve in after school programme within a shared single display. Each student will allocate with one mouse and they can perform exercises within a single desktop and single display environment.

The design principle is based on CSCL design with certain degree of flexibility as a form of peer learning. Internally, RimbaIlmu is equipped with a simple core management service which is briefly described below:

Coordination- it presides over tasks management and tasks to manage multiple users concurrently. It predetermines the order of tasks and provides the decision of order as to who performs specific tasks and roles in the activities.

Game flow- Once the game commences all the players will have to complete the tasks given. Players are not allowed to exit or quit pre-maturely from the game.

Interactivity- To handle multiple participants' responses synchronously, rules are designed to ensure equal privileges for every participant to ensure a fair distribution of chances.

On the front-end, RimbaIlmu consists of four modules namely the class management, the user management, the subject management, and the console management modules. The class management module allows the teacher to add in the class details like class name, number of students prior to using of it. The user management module manages the students’ identity allowing them to perform the necessary actions such as to complete their profiles or delete their profiles. The subject management module allows the students to select their preferred subject to learn and load the respective activities or tasks. Finally, the console module will provide the workspace for the students, designed with an interactive shared single display environment. The four modules can be further modelled in use case as shown in Figure 1. Figure 1 shows the use case diagram of the RimbaIlmu. It shows various activities to control the class, setting up Bahasa Malaysia content, user management, activity management. The next section delves deeper into the design of the console module. This is important to promote the engagement in RimbaIlmu.

![Use case diagram of the RimbaIlmu](image1)

**Prototyping of RimbaIlmu**

For the capacitive accelerometer studied in this paper, the g-function will be formulated as

![The home page of Rimba Ilmu](image2)

![RimbaIlmu as a shared single display environment](image3)

As shown on Figure 2, the prototype greets the players with a main page adorned with a nature's theme to provide a familiar environment and a sense of connectivity for the students. Figure 3 shows the view of a RimbaIlmu projected on a projection screen with the players exploring the game. The current version of RimbaIlmu can support to a maximum of four players concurrently.

To give a sneak peak of one of the activities, Figure 4 presents a screenshot of the workspace. It is based on the interactive design mentioned in Section 3.2 using intermediary interaction style in a shared single display. In general, the students’ workspace is divided into separate regions, and on the same screen, it displays a dashboard of the players’ status and the gameplay such as the points obtained, status of the player’s turn, the timer, and the access to the control buttons. The players are allowed to pause the gameplay to do reflection or discussion. As shown on Figure 3, four players are loaded in this shared single display, each with its own region. Each player’s profile is represented by different animals as highlighted in the red box. Every player is given an autonomous control over their own respective mouse. The player 1 to player 3 assume the student’s role and are only able to move within their own regions, while player 4, acting as the assessor is able to move across all the regions to check the answers.

![Screenshot for the intermediary interaction style in RimbaIlmu](image4)
EVALUATION OF RIMBAILMU

Rimbailmu has been evaluated in two education technology workshops between May 2013 and October 2013. The 1st colloquium of learning science was the first attempt where we conducted a demonstration on Rimbailmu with the audience. The audience consisted of postgraduate students of learning science and most of them are teachers as profession. They were given the opportunity to test the system and feedback forms were distributed thereafter. The feedback provided in barchart as shown in Figure 5.

An informal observation was carried while the participants were testing the system. The main objective is to obtain as much information as possible on the users' experience such as to gauge the level of engagement, attention and enjoyment. They appeared to be enthusiastic about the concept introduced and eager to explore more on the system by trying out some other modules. This proves that it had triggered their affective response and they were inquisitive in learning more, finding it engaging. Some were engaged in a discussion while trying to solve some challenging tasks, which indicates that peer collaboration was taking place.

In sum, the audience found Rimbailmu useful especially for students in the rural area with limited resources (refer to Q1, Q2, Q4 and Q5). The audience agreed on having different interaction design to cater to various learning style of students as shown in Q3 and Q4. One of the negative comments on Rimbailmu, which is worth highlighting is the need for technical support. As it is a new technology, it requires some guidance to be familiarized with the system as well as in handling the interaction. The rules of the games have to be clearly conveyed in a concise and simplified manner. In addition to that, limited time was allocated; hence audience did not have ample time to explore on it.

From the feedbacks collected and analysed, Rimbailmu was improved to produce the second version. In September, we conducted a workshop on testing the enhanced Rimbailmu among the teachers and teachers' trainers in the International Conference on Teaching and Learning Through Games, iCoTLG, in Kuching, Sarawak, East Malaysia, as shown in Figure 6. The aim of this conference was to gather the teaching experts within the professional institutions, schools as well as NGOs to exchange intellectual knowledge and ideas related to education with hope to internationalize education in Malaysia.

It was a two-day conference in which we presented the Rimbailmu in a workshop session for 1.5 hours with the following agenda. 1) Discourse on the Theories of Collaborative learning, Game-based learning, and Shared Single Groupware technology 2) Introducing the Rimbailmu system and its features 3) Demonstrating Rimbailmu 4) Practicum with participants 5) Collecting feedbacks and ideas for improvements.

All the participants of the workshop have certain degree of ICT proficiency. Due to the limitation of resources, we were only able to setup 3 groups of participants for the testing of Rimbailmu. Participants were grouped randomly based on volunteering basis. A short briefing showing a walkthrough of Rimbailmu was given along with the training as shown in Figure 7. Excitement was building up among the participants as the training was conducted, as shown in Figure 8. This is a positive key indicator to show that educators are keen in a new method of teaching through education technology and they are ready to embrace the paradigm shift in education from the status quo.
As with the first workshop, informal observation was carried out and the audience showed positive response. They were deeply immersed in the system to the extent that they were unaware of the time. They realised the potential of this system in motivating students and in engaging them, and how far this can go beyond the norms of classroom education.

The participants have assigned with 11 question after trying on the Rimballmu. Also, they were given open ended questions for a more constructive feedback and suggestions on the improvement of Rimballmu.

<table>
<thead>
<tr>
<th>Questionnaires</th>
<th>Test case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1, Q2</td>
<td>The acceptance of shared single groupware</td>
</tr>
<tr>
<td>Q5, Q6, Q9</td>
<td>Understanding the interaction styles</td>
</tr>
<tr>
<td>Q3, Q4, Q7, Q8, Q10, Q11</td>
<td>User interface design (general design aspect of Rimballmu)</td>
</tr>
</tbody>
</table>

Table 1 The distribution of the test sample.

Overall, the Rimballmu had received very good responses from the participants. From the survey results, Figure 8, the concept of single display groupware through Rimballmu development was accepted by the audience without any apprehension. All the participants found that the system was user-friendly and not cumbersome. The technology is uncomplicated and easy to setup. Meanwhile, almost all of the participants (98%) did not face any difficulty in using the Rimballmu as a shared display technology. However, one participant did encounter an error in the system.

The interaction styles for the shared single display in Rimballmu had also received good feedbacks from the participants. 18 participants were able to follow the game flow, the navigation and could easily adapt to the controls and display. 19 participants were quick to understand the activities designed within the shared single display environment. Conclusively, most participants agreed that Rimballmu is interactive, simple and easy to use, suitable for novice users with minimum guidance and the content is sufficient to reinforce the concepts taught in class.

From the open-ended questions, we receive the suggestions to:
Participant1-“deploy the Rimballmu in rural school.”
Participant2- “explore the Rimballmu to other subjects like Science and Mathematic”, “promote the usage of Rimballmu within a remote community during the discussion or meeting among their communities.”
Participant3- “Rimballmu needs to cover different discipline of knowledge.”
Participant4-“Rimballmu can be used for problem based learning.”
Participant5- “we should control the number of mouse in order to reduce the confusion.”
Participant6- “instructions need to be clear and consistent.”

CONCLUSION AND FUTURE WORKS

The main motivations for this research are test the feasibility particularly testing the engaging of an interactive single display groupware using MultiMice technology for schools with limited computer facilities. After exploring various related works, we have produced 9 different interaction designs and presented one of the prototypes called Rimballmu. Based on the results collected from the evaluations conducted, the results had proven that Rimballmu is an able to engage with the community (e.g. Students, teachers) with positive affective impacts and motivational values. Also, the deployment of Rimballmu is low and it reduces efforts to maintain lots of computers in a long run. Meeting its core objective, Rimballmu is didactic, harnessing collaborative peer learning and interaction via single shared technology. However, it is still at the continuum of improvement. To further enhance the efficacy of this approach, more focus should be placed on enriching the contents to align them with clear pedagogic objectives. According to (Prensky, 2001), effective educational game design can be achieved through the equilibrium of fun and educational values. Ideally, it can also include additional features such as in-game assessment and grading as a means to evaluate the transformation of the students’ experience into knowledge. Apart from that, to move into a deeper learning experience, the activities should be designed to stimulate and
encourage learners to reflect, develop new ideas and opinions.

The evaluation was carried out on a specific set of learning games on adults on a modest scale. The hypotheses were drawn based on their experience as educators and after testing the system. However, the findings gathered can be strengthened if applied to a larger group of students, particularly from the rural areas with diverse cultural and academic background. Learning is a complex process, influenced by the learner's familiarity with the setting, their prior knowledge, their cognitive level and the cues for learning (Rennie and Maclafferty, 1995)(Rennie and Maclafferty, 1996). The effort of future research on this frontier will undoubtedly provide more insights into learners’ needs in the rural areas, enabling better adaptation of the system to their cognitive standards and to exploit the potential of the technology.

ACKNOWLEDGEMENT

The authors would like to acknowledge ministry of education to fund for this project under exploration research grant scheme, ERGS. Also, the authors would like to acknowledge the interns, Lim Zi Xuan and Boon Hui Kee for their contribution in developing RimbaIlmu.

REFERENCES


MODERN INFORMATION OVERLOAD
Hamda Binghubash Almarri, Radmila Juric
Department of Computer Science, Faculty of Science and Technology,
University of Westminster, London, UK

ABSTRACT
This is a position paper which examines the problem of Information Overload (IO) and gives an overview of various attempts to resolve it. We argue that instead of fighting IO we should start learning how to live with it. It is unlikely that in our modern information age, where each of us is a producer and consumer of information, the amount of data and information we generate would decrease. Furthermore, when managing IO we are confined to algorithms and policies of commercial search engines which create results that also add to IO. Our position is that we should initiate a change in thinking. Firstly we have to give greater power to users when we address the relevance and accuracy of Internet searches, which helps in IO. Secondly, however powerful search engines are, they do not process enough semantics at the moment when search queries are formulated. Consequently we have developed numerous strategies based on users’ past behavior, in order to predict relevance of Internet search results. These did not alleviate the IO problem. By addressing a particular moment in Internet searches and defining/manipulating semantics which describe that moment, we would be in a better position when managing modern IO.

INTRODUCTION
In our digital information age, Information Overload (IO) (Bawden et al., 1999), (LaPlante, 1997), (Schick et al., 1990), (Strother et al., 2012) has been our constant companion since we started producing excessive amount of information on a daily basis. In order to manage information, we heavily depend on search engines which are supposed to help in retrieving the right information at the right time. At the same time search engines overload us with multiple search results which do not necessarily meet our expectation (Kovach and Rosenstiel, 2010), (Doomen, 2009), (Lincoln, 2011) and contribute towards IO in any environment. Furthermore, IO also exists everywhere and it is not confined solely to retrievals from the Internet (Russell et al., 2007), (Baeza-Yates et al., 2007), (Farhoodmand and Drury, 2002), (Simperl et al., 2010), (Lohr, 2007), (Lee and Lee, 2004), (Ulijn and Strother, 2012). If we wished to have a brief overview of works which attempted to resolve IO than we have to understand two things.

Firstly, IO is not a new term in information science because it appeared almost 40 years ago when we tried to retrieve information from very large and numerous databases containing mostly library and bibliographical data. For decades we used traditional Information Retrievals (IR) (Manning et al., 2008) (Norton, 2000), (Cortes et al., 2007), (Baeza-Yates et al., 2007), (Su et al., 2010) in order to reach scattered data around repositories, including traditional databases. These repositories contributed towards IO in the 80s. IR were powerful systems which worked well, because repositories of traditional computing systems from the 80s and 90s were very well structured. One of the first IR systems was created for heterogeneous systems in libraries in the mid-80s (Rijsbergen, 1979), (Saracevic et al., 1988), (Smithson, 1994), and they did address IO.

Secondly, the nature of IO has changed because we have changed the way we perform computing today. In the second decade of the 21st century we generate enormous amounts of data as we live. We tend to save almost all data we create and we use the computational power of modern devices in order to compute at the time we want. On top of this we deal with the Internet on a daily basis and almost all activities in our everyday life are supported by Internet searches. From this perspective IR systems from the 80’s cannot exist anymore: we have moved to Internet searches, triggered by user queries, their results and interpretations. Consequently, we may argue that in modern computing, search engines (Schwartz., 1998), (Seymour et al., 2011) have replaced powerful IR systems from the past (Broder, 2007), but we should agree that behind all these search engines today are new and powerful algorithms which do the job (Baeza-Yates, 2006) (Cortes et al., 2007), (Baeza-Yates et al., 2007), (Su et al., 2010). They are not without faults (Henzinger, 2004), but they do respond successfully to user’s requests for the retrieval and answer their queries.

Our major concern is that we frequently find ourselves in situations where results of Internet searches might not be exactly what we expected. They might even be irrelevant to our query and we might also receive too many of them.

The paper is organized as follows. In the next three sections we outline different ways of addressing IO in the last 2 decades. We talk about Recommender Systems (RS)
and their techniques, we talk about tagging, annotations and folksonomies on the Internet and ranking of Internet searches which have become an important vehicle in managing enormous amounts of information on the internet though search engines. Problems encountered when addressing IO are explained in the section which follows and in the final section of the paper we talk about future trends in research which may alleviate IO.

RECOMMENDER SYSTEMS

One of the most exploited solutions which have addressed the new era of IO in the last 15 years belongs to the category of (RS) (Felfernig et al., 2007), (Ullman, 2012), (Adomavicius and Tuzhilin, 2005), (Ricci et al., 2010), (Baeza-Yates and Raghavan, 2010). RS initially focused on IO in electronic commerce; we are all familiar with recommendations directed towards buyers of Internet products or services. And there are numerous other examples of RS, which do not adhere to the domain of e-commerce RS. We choose to mention just a few. (Hornung et al., 2007) have developed a RS that suggests a list of correct process fragments for business process modeling in order to reduce the number of structural errors in the process. (He and Chu, 2010) give a new paradigm in RS suitable for e-commerce, but based on Social Networks (SN), i.e. it utilizes information from social networks in order to make personalized recommendations when buying products. Consequently, we can see in (Hoens et al., 2010) that privacy issues in RS based on SNs and social interaction can be addressed and include aspects of trustworthiness and relevance of recommendations. Similarly, (Bourke et al., 2011) generate recommendation by incorporating social information in their recommendation process. (Tayebi et al., 2011) reduced crime prevention in the field of low enforcement and intelligent agencies, through their novel approach to crime suspect recommendations. All these works have open doors to a new type of RS which can be based on personal information (Nunes and Hu, 2012). It is encouraging to see that in healthcare we also tried to build RS, having patient privacy in mind when recommending appropriate physicians to diagnose and treat medical conditions.

There are also many attempts to create group RS (Jameson and Smyth, 2007) because they address decision problem in recommendation by more than one user. (Guzzi et al., 2011) propose an interactive multi-partying critique for group recommendations, (Shunichi Seko, 2011) use behavioural tendency and power balance between group members to create group recommendations and (Tschersich, 2011) deal with unavailable information in mobile group RS.

It is no surprise that the pervasiveness in software systems, which required context awareness in software applications (Abowd et al., 1999), brought forward the development of context aware RS (Adomavicius and Tuzhilin, 2008) (Adomavicius et al., 2011) which focus on contextual information (Anand and Mobasher, 2007). It is very difficult to define context in modern computing because it is not solely confined to a location where we run context aware applications. Therefore user’s queries have become an important source of contextual information for creating context aware recommendation (Hariri et al., 2013).

Researchers have looked at various aspects of RS, which range from predicting performance of recommendations (Bellogin, 2011), (Bellogin and Castells, 2010), (Hauff et al., 2008), (Katz et al., 2011), testing the RS on missing data (Steck, 2010) and solving the problem of diversity, similarity and accuracy in RS (Zhou et al., 2010), (Jojic et al., 2011) to recommending item popularity (Steck, 2011) and interactions between users and RS (Knijnenburg et al., 2011).

At the heart of all these RS are various techniques, often associated with filtering of data. One of the most famous and frequently used is collaborative filtering (CF) (Goldberg et al., 1992), (Schafer et al., 2007), (Herlocker et al., 2004), (Herlocker et al., 2004), (Zhao et al., 2011b). Content based filtering (CBF) was elaborated in (Balabanović and Shoham, 1997), (Pazzani and Billsus, 2007) and consequently we could see that they have been placed together in various frameworks (Pazzani, 1999) and in hybrid RS (Burke, 2002). These techniques have also been combined with information we have in SN (Kautz et al., 1997), (Yu et al., 2011) and in knowledge-bases (Burke, 1999), (Tran, 2007). CF techniques are also enriched with personality information (Hu and Pu, 2011), adaptive information sources (Lathia et al., 2009), and personalised tag recommendations in folksonomies (Kim and El Saddik, 2011).

It is important to note that RS techniques may be used for a range of predictions in RS: from creating domain-specific and mediated predictions (Rosenthal et al., 2010), rated predictions (Tang et al., 2013), (Steck, 2013), (Campos et al., 2011) to predicting personalised item distribution (Koren and Sill, 2011), social tags for cold start book recommendations (Givon and Lavrenko, 2009) query intent prediction (Baeza-Yates, 2010) and predicting the desirability of social match (Mayer et al., 2010) and personality traits (Gao et al., 2013b).

TAGGING / ANNOTATIONS AND FOLKSONOMIES

There are other attempts to make RS more modern and its filtering techniques more responsive to user queries. They come in the format of tagging, including social tagging, annotations and folksonomies. However, tagging in general appeared as a consequence of having Web 2.0 (O'Reilly, 2009) applications (Andriole, 2010), (Churchill, 2009), (Wrightemail et al., 2008), where users
are becoming both producers and consumers of information on the Internet. Web 2.0 platforms allow users to post various resources and tag them freely according to their own perception on “what is an appropriate tag for a particular Internet source”. We have been aware of numerous problems in free tagging (Golder and Huberman, 2006), but tagging is good for satisfying user’s needs when publishing and retrieving from the Internet. This can also take us towards personalised tag recommendation (Landia and Anand, 2009), (Kim and El Saddik, 2011). In (Musto et al., 2009) we can see a very good example of tag recommenders in the system called STaR which suggests a set of relevant keywords for an annotation. Consequently, social tagging has rapidly spread, creating an additional dimension to the world of RS techniques. The traditional two dimensions (user-items) have been augmented with “tags” as a third element, i.e. the third dimension, which converted RS into tag-aware recommendation (Zhao et al., 2008), (Trousson et al., 2009).

The hierarchical nature of social tagging systems used in RS (Balby-Marinho et al., 2011) (Klasnja-Milicevic et al., 2010) might not be the best counterpart to our traditional RS techniques and therefore new approaches, as suggested in (Trousson et al., 2009) have encouraged the use of spreading activation techniques, known from (Crestani, 1997), which would allow the extension of the traditional user-item environment into a context where we can tag events (Olston and Chi, 2003) (Hussein and Ziegler, 2008). It is also worthwhile to extend user-item environments towards the context where we pay more attention to their interconnectivities and create some kind of semantic relationship between users and items (Barragáns-Martínez et al., 2010) (Almarri and Juric, 2013). In one of the examples, Queveo.tv system from (Barragáns-Martínez et al., 2010), contains a recommender component which uses CF, CBF, social content based and collaborative recommendations. Collaborative tagging, sometimes associated with folksonomies (Vander-Wal, 2007) is an additional mechanism of sharing, annotating and searching for online Internet resources.

A collection of user annotated resources with their tags can be described in folksonomies as a network of interrelated users, resources and tags, which is very valuable for understanding the nature of Internet sources though folksonomies (Mathes, 2004). Therefore the possibilities of using them are numerous and well explained in (Gemmell et al., 2009b). Consequently, by paying more attention to collaborative tagging/annotations on the web (Macgregor and McCulloch, 2006) we can significantly reduce the time we need for retrieving relevant information and address IO. For illustration of the use of folksonomies in IR (Hotho et al., 2006b), the reader should pay attention to FolkRank, a formal model and search algorithm for folksonomies (Hotho et al., 2006a) which computes a web page rank (Gemmell et al., 2009a), (Kim and El Saddik, 2011). More information on tag recommendation in folksonomies can be found in (Jäschke et al., 2007). (Lipczak, 2008) is gearing tagging through folksonomies towards individual users. The question of using information presented by a folksonomy and the item metadata, in order to boost the traditional CF algorithms is given in (Bogers and Bosch, 2009). However, in all cases we are able to exploit social tagging and folksonomies in order to harvest social knowledge though folksonomies (Wu et al., 2006) (Limpens et al., 2008).

**RANKING OF INTERNET SEARCHES**

We have numerous examples which address IO when performing Internet searches. For this purpose:

- search engines use ranking (Zhao et al., 2011a), (Page et al., 1999), (Dong et al., 2010) mechanisms before they deliver search results and
- algorithms, which run upon the Internet search results and re-rank them.

Readers interested in ranking of search engine results and Internet searches in general, should be aware that the definition of relevance and its levels or degrees are crucial factors when deciding whether ranking techniques are successful or not (Järvelin and Kekäläinen, 2000), (Burges et al., 2005), (Borlund, 2003). Therefore whenever we rank Internet search results we should be aware that

(i) the issue of relevance of search results cannot be ignored and

(ii) the degree and level of relevance should always be determined by user’s queries, i.e. by keywords which are given by the user and which are converted in a “query” by a search engine.

The most published, analyzed and criticized ranking mechanism is the one which runs within the Google search engine (Su et al., 2010). It cannot be completely revealed to the public, because it still remains the company’s business secret. We can only guess through experiments and experiences how Google ranking works. Various authors analyze Google ranking algorithm differently. (Beel and Gipp, 2009) look at the impact of citation counts in the Google Scholar’s ranking. (Dong et al., 2010) are interested in recency ranking in web search in terms of taking the freshness of ranked documents. (Chidlovskii et al., 2000) suggest ranking which is built according to the user and his/her community profiles, and (Lin et al., 2010) look at diversification of search results in order to improve web search effectiveness. They think that we have to strike a balance between relevance and diversity in Internet searches. Re-ranking of search results in (Collins-Thompson et al., 2011) had particular purpose: they wish to address the reading level of children when searching the web and adjust ranking to their needs.
There are other interesting works. (Yong et al., 2008) use machine learning approaches in the ranking of web pages, (Agichtein et al., 2006) suggest incorporating user behaviour in web search ranking and (Stojanovic et al., 2003) and (Chau et al., 2012) use Semantic Web Technology (SWT) in order to address relevance and rankings. It is important to note that the SWT helps to interpret the meaning behind user queries environments where they are created (Almarri et al., 2013) (Almarri and Juric, 2013). The same ideas help to address relationships between search results by exploiting metadata of semantic web sources, which was proposed in (Aleman-Meza et al., 2005) and (Anyanwu et al., 2005).

PROBLEMS IN ADDRESSING IO

Our previous sections have indicated that we might be far from diminishing IO and securing results of Internet searches which satisfy user’s expectations (Mahoney et al., 2009), (Wang et al., 2009), (Addis et al., 2010), (Girard and Allison, 2009). IO has changed in the last decade because the way we create and consume information has changed. The amount of data we create on a daily basis will constantly grow and IO is not something that we can or wish to control (Pollar, 2004), (Westmead, 2013). Today we:
- primarily want to “grab” information from the Internet at the moment we need it and
- tend to retrieve the content of repositories available on the Internet, which are not necessarily structured.

The impact of technologies, which enabled users to become producers and consumers of information, is imminent and we propose to initiate a shift in thinking on ‘how to create modern and more effective ways of searching” when trying to find solutions for managing IO.

Some critics would say that our division of the literature review into RS, tagging, annotations, folksonomies and ranking of search results, might not be the best way of underpinning new ideas for addressing IO. However, the previous section follows the exact chronological order of events which happened across fields of IR, RS and Internet searches in the last 20 years.

To summarize:
1) We should take into account that the nature of IO and the power of traditional IR systems have CHANGED. We cannot take for granted that a technique and a “system” which claimed to address IO in the past, would work today. For example, it would be inappropriate to claim that RS techniques, developed in the 90s, for managing a surplus of information in structured repositories, would work in modern retrievals from the Internet. It does not mean that these techniques are not applicable at all. It means that we should find out if they can address modern IO and contribute towards our shift in thinking when managing it. Furthermore, IO today has become closely associated with Internet searches and results from search engines very often overload us.

2) We should focus on interpreting the semantics of the environments where we experience IO and consequently model both:
- user’s intentions when creating or using information and
- characteristics of the environments where IO happens, i.e. characteristics of information in that environment.

We are aware that an abundance of information available around us is accessible through search engines. If so, then both: the amount of information around us and the results of Internet searches are the main reasons for having modern IO. The collected and interpreted semantics of the environment where Internet searches happen could help us to be more precise and deliver relevant or at least the best possible search result to users.

If we carefully look at techniques found in RS we can see a few major obstacles in using them to improve the results of our Internet searches and reduce modern IO:
- They are all focused on building profiles of users and items, which may be recommended to users, based on user PAST behavior and item’s ranking.
- RS sometimes use CF and CBF algorithms, which measure similarities between the item’s rankings according to keywords which appear in user’s reviews of a particular item. The results of measuring these similarities are then used in building a “better” profile of the user.

However efficient these ideas may have been in early days of e-commerce by telling the user “this is what you might like”, they are very unreliable now if we use them to make recommendations when users search the Internet. Relying exclusively on the past behavior, when modeling someone’s profile and basing recommendations on that profile is risky. Users often change their mind and interests, sometimes instantly when searching the Internet and their profiles might have been built wrongly. The reason is trivial, but powerful: buying a set of books on Amazon on Spiritual Healing, for friends and families, and viewing them before buying do not make the buyer a spiritual healer, it should not be interpreted as the buyer’s “interest” and should not be built into his/her profile because it is wrong. Furthermore one attempt of viewing certain items on the Internet to gather the information does not mean that they are in the focus of anyone’s interest. It can happen accidentally!

In the modern information age, users are in charge of computational environments (Shojanoori, 2013) (Shojanoori et al., 2012). They very often manage the data involved in such computations because they are producers and consumers of information. It is almost impossible to predict exactly, through any of the available Artificial Intelligence (AI) or RS techniques, based on the
knowledge we collected from the users past behaviour, how users would react from situation to situation when generating and retrieving data at the same time and manipulating information on the Internet.

Tagging, annotations and folksonomies are a huge step forward towards managing excessive information in our modern lives. They have allowed users to be involved in the classification of data they generate and consequently secure more relevant results of searches in terms of satisfying user’s expectations. If the user is given the power to classify information he/she creates, then any type of retrieval is expected to give more “relevant” search results to the user. However, there are a few issues with tagging/annotations/ folksonomies:

a) They have been introduced as a consequence of (i) extensive involvement of users in online social networking and (ii) the proliferation of Web 2.0 applications which required tagging. In other words, tagging was not explicitly introduced for addressing the deficiency of RS techniques and for reducing IO. From that respect we should not assume that they can be efficient in solving the problem of IO. They may even aggravate it.

b) Tagging/social tagging/annotations/folksonomies have become very complicated because they address relationships between items, users and tags. It is difficult to think that we can manage the complexity of such relationships through folksonomies. They do not have enough “space” for describing fully the semantics of relationships in such situations which may make them too complicated. Modelling relationships in computer science has always been sensitive, if not a problematic issue, particularly if we do not reserve a special modelling element where the semantics of relationships can be “stored” in folksonomies.

Rankings of search results are present in all search engines today and there is no doubt that they may reduce IO by using various criteria. However there is one important problem there. Ranking algorithms have been, and will remain business secrets of companies which are in charge of search engines. As mentioned before, we can only “guess” how Google ranking works (Google, 2012). As a consequence we find a proliferation of companies today which claim that they will make your web presence “Google-ranking-friendly”, i.e. your URL “might become highly ranked”. In other words, it appears that your URL might never be retrieved if the Google ranking algorithms does not “recognize you”, which will then affect the relevance of information retrieved by the Google engine. Consequently, Internet retrievals through Google, without understanding exactly how its ranking works, will not reduce IO. It may even add to it. How could a ranking system in widely used search engine(s) fit the expectation of all users? No wonder we have smog (Shenk, 1998) and glut (Bawden, 2009) in results of retrievals through modern search engines. We believe that ranking/ratings of Internet searches are not very suitable methods of solving the problem of “relevance” in Internet searches.

FUTURE TRENDS

Our position which defines future trends in IO research is in the bullets below. We

1) Suggest giving more power to users! They are the ones who are producers and consumers of information at the same time. Therefore if we would strive to create a new computational model which will help in resolving problems 1)- 2) and a) - b) from the previous section then we must demonstrate how and where we give the power to the user in order to address the problem of finding the most relevant Internet search.

2) Address a “moment” in a particular situation when Internet searches happen. This immediately implies that the computational model form 1) above might not be interested or involved “in the past”. It would build computations based on capturing as much semantics as possible in that “moment”. The reason for that is simple. Remembering results of retrievals from previous “moments” might not be advisable because (a) each “moment” carries different semantics in terms of the reasons and needs for retrieval and (b) users very often change their mind while searching the Internet. We have to avoid storing the semantics of the results of past retrievals, because it may be wrong information for the next “moment”.

From these two perspectives it we should be able to decide which Internet search results are relevant in which circumstances. However, if we claim that we should understand the circumstances when Internet searches happen, then it is obvious that we should have a technology which can help us to collect, store and manipulate the semantics in such environments. It is prudent to use the SWT stack because it allows us to interpret the meaning of a particular situation where retrievals happen. We can exploit a set of rich languages which give us opportunities to reason upon the semantics of the situation in a particular Internet search. The computational model deployed with the SWT would be a step forward in ensuring that the user receives the most relevant search results. Consequently, SWRL enabled OWL ontologies might be perfect for implementing the computational model which could be deployed in any Integrated Development Environment (IDE) for Java and Andriod based platforms (Patadia et al., 2011)

In our previous publication (Almarri and Juric, 2013) we introduced a generic model for recommending online learning sources. Figures 1 and 2 are excerpts from the proposed model. At the same time, these two figures address 1) and 2) above. This might not be a trivial task to explain here, but if the reader looks at the published model for recommending online learning sources from (Almarri and Juric, 2013) it will be easier to see that
(i) Giving greater power to the user is illustrated in Figure 2. User preferences dictate the way we describe Internet learning sources through its characteristics CS
k. They also ensure that the relationships between these characteristics and user’s preferences are consistently used in reasoning which determines which Internet learning sources are recommended.

(ii) We show in Figure 2 that at every moment we may have a different set of CS
k classes which describe characteristics of learning sources. We may also use horizontal hierarchies of CS
k subclasses in order to exercise more precision when defining CS
k. All these classes in Figure 2 are changeable, which means that at every moment we may have any number of them and we should be able to choose between CS
k and its subclasses and relationships has_SC
k and has_S_Cm
k. Consequently, the model itself is changeable because we do not adhere to a strictly predefined set of characteristics CS
k when defining Internet sources because we leave to users to decide how to tailor their preferences, which would then trigger the existence of SC
k classes.

We are currently testing various instances of the model proposed in (Almarri and Juric, 2013) in situations where the computational model derived from it alleviates the IO problem. We mostly focused on the relevance and accuracy of Internet searches as elaborated in the section where we listed problems of addressing modern IO.

REFERENCES


Chidlovskii, B., Glance, N. S. and Grasso, M. A., 2000, "Collaborative Re-Ranking of Search Results," In Proceedings of the 17th International Conference on Association for the Advancement Artificial Intelligence Workshop (AAAI '00) 30 July - 1 August 2000, Austin, Texas, USA.


Mayer, J. M., Motahari, S., Schuler, R. P. and Jones, Q., 2010, "Common attributes in an unusual context:


O'Reilly, T., 2009, "What is Web 2.0," In O'Reilly Media, Inc.


13th international conference on Intelligent user interfaces (IUI '08), 13 - 16 January 2008, Canary Islands, Spain, pp.413-416.


FIGURES

Fig. 1 Choice of Constraints for each CSk and SOURCES

Fig. 2 Proposed Generic Model (The Reasoning Process with Semantic Overlapping)
RAISING AWARENESS OF DENTAL PAIN THROUGH TWEETS

Radmila Juric¹, Inhwa Kim², Ana Sinkovic³, Igor Tesanovic⁴
¹Faculy of Science and Technology, University of Westminster, London, UK
²Samsung SDS Europe Ltd, Eastworth House, Chertsey, KT16 8SH, UK
³Dental Practice Gornja Stubica, Zagreb, Croatia
⁴Metronet d.o.o. Zagreb, Croatia

ABSTRACT

In this paper we give the result of our qualitative analysis of tweets which focus on dental pain for the purpose of finding if a) Twitter is being used as a platform for disseminating information about dental pain amongst sufferers and dental professionals and b) we can raise awareness about the danger of dental pain amongst the wider population by tweeting and re-tweeting, i.e. by creating real time information. Our study showed two distinctive results. Dental pain sufferers are very active in their attempts to express distress caused by dental pain and their inability to cope with it. It is disappointing that there are very small numbers of (re)tweets where dental professionals engage and raise awareness about the danger of dental pain. This research has also found that we can use analytic tools run upon tweets in order to detect and engage in health-oriented question-and-answer tasks on Twitter and think about new ways of creating probabilistic topic modeling from tweets in future.

INTRODUCTION

Online services obtained through social media and general purpose search engines have become an inseparable part of our everyday lives. It is no surprise that information related to healthcare is massively searched across the Internet (Hawn, 2009) and, at the same time, widely shared through special social networks (SN) which connect patients and healthcare professionals (Fox and Jones, 2009; Grajales III, et al., 2014; Greene, et al., 2011). The spread of micro-blogging, for the purpose of disseminating and consequently gathering and analyzing information available on blogging platforms and Twitter (Baumann, 2009; George and Dellasega, 2011; Rice, 2006) has also revealed how much we share health information. The new world of medical tweeting is our reality (Shariff and Desai, 2012); the analysis of the State Health Department tweets (Thackeray, et al., 2013) showed that the majority of tweets were about personal health and not about a state health department as an organization, which is consistent with other studies (Dumbrell and Steele, 2013; Neiger, et al., 2013). We should agree that social media are reshaping Healthcare (Eytan, 2008; Hawn, 2009) and may create happier patients, cut costs and improve quality of healthcare data through stronger multiple online communications. It also has an impact on Public Health (Chou, et al., 2013; Dredze, 2012; Thackeray, et al., 2012). Analyzing the content of social media is becoming prevalent in the Public Heath domain (Paul and Dredze, 2011). It ranges from creating public health applications based on Geo-location systems in order to analyze messages available on Twitter (Dredze, et al., 2013) to defining frameworks which can detect public health trends with Twitter (Parker, et al., 2013) and public health concerns using Twitter sentiment classification (Ji, et al., 2013).

In this paper we focus on the dissemination of data related to dental pain, which is available on Twitter, for two reasons:

1) There have been quite a few successful attempts to analyze twitter data for the purpose of digital surveillance and tracking of infectious diseases (Culotta, 2010; Lamb, et al., 2013) which encouraged us to believe that Twitter can be useful in addressing health-oriented question-and-answer tasks (Paul, et al., 2011);

2) Dental pain is very often neglected in analysis of problems associated with pain in general medicine. We do not have too many published papers which focus on orofacial pain, but this type of pain has a significant impact on the quality of life and deserves attention. The work of (Heaivilin, et al., 2011) has encouraged us to look again at Twitter data on dental pain, by using software analytics tools available in 2014, which are suitable for extracting relevant Twitter data and perform a new analysis of them. We can therefore achieve two things. We can contribute towards health-oriented question-and-answer tasks from 1) above, which focuses on dental pain, and pave the way towards new types of probabilistic topic modeling (Brown, et al., 2011) from Twitter data.

In this research we expected to get answers to quite a few obvious questions: are people tweeting about dental pain? If so, do they tweet because they need advice or expect something else from re-tweets? Do they tweet because they wish to describe what is happening to each individual at the moment when he/she tweets? Is the
dental pain so specific that it creates a need for a quick fix through the Tweeter and eliminates visits to dentists? Do people ask questions related to dental pain on Twitter?

We know that Twitter has been very successful in tracking various public opinions and that it encourages the development of cost-effective health monitoring approaches. However, in the case of addressing dental pain, Twitter data may be able to answer the following research questions in order to address 1) and 2) above:

RQ1: Is Twitter, as a micro-blogging environment, used for reporting problems related to dental pain and at the same time enabling individuals to distinguish between orofacial and all other pains in the head?

RQ2: Do people tweet only about PAIN, i.e. do they include in their tweets “terms” which describe dental pain in detail and problems which cause it?

RQ3: Do tweets ask for advice and if so, which advice is given in tweets and re-tweets? Do tweets ask questions on dental pain and if so, which questions are prevalent?

RQ4: Can we find tweets with professional advice for dental pain and if so which types of advice are prevalent?

In order to address these research questions we have adopted the following methodology. We primarily conducted experiments which extracted tweets in order to answer RQs. We performed a qualitative analysis upon the results of experiments and therefore the choice of experiments was dictated by the research questions RQ1-RQ4. Furthermore, the extraction of relevant twitter data was based on the criteria:

a) The availability of software tools which could extract tweets suitable for each experiment;

b) The “search words” defined by a dental healthcare professional which fit each experiment and can be used in software tools from a).

The experiments which extracted twitter data were conducted during 72 hours in March 2014. A software tool which proved to be efficient in extracting tweets was Tweepy (http://www.tweepy.org/). We exported tweets extracted with the tool into spreadsheets and performed a quantitative analysis of their content to answer RQs.

The paper is organized as follows. In the next section we explain the motivation for this research and analyze a few papers which may belong to our related work. In the Data and Methods section we describe our experiments and show the way they were conducted. In the Result section we give the results of our qualitative analysis and in Conclusions we assess if and how our research questions RQ1-RQ5 have been answered. We also look at the possibility of addressing issues outlined in 1) and 2) above in future work.

RELATED WORK

We have been motivated to conduct this research after reading the paper on public surveillance of dental pain via Twitter (Heavilin et al., 2011). The authors used their own way of coding tweets in order to categorize them after they have been selected and examined. The coding is very impressive and may be reusable in future studies, particularly if there will be more work which will address health-oriented question-and-answer tasks (Paul et al., 2011) as outlined in 1).

However, we wanted to examine tweets on dental pain completely differently. Firstly we decided to use modern software tools which can extract (i) relevant tweets in a certain period of time, i.e. during 12 hour time-slots and (ii) according to criteria defined by a dental professional. In other words, we extracted tweets on dental pain which appeared in the specified time slot (in real-time) and then analyzed them manually. Our selection of tweets through Tweepy depends on both: selection criteria specified by the dental professionals and research questions we must answer. We also wanted to pave the way for a possible automated extraction of a “dental topic” which may be tweeted (see a) in Introduction) and therefore we were more interested in defining an ad-hoc selection criterion for the extraction/analysis of tweets.

We should be able to compare our results with those of (Heavilin et al., 2011) and debate them further.

There are a few other interesting papers which have been frequently cited. They all show what power twitter data may have in supporting health care and public health in general. We learned from all of them, but could not use their ideas in our research for various reasons. In Lamb et al. (2013) Twitter data is used to track flu infections, which is feasible through locations available in tweets. However they improve previous attempts of flu surveillance by distinguishing tweets which give facts about the spread of flu from tweets which illustrate only concerns or carry information about “flu” not directly relevant for surveillance. Similar work from (Lee, et al., 2013) uses spatial, temporal and text mining on Twitter data for real time surveillance of flu. A slightly older work of (Culotta, 2010) was motivated by Google’s flu trend service and tried to see if there is a correlation between messages on Twitter and influence statistics reported by the Center for Disease Control and Prevention in the US. In (Scanfeld, et al., 2010) twitter is used for dissemination of information on evidence of misunderstanding or misuse of antibiotics, in (Trigo, et al. (2013) it has been used together with HL7 for cardiovascular patient follow up and in (Park, et al., 2014) Twitter was used for collecting patient reported outcomes in patients with celiac disease.

DATA AND METHODS

Data: Search Words for Dental Pain

A key challenge in this research was to define exactly which words in tweets would become “search words”
relevant for the extraction of tweets on dental pain. We firstly asked a dental professional to define exactly what would be of interest to patients and dental professionals in order to address 1) and 2) from the introduction.

Dental pain is essential in diagnosis of dental problems and as such it plays an important role as a warning signal of serious problems behind it. Categories of dental pain depend on its duration and intensity, which can be (in) tolerable according to the patient’s own perception. The longer the pain the worse outcome and even if it stops it does not mean that the problem is “solved”. It is often the case that the patient should be treated immediately when the long term dental pain stops, contrary to the belief of frightened patients who have suffered long term dental pain. Furthermore, not all dental pain is related to dental problems, and dental pain might be confused with some other pain in head. Patents who fear dentists tend not to suspect having dental pain. They would rather see any other medical professional including GPs than their own dentist.

Raising awareness of these issues through tweeting would be beneficial for both patients and dental professionals and “search words” used in the selection of tweets for our analysis should also answer our RQ1-RQ4. We firstly look at all tweets and re-tweets which are given below. Tweets for our analysis should also answer our RQ1-RQ4. The extraction of professionals and “search words” used in the selection of would be beneficial for both patients and dental professionals and “search words” used in the selection of tweets for our analysis should also answer our RQ1-RQ4. We created 5 different groups of “search words” which are given below.

1) We firstly look at all tweets and re-tweets which simply mention words “toothache” or “tooth ache”. Tweets which contain these two words would give us a partial answer to RQ1 and RQ2.

2) The second set of search words again included “toothache” or “tooth ache” but we added numerous other types of commonly used words by patents which could describe the dental pain better, such as “sore gums, tooth decay, infected gums, sensitive teeth, sensitive gums, gingivitis, bad tooth nerve, mouth ulcer, abscess, receding gums, chipped tooth, broken tooth, growing wisdom tooth, wisdom tooth hurts, tooth pain, painful tooth”. The tweets containing these words could help us to give more precise answers to RQ1 and RQ2.

3) The third set of search words again include “toothache” or “tooth ache” but we added a list of words which might have been used by dental professionals in order to describe dental pain and its sources. The are “odontalgia, dentalgia, odontodynia, odontogenic pain, non-dental diseases, inflammation of the pulp, dental caries, tooth decay, dental trauma, denit hypersensitivity, gingivitis, gum infection, trenchmouth, pulpitis ,pulp necrosis, death of the pulp, periapontitis, abscesses, periodontitis, gum disease, maxillary sinusitis, angina pectoris”.

4) The fourth set of search words again included “toothache” or “tooth ache” with added words which describe problems which are NOT related to dental problems, but might be confused with them. They are “sinusitis, angina pectoris, migraine, muscle pain, neuralgia, herpes, headache, bruxism”. This was important because it might signal that there is advice on Twitter where dental professionals or patients who have experienced similar problems might have raised awareness on the complexity and role of dental pain.

5) The last search words focus solely on the possibility of finding advice for dental pain in (re) tweets. We again used “toothache” or “tooth ache” and added the following words: “call dentist, make appointment, cold water in mouth, nonsteroid anti-inflammatory drugs, aspirin, ibuprofen, naproxen, topical treatments, toothpaste for sensitive teeth, protective varnishes, dental floss, rinse salt water, rinse medicated mouth wash, antiseptic solution, painkiller, antibiotics, tooth extraction, oral hygiene, local sedative dressing” This was important for answering RQ4 and finding out the extent to which professionals are involved in tweeting on dental pain and why.

Method
The method we used consists of 5 experiments which selected tweets in 12 hour time-slots and according to “search words” from 1-5 from the previous section. The selected tweets were exported in separate spreadsheet documents and then manually analyzed in order to answer research questions.

We used search words connected with “OR” logical operator in Tweepy because we wanted to collect all possible tweets which could help in answering RQs. This means that we had a duplication of tweets in some experiments. However, to eliminate a surplus of information we used search mechanisms within each spreadsheet, which was sometimes performed automatically in order to interpret findings efficiently. It was expected that experiments with numerous “search words” connected with OR logical operator, added to our “toothache” and “tooth ache” words would extract a significant number of tweets.

The results of our manual inspection of selected tweets for each experiment as a part of our qualitative analysis was then summarized in short texts which also highlighted whether our RQ1-RQ4 had been answered.

RESULTS AND DISCUSSIONS
In this section we describe each experiment separately by repeating which “search words” are used and summarizing results of our analysis upon selected tweets.

Experiment 1
We extracted 730 tweets and re-tweets using search words as in 1) from the Methods section. We summarize our analysis in the following bullets:

- The most striking outcome after reading all selected tweets was the variety of ways people describe dental pain and the strength of words used. If we ignore a huge amount of unpleasant wording and tweets which made no sense, because people who tweeted were in pain, then the picture of the intensity of pain and the way it affects someone’s life was vivid and powerful.
- 176 tweets were either very long or contained strong words which described the pain, compared it with other pains or describe desperate thoughts a toothache sufferer may have.
- There was one re-tweet which appeared numerous times in different posts, which says: “in 1954 a man committed a suicide by jumping from the Golden Gate Bridge leaving a suicide note saying: absolutely no reason except have a toothache”.
- We did not expect to find any advice or statement of action in re-tweets, but we have managed to collect a few of them as tweets. Re-tweets either confirmed the misery of dental pain or taught a lesson, such as “even the sweetest things on Earth hurt”.
- Tweets (not re-tweets) which give advice were only few and they were created by nonprofessionals. We read advice to take “simple painkillers” to “use cold water to ease the pain”.
- We found 26 tweets which directed us to links / websites where people can get advice on how to manage tooth ache. We could not see if the person(s) who tweeted these links was(were) actually owner(s) of the tweeted websites.
- There were no re-tweets in the first experiment which could have followed one particular tweet where a toothache sufferer said “I had toothache for one month”! In other words no-one found it necessary to re-tweet and tell the person how dangerous long term toothache can be.
- One tweet was quite clear and asked: “how is a toothache less medical than headache”.

Experiment 2

In this experiment we use search words as in 2) from the previous section. We expected to collect tweets which may have had more special words in explaining dental pain and possibly revealing its causes. In other words we ignored tweets which had content similar to the one from experiment 1 and looked for tweets and re-tweets which may give a better explanation of the pain and its reasons. These explanations might have come from patients who have already had professional help or from dental professionals themselves. There were 1931 tweets extracted in this experiment but the results of the analysis were disappointing:

- Only a few tweets (21) used words like: wisdom tooth, tooth decay, gums, sensitive teeth and chipped tooth in the description of the situation they wanted to tweet about. Not all of them were related directly to dental pain.
- We have found a few QUESTIONS people asked through their tweets such as “How do I stop wisdom tooth pain” or “Tooth is trying to abscess. Advice anyone?” or “Are you combining pain killers with anti inflams?” Their number was very low (14).
- There was more advice collected in the tweets of this experiment (164) than in the previous one. We did not primarily want to look at the presence of advice, but they appear because the “search words” for experiment 2 included terminology which may do both: describe dental pain better and include its reasons, and enrich advice given in tweets and re-tweets. None of the advice found in experiment 2 was in re-tweets, i.e. they are not re-tweets related to someone else’s is plea for help. They usually give links which takes you to “advice” (usually an URL) in order to address a variety of dental problems described through our “search words”.

Experiment 3

In this experiment 998 tweets were extracted with search words defined in 3) from the previous section. We expected to collect tweets which might have been used by dental professionals in order to describe dental pain, its sources and advice, or tweets created by toothache sufferers in order to obtain advice from professionals.

- The results showed that 187 out of 998 tweets are advice related. More than half (115 out of 187) tweets were added through retweets (RT) instead of direct update. A few sentences were repeating in tweets (e.g. @Fact: Kissing helps to prevent tooth decay, @DMartestQB: Semen actually fights against tooth decay).
- Only 3 out of 998 tweets are update questions, which means that individuals/sufferers may not be using Tweeter for the purpose of getting a response to or advice on their questions.
- 21 out of 998 tweets included professional words. This might indicate that tweeting is being used by ordinary people rather than professionals.

Experiment 4

In this experiment we collected 48,444 tweets according to “search words” for experiment 4. This was expected because the search words covered such a range of health problems which are bound to be in the focus of tweeting. Obviously we examined such a vast number of tweets through automatic search of keywords in each line of the spreadsheet, populated with Tweepy results. However, at this stage in our experiments we did not have
expected as before and therefore it was not a big surprise that:

- There was not a single tweet which connected toothache with any other “search words”
- The word headache appeared in one tweet only (which was found in all previous experiments) where a person describe a misery of having “toothache, headache, cold,…”
- The problem of bruxism appeared more frequently in tweets than any other search word, but it was not related directly to dental pain. There were quite a few tweets with advice on what to do and how to treat bruxism because of its impact on general health, but none of them explained that bruxism may be mistaken for dental pain.

Experiment 5

In this experiment we focus on “expected” advice one may give in a (re)tweet as the answer to dental pain. As in the previous experiments, we collected a greater number of tweets (7822). At the same time we tried to run a separate set of small experiments where we paired one or two search words with either “toothache” or “tooth ache” and connected them with logical AND operator when running Tweepy. We found no selected tweets in the first few cases which signaled that people simply might not (re)tweet in order to give advice. Our results are:

- Not all “search words” appeared in the extracted tweets.
- There were no tweets which combined any of our search words with “toothache” and “tooth ache”.

CONCLUSIONS

One of our goals in this study was to examine if we can find quickly whether the twitter community extensively communicates experiences of dental pain in real time and whether we can efficiently search twitter data using simple software tools. The answer to both of these questions is YES. Our experiments also show that the choice of search words may correspond to question we wish to answer and can be easily incorporated in tools as simple as Tweepy. Therefore we can carry on with such studies and involve more dental professionals in order to either change “search words” or motivate them to get involved in tweeting if we wish to raise awareness of how dangerous dental pain can be. The compactness of the environments where we create tweets and their presence on all our mobile phones makes the Tweeter really a prefect medium for an efficient and quick exchange of valuable data on dental pain.

Our research questions have been partially answered. Results of Experiment 1 really justify this study and proved that Twitter, as a micro-blogging environment, is used for reporting extensively about problems related to dental pain. However, we have found little evidence in experiment 1 that individuals distinguish between orofacial and all other pains in the head, which was also the result of experiment 2. However, the same experiment has revealed that search words which are used for getting tweets which distinguish orofacial pains actually selected the tweets which either post a question or advice. Therefore the answer to RQ2 is NO. If tweets mention “terms” which describe dental pain in detail and problems which cause it, then they are always part of advice or questions posted within the tweet. Consequently, the answer to the first part RQ3 is YES, but we could not find re-tweets which give advice and we could not find any prevalent question on dental pain in Experiment 2.

Experiments 3,4 and 5 enrich answers to RQ3 and RQ4 . Experiment 5 shows that dental professionals are NOT tweeting at all, in terms of giving advice, and therefore these results underpin our expectations that the answer to the first part of RQ4 is NO. Experiment 4 confirms that there are no re-tweets which could warn dental pain sufferers about misinterpreting dental pain and (dis)connecting them with other pains in head, which can be dangerous. Therefore the answer to the second part of RQ4 is that there are tweets which may give professional advice on various pains in head, but this advice does NOT connect/disconnect such pains with dental pain.

The results of our study partially overlap with the findings from (Heaivilin et al., 2011). Being a part of the global community and creating real-time data, Twitter platforms are really a valuable source of information for numerous monitoring in our society. From this point of view our results which showed how distressful and unbearable dental pain can be, confirm what was discovered in 2010, using a completely different methods of collecting and analyzing Twitter data on dental pain. In addition real time data on dental pain generates very accurate expressions and descriptions of experiences of dental pain sufferers, which cannot be obtained by any other means of collecting data. Consequently we assume that our twitter posts showed the high levels of distress as mentioned in (Heaivilin et al., 2011). However, we have discovered something which contradicts their findings. In our study there is very little to support their view that there exist sufficient self-help strategies in managing dental pain. We discovered much more prevalent directions towards tweets which contain links/URLs for either self-help or connecting sufferers with dental professionals who can help. (Heaivilin et al., 2011) categorize these tweets as Spam, but their prevalence was much higher in our study and as such indicates how the use of the Twitter platform has changed. We agree that these tweets target potential dental pain sufferers because they do not appear as particular re-tweets, but we should advocate them if we aim to raise awareness that dental pain should not be ignored at all. If dental professionals are too busy or not motivated to re-tweet on “toothache” then Twitter might be an ideal platform to to pass the
message to toothache sufferers: “dental pain may be very dangerous”.

Our future work firstly includes fully addressing 2) from the Introduction. We can use semantic reasoning in order to learn about a topic on dental pain which is tweeted in real time. This would help us to avoid being dependent on search words and might address the limitations of our method of extracting relevant tweets. Secondly, results of running the Tweepy tool according to our search words upon real time tweets, could be directly used in such semantic reasoning and would result in real time surveillance available for dental professionals, which is ideal for addressing health oriented question-and-answer tasks, as mentioned in 1) from the introduction. Dental professionals do not re-tweet on tooth ache: we hope that this paper might change their perception on what they may achieve with simple tweeting using their hand held devices. Even if dental professionals do not tweet in future, we could still use the semantic reasoning in order to address “dental pain” issues available on Twitter in real time. This would mean that the reasoning would produce advice from health professionals, triggered by the detection of tweets on dental pain, which may be used in environments outside Twitter.

REFERENCES


Hawn, C., 2009, "Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care,” Health Affairs, Vol. 28, pp. 361-368.


Paul, M. J., and Dredze, M., 2011, "You are what you Tweet: Analyzing Twitter for public health,” ICWSM.


THE REASONING PROCESSES FOR SELECTING TEACHING PRACTICES FOR STUDENTS WITH IMPAIRMENTS

Eiman Almami, Radmila Juric, Mohammed Zaki Ahmed
School of Computing and Mathematics, Plymouth University, PL4 8AA, Plymouth, UK

ABSTRACT

In this paper we show two SWRL enabled OWL models which allow reasoning for the purpose of determining which teaching practices are suitable for students with various impairments, when trying to achieve learning goals. We use the same basic OWL model which describes the semantic of the environment where teaching of pupils with impairments takes place. However, the models differ because they have different ways of creating constraints and the reasoning process upon the OWL ontological classes and their individuals. They may also have different implications on applications built upon these models. The paper discusses these differences and highlights their advantages and limitations.

Introduction

This paper is a continuation of our work on selecting the most suitable teaching practices, including instructional (Van Merriënboer, 1997; Reigeluth, 1999; Meisel, et al., 2003; Cheung, et al., 2010; Sicilia, et al., 2011) for students with various impairments, who would like to achieve a particular learning goal (Almami and Juric, 2011a; Almami and Juric, 2011b; Almami, et al., 2012; Almami, et al., 2013). We have shown in previous publications that the technologies from the Semantic Web Technology stack can be used for decision making, through reasoning, on the best possible teaching practices in various circumstances. Consequently, we created various SWRL enabled OWL ontologies, which

(a) store semantics of learning environments and
(b) enable reasoning upon their main concepts in order to make decisions on best teaching practice.

All our OWL models focused on a particular question (i.e. a competency question in OWL terms) which was supposed to be answered through the reasoning. However, none of them defined exactly which OWL design principles need to be followed in order to achieve (a) and (b) above. OWL is a powerful language and SWRL enabled OWL ontologies provide numerous possibilities for creating reasoning which support decision making in environments described through OWL concepts (Kataria, et al., 2008; Shojanoori, et al., 2012; Almami, et al., 2012; Shojanoori and Juric, 2013; Shojanoori, 2013; Almarri and Juric, 2013; Almarri, et al., 2013; Shojanoori, et al., 2014). Whether we use solely OWL hierarchies for expressing the semantics of learning environments and secure decision making through the reasoning or rely more on OWL constraints for the same purpose, we have been able to claim that we made “semantic decisions” on “which teaching practices can be used in situations where pupils with various impairments are trying to achieve learning goal(s)”.

In this paper we show two different OWL models which secure reasoning upon their concepts in order to answer the simple competency question above. The reader should be able to examine the differences in both OWL models and the reasoning processes in order to evaluate which model is:

(i) more suitable/flexible for the implementation of the reasoning and
(ii) easier to re-use and apply in order to answer some other competency questions.

This paper is organized as follows. In the next section we define two distinctive modeling principles. In each design we show:

a) Ontological OWL classes and Constraints
b) The Reasoning Process and SWRL rules and
c) Screenshots of all reasoning results.

Each of these subsections has Figures and Tables which are explained through the adjacent text. In Conclusions we evaluate our designs by looking at the way they address (i) and (ii) above. We also look at their applicability in potential computational model which can be used in any situations where instructional practices are dynamically generated in order to address needs of pupils with impairments.

THE MODELLING PRINCIPLES

Design 1: Ontological Classes and Constraints

In Figure 1 we give a set of main ontological classes created by the Protégé tool. Some of them are self-explanatory: PERSON, IMPAIRMENT, GOAL and
PRACTICE store individuals which describe the semantics relevant to people, impairments they may have, learning goals they wish to achieve and teaching practices essential for achieving such goals. These four classes have origins in the competency question from the introduction. However, they may fit in any other competency question where relationship between people’s impairments, learning goals and teaching practices are important. Table 1 gives a selection of OWL individuals which may populate these four classes.

OWL classes which store the results of reasoning, in order to answer the competency question, are named accordingly: PERSON_IMPAIRMENT class stores the semantic which describes which impairment(s) a particular person may have; IMPAIRMENT_PRACTICE stores the semantic which describes which practice(s) is related to a particular impairment and GOAL_PRACTICE stores the semantic which describes which practice(s) are important for achieving a particular learning goal.

In order to secure reasoning in general, we need to define constraints upon OWL classes. Tables 2 and 3 shows examples of constraints we use in Design 1. They are all object properties between two OWL classes, but defined upon their individuals, which participate in the relationship defined through the property. For example, from Table 2 we can see which practices are used for which impairment and which practices are good for which learning goal. Table 3 focuses on describing person’s impairments and learning goals, i.e. it shows impairment(s) for a particular person and which learning goals he/she wishes to achieve.

Design 1: The Reasoning Process

Figure 2 shows the reasoning process where all classes, individuals and object properties are used in order to answer the competency question. The diagrams’ red lines are object properties, and blue lines shows the reasoning process which moves individuals from IMPAIRMENT to PERSON_IMPAIRMENT class, individuals from PRACTICE to IMPAIRMENT_PRACTICE class and individuals from IMPAIRMENT_PRACTICE to GOAL_PRACTICE class. The reader should read the diagram form Figure 2 together with the content of SWRL rules which enable the moving of these individuals. These three SWRL rules, which are denoted as black lines in Figure 2, are given in Figures 3, 5 and 7. Their results are in screenshots given in Figures 4, 6 and 8.

Design 1: The Results of Reasoning

The final result of reasoning (“which practices…..”) is available as a set of individuals in the GOAL_PRACTICE class, available in Figure 8: a pupil who has autism, shows a serious lack of organizational skills and wishes to learn colors (see Table 2), will learn them if the teacher uses “practices” listed as individuals of the GOAL_PRACTICE class in Figure 8.

However, these results are achieved through three different rules: Rules 1, 2 and 3, available in Figures 3, 5 and 7. They obviously show that in this reasoning process we “reason upon the result of reasoning”. This is why we used colors in Figure 2: the green color denotes the main ontological classes and the orange color denotes classes where we store the results of reasoning: PERSON_IMPAIRMENT, IMPAIRMENT_PRACTICE and GOAL_PRACTICE. These “orange” classes are populated after running a particular SWRL rule and then used in the next rule (see “black lines” in Figure 2 which show classes involved in reasoning for each particular rule).

Design 2: Ontological Classes and Constraints

For design 2 we use the same competency question given in the introduction and performed reasoning in order to answer it. However, we change the ontological model, constraints and reasoning in order to achieve the same goal as in Design 1.

Figure 9 shows the main ontological classes. They are identical to classes from Figure 1, but without classes (PERSON_IMPAIRMENT, IMPAIRMENT_PRACTICE and GOAL_PRACTICE) where we stored results of reasoning performed in Design 1.

Tables 4 and 5 give another set of constraints defined upon classes from Figure 9 and their individuals. The difference between these two tables and the tables which describe object properties in Design 1 are obvious: we have an additional object property “requires” (see Table 4) which is expected to connect individuals of the PERSON and PRACTCE classes. All object properties in Design 1 are asserted, which means that they are defined in advance in order to grasp the semantics of the environment where the competency question can be answered. In design 2 the “requires” object property will have to be inferred for two reasons:

(I) We do not have a class in the OWL model which would store the results of reasoning and therefore no individuals of the PRACTICE class will be moved.

(II) The result of reasoning should be an object property which will connect person (his/her impairments and learning goals) with practices which would secure the achieving person’s learning goals.

Design 2: The Reasoning Process

Figure 10 shows the reasoning process for Design 2. It is obvious that object properties is-for, address, wishes-to-achieve and has have been asserted and “requires” property will have to be inferred. Therefore Figure 10 should be read in conjunction with Tables 4 and 5. Consequently, there will be only ONE SWRL rule, given
in Figure 11 which shows the exact inference of the object property.

**Design 2: The Results of Reasoning**

In Design 2 we are in different situation in terms of finding answers to the competency question. Firstly, we cannot simply retrieve individuals of the OWL class which might have stored the result of reasoning because we did not move any individual across the OWL model in order to answer the competency question. Secondly, if a particular object property constraint has been inferred (“requires”), then we will have to find out which individuals it “connects” from the classes involved in the reasoning. Therefore Figure 12 shows how the property “requires” has been inferred. It shows that 4 practices are inferred for the pupils who have autism and wish to learn colors. The practices are self-explanatory in Figure 12.

**RELATED WORK**

It is difficult to find a related work which connects an ad-hoc creation of teaching practices for students with impairments and reasoning processes performed by SWRL in order to create dynamically these practices. There are so many different approaches which use computer systems and specific “computations” in order to create assisted environments which can help people with difference in learning. If we look at the deployment of ontologies for the same problem domain, modelling a knowledge base is the core of numerous research proposals. In other words, decision making upon the best possible practices in teaching and learning is done by reasoning upon the knowledge stored in such repositories. In (Chu and Chen, 2009), the authors model knowledge though ontologies in order to support for students with mild disability in learning mathematics. There are formal ontologies which are used for semantic description of metadata in environments which contain educational resources as in (Lama and Sanchez, 2009). However, the reader might be interested in looking at relatively old selection ontologies in teaching/learning from (Wilson, 2004) in order to grasp the prevalence of these ontologies for education purposes.

An interesting work from (Nganji, et al., 2011) uses Semantic Web Technologies to create assistive learning environments for disabled students. Their ADOOLES ontology categorizes some common types of disabilities, assistive mechanism in learning and student’s goals. The work does not detail how the knowledge has been modeled and which constraints are being used in order to provide correct semantics and reasoning. Therefore it was difficult to compare their ideas with the proposal in this paper. However the authors have made it clear that ADOOLES is a knowledge base that serves as a repository of meaningful information. In the architecture they propose, student’s request is received by a Web server which in turn invokes a reasoned to query ADOOLES. This might mean that our design 2 could use the infrastructure prepared for ADOOLES, but the OWL model and reasoning has to be replaced completely with our solution. However, our ad-hoc creation of object property in Design 2 would mean that we will NOT build a rich knowledge-base: we would delete the results of previous inferences (object properties in Design 2) for each particular situation we model in learning environment. This is the biggest difference between our OWL/SWRL enabled models and all other OWL models which build knowledgebase. (Sicilia, et al., 2010) attempt to model instructional-design theories with ontologies: they model collections of methods represented as a combination of rules and concept constraints that express the recommendation for the final arrangement of learning activities and resources. They use rules to verify the matching of semantics of the instructional theory with the description of the situation where the learning happens. The idea is similar to (Nganji, et al., 2011), but different to ours. We use ontological matching, but it is changeable from “situation-to-situation” and reflects upon the semantics of that situation. Our design 1 is heavily dependent on such matching in order to move individuals across the ontology, but in design 2 our matching results in an ad-hoc definition of constraints. Therefore it is very difficult to compare their ideas with our designs.

**CONCLUSIONS**

In this paper we have improved our previous OWL models which enabled reasoning in order to decide about the best possible teaching practices for achieving learning goals of pupils with various impairments. Our two design models are quite different from the previous SWRL enabled OWL ontologies. The change we illustrate in this paper is twofold:

a) We moved away from our initial idea of using data type properties in order to express the semantics of learning environments and

b) We give two different design principles which enable reasoning with SWRL upon OWL concepts.

Both designs give the same result. Our intentions were not to push forward designs form this paper and ignore previous solutions for determining the best possible teaching practices. However, we would like to draw reader’s attention to various possibilities when defining reasoning with SWRL enabled OWL ontologies. The OWL models (designs) from this paper may be able to resolve the issue of dynamically allocating teaching practices in any learning environment for two reasons:

- “requires” property can be inferred (design 2) and
• SWRL rules form design 1 can be created in advance and chosen when necessary.

Consequently, we have shown that the changes specified in a) and b) above are feasible and may bring advantages when performing SWRL enabled OWL modeling. We have lost a significant amount of hard coding in our SWRL rules given in this paper by removing data type properties from our OWL models in both designs. We have also offered a choice of OWL design models with different OWL concepts and the reasoning process.

When answering questions (i) and (ii) from the Introduction, we conclude that Design 1 might have been more flexible and reusable solution because it is based on the reasoning which moves OWL individuals. Design 2 has flexibility in inferring constraints, but we have no information yet (at the time of writing this paper) if property inference could affect the performance of software applications which host them.

It remains to be seen if, in the full scale software applications based on these reasoning models, their performance will be significantly dependent on the design model. We also need to investigate if the inference of object properties will be an obstacle or advantage when focusing on the reusability of software applications which are based on these two designs.

REFERENCES


47th Hawaii International Conference on System Sciences (HICSS). Waikoloa, Big Isla.


FIGURES AND TABLES

Fig. 1 Ontological classes for Design 1

Fig. 2 The Reasoning Process in Design 1
Fig. 8 Screenshot of SWRL Rule 3 Results in Design 1

Fig. 9 Ontological Classes for Design 2

Fig. 10 The Reasoning Process in Design 2

Fig. 11 Screenshot of SWRL Rule in Design 2

Fig. 12 Screenshot of SWRL Rule Results in Design 2
Table 1: Individuals of classes in Design 1

<table>
<thead>
<tr>
<th>CLASS</th>
<th>INDIVIDUALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSON</td>
<td>Hismi, Iman, Jhon</td>
</tr>
<tr>
<td>IMPAIRMENT</td>
<td>lack_of_organisation_skills, autism, asperger_syndrome</td>
</tr>
<tr>
<td>GOAL</td>
<td>learning_colours, learning_shapes, learning_social_interaction_skills, learning_categories, learning_numerals, learning_sequencing_and_order</td>
</tr>
<tr>
<td>PRACTICE</td>
<td>use_sorting_object, use_calenders, use_schedules, use_lists, use_checklists, use_basic_matching_for_colour, maintain_lists_of_assignments, place_pictures_on_containers, use_social_stories, use_colour_coding, use_rewards_and_motivation</td>
</tr>
</tbody>
</table>

Table 2: Object Properties for Design 1

<table>
<thead>
<tr>
<th>CLASS</th>
<th>INDIVIDUALS</th>
<th>ObjectProperty</th>
<th>INDIVIDUALS</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td>Domain Class</td>
<td></td>
<td>Range Class</td>
<td>Range</td>
</tr>
<tr>
<td>PRACTICE</td>
<td>use_basic_matching, use_sorting_object, use_colour_coding, use_rewards_and_motivation use_calenders use_schedules use_lists use_checklists use_rewards_and_motivation use_picture_cues_in_locker</td>
<td>difficulty_with_change_in_routine</td>
<td>Autism</td>
<td>IMPAIRMENT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>addresses</td>
<td></td>
<td>difficulty_reading_the_emotions_of_others asperger_syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Difficulty_understanding_complex_language</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lack_of_organisation_skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>language_and_communication_skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_colour</td>
<td>GOAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>is_for</td>
<td>learning_social_interaction_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_literacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_writing</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3 Object Properties for Design 1

<table>
<thead>
<tr>
<th>CLASS Domain</th>
<th>INDIVIDUALS Domain Class</th>
<th>ObjectProperty</th>
<th>INDIVIDUALS Range Class</th>
<th>CLASS Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSON</td>
<td>Hism</td>
<td>wishes_to_achieve</td>
<td>learning_shapes</td>
<td>GOAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_colour</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Iman</td>
<td></td>
<td>learning_social_interaction_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_literacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_writing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jack</td>
<td></td>
<td>has</td>
<td>IMPAIRMENT</td>
</tr>
<tr>
<td></td>
<td>Joe</td>
<td></td>
<td>difficulty_with_change_in_routine</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>difficulty_reading_the_emotions_of_others</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>asperger_syndrome</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teta</td>
<td></td>
<td>language_and_communication_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lack_of_organisation_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Difficulty_understanding_complex_language</td>
<td></td>
</tr>
</tbody>
</table>
Table 4 Object Properties for Design2

<table>
<thead>
<tr>
<th>CLASS Domain</th>
<th>INDIVIDUALS Domain Class</th>
<th>ObjectProperty</th>
<th>INDIVIDUALS Range Class</th>
<th>CLASS Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSON</td>
<td>Hism</td>
<td>wishes_to_achieve</td>
<td>learning_shapes</td>
<td>GOAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_colour</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Iman</td>
<td></td>
<td>learning_social_interaction_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jack</td>
<td>has</td>
<td>learning_literacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_writing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Joe</td>
<td>has</td>
<td>difficulty_with_change_in_routine</td>
<td>IMPAIRMENT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>difficulty_reading_the_emotions_of_others</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>asperger_syndrome</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>language_and_communication_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lack_of_organisation_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>difficulty_understanding_complex_language</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teta</td>
<td>requires</td>
<td>use_sorting_object</td>
<td>PRACTICE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_schedules</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_colour_coding</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>maintain_lists_of_assignments</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_checklists</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_rewards_and_motivation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_social_stories</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_calenders</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_basic_machine_for_colour</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>place_pictures_on_containers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>use_lists</td>
<td></td>
</tr>
<tr>
<td>CLASS Domain</td>
<td>INDIVIDUALS Domain Class</td>
<td>ObjectProperty</td>
<td>INDIVIDUALS Range Class</td>
<td>CLASS Range</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>-------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PRACTICE</td>
<td>use_sorting_object</td>
<td>is_for</td>
<td>autism</td>
<td>IMPAIRMENT</td>
</tr>
<tr>
<td></td>
<td>use_schedules</td>
<td></td>
<td>difficulty_with_change_in_routine</td>
<td></td>
</tr>
<tr>
<td></td>
<td>use_colour_coding</td>
<td></td>
<td>difficulty_reading_the_emotions_of_others</td>
<td></td>
</tr>
<tr>
<td></td>
<td>maintain_lists_of_assignments</td>
<td></td>
<td>asperger_syndrome</td>
<td></td>
</tr>
<tr>
<td></td>
<td>use_checklists</td>
<td></td>
<td>language_and_communication_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td>use_rewards_and_motivation</td>
<td></td>
<td>lack_of_organisation_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td>use_social_stories</td>
<td></td>
<td>Difficulty_understanding_complex_language</td>
<td>GOAL</td>
</tr>
<tr>
<td></td>
<td>use_calendars</td>
<td></td>
<td>learning_shapes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>use_basic_maching_for_colour</td>
<td></td>
<td>learning_colour</td>
<td></td>
</tr>
<tr>
<td></td>
<td>place_pictures_on_containers</td>
<td></td>
<td>learning_social_interaction_skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td>use_lists</td>
<td></td>
<td>learning_reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_literacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_writing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>learning_numeracy</td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT

According to the Korea Census Bureau, the number of elderly people above the age 65 will reach more than 10 millions in 2026. It means that more than 21% of the population in KOREA will reach over 65 years old. It becomes extremely important to take care of the elderly people for the case of health emergency, such as fall or heart attack. In this paper, a fall detection system for the elderly is developed under the USN environment. In order to detect the fall situation, 3-axis acceleration sensor (MMA7331) and slope detector are used in the system. The algorithm to decide the directions of the falls is proposed. The direction of the fall is decided by examining the slope of the Y and Z direction voltage measurements. It is shown that the proposed algorithm successfully detects the forward and backward fall with probability of 88% and 90%, respectively.

INTRODUCTION

In 2007, the Joint Commission on Accreditation of Healthcare Organization(JCAHO) reveals “reduce patient risk of harm resulting from falls” as the important sub-goal [1]. The elderly fall events have become the major public health issue in the world these days. The concerns of the elderly falls grow as the aging population increases. [2]. Falls have to account for a significant portion of injuries in hospitalized patients. Fall injuries in a hospital are often severe, due to the nature of the underlying medical condition [3]. According to the Korea Census Bureau, the number of elderly people above the age 65 will reach more than 10 millions in 2026. It means that more than 21% of the population in KOREA will reach over 65 years old. Therefore, fall prevention issues have become an important for the safety of the elderly in recent years all over the world including KOREA [4].

Wireless health monitoring in health care for the elderly is in great demand as the result of the elderly falls becomes fatal. Numerous researches have conducted to detect the elderly falls [5]-[7]. In order to detect the elderly fall in real time, Bluetooth module is used in [8]. A two-threshold MATLAB algorithm for fall detection method is experimented. The algorithm uses mainly triaxial gyroscope data measured from the waist to distinguish between fall, possible fall, and activity. The decision between fall and possible fall is done by the posture information from the waist-and-ankle-worn devices ten seconds after the fall impact [9]. A fall prediction algorithm that can predict whether the person will fall within one-walking-step is proposed in [10]. The fall prediction is different from the fall detection, and it is intended to predict a fall before it occurs. The neural network assisted algorithm uses accelerations and angular rates as inputs. There are numerous papers that the falling cases are determined by the measurements of the accelerometer [11]-[13]. Robust detection of events and activities, such as falling, sitting, and lying down is a key to a reliable activity monitoring system. In [14], a fall detection and activity classification system using wearable cameras. A fall detection algorithm employing histograms of edge orientations and strengths is presented. An optical flow-based method for activity classification is also introduced. An acoustic fall detection system that employed a microphone array and beam-forming to provide automatic fall detection is introduced. In order to overcome the difficulties in detecting the fall signal in environment where interference comes from the fall direction, two blind source separation methods for extracting the fall signal out of interferences are proposed. The fall classification is improved by the algorithm [15].

In this paper, fall detection algorithm that detects the actual fall is described. In the previous research, the movement detection of the normal sitting on a chair and standing up from a chair has shown in [16]. The electronic system that decides the fall detection is developed. The system includes a 3-axial acceleration detector and a slope detector. In order to decide two types of fall, the slopes of voltage measurements in Y and Z directions are observed. The Y-direction represents the front or back direction whereas Z-direction represents up or down direction. The proposed algorithm detects the fall as well as the direction of the fall such as forward fall and the backward fall. Also, the system can also measure the heart beat rates of the elderly that can be used in the future to enhance the detection probability of the fall.
FALL DETECTION SYSTEM

Hardware Description

The fall detection system consists of a wireless waist worn measurement unit with a tri-axial accelerometer and 2-directional slope detector. As shown in Fig. 1, the fall detector system is worn around the waist of the person. The system can measure the acceleration in 3 directions and the slope in 2 directions. The system has the capability of counting the number of walk, detecting the real fall and the possible direction of the fall such as the forward fall and the reverse fall. The location and the fall detection information is sent to the main computer by the Zigbee module. The location of the fall and the direction of the fall can be reported to the proper authority such as family member, nurse or the family doctor.

Fig. 1  The Fall Detection System

As shown in Fig. 2, the fall detection system hardware block is controlled by Atmega128. The fall detection system includes the accelerometer(MMA7260Q), heart rate sensor(RP-320) and Zigbee module(ZBS-200). There are two different voltage sources such as 5v and 3.3V. The MMA7260Q, RP-320 and ZBS-200 are powered from a 3.3V power source. In order to measure the acceleration related data, MMA7260Q from the Freescale Semiconductor is used. The MMA 7260Q accelerometer has signal conditioning, a 1-pole low pass filter, temperature compensation and g-select which allows for the selection among 4 sensitivities. The accelerometer reads the acceleration on 3-axis and sends the data to the Atmega128. In order to connect the Atmega128 to the Zigbee module, the UART interface is used. The Atmega128 coordinates all the modules in the fall system. The system decides whether a fall is detected or not. It also can decide the direction of the falls such as forward or backward fall.

Fig. 2 Block diagram of a fall detection system

Fall Detection Algorithm

The falling scenario is shown in Fig. 3. As explained before, the falling detection hardware is located at the waist. And the actual falling experiment is conducted on the bed for the safety. The X-direction corresponds to the left or right direction from the person who wears the hardware. The Y-direction corresponds to the front or back direction of the person. Finally, the Z-direction corresponds to the up or down direction of the person. The detection algorithm is shown in Fig. 4.

The fall detection algorithm works as follows. When the fall occurs, a sudden change of acceleration value occurs in 3 directions. The algorithm detects the sudden variations in the measured data in 3 directions. Then, the register in the accelerometer is turned on for 10 seconds to measure the variations in the measured voltages in the Y and Z directions. At this time, the algorithm decides the direction of the fall by looking at the slope of the measured voltages in the Y and Z directions.

Fig. 3  The Picture of the Forward Fall
The typical X, Y and Z direction acceleration values for the forward fall are shown in Fig. 4. The slope of the Y-direction is calculated by dividing the voltage difference by time difference. The slope of the forward fall Y-direction is calculated as +8.3. Similarly, the slope of the forward fall in Z-direction is calculated as -10.2. Note that the slope of the Y direction and Z direction in the forward fall are positive and negative, respectively. A similar result for the backward fall is shown in Fig. 5. Note that the slope of both Y-direction and Z-direction are negative. The slope of the Y and Z directions are calculated as –7.5 and -9.7, respectively.

As shown in Fig. 4 and Fig. 5, a sudden change in the measured voltage values in X, Y and Z directions. When the actual fall is detected, the algorithm is calculating the slope of the Y and Z measurements. The algorithm decides the direction of the fall as forward fall when the slope of the Y direction and Z direction are positive and negative, respectively. The algorithm decides the direction of the fall as backward fall when the slopes of both Y and Z are negative. For example, as shown in Fig. 4, the slope of the Y-direction for the forward fall is positive. The slope of the Z direction for the forward fall is negative. For the backward fall, both the slope of the Y and Z directions are negative as shown in Fig. 5.

MEASUREMENTS RESULTS

In order to measure the voltages in X, Y and Z direction, 3 different persons with different heights and weights are participated in the experiment. A total of 180 fall attempts are performed. A total of 90 forward fall attempts were performed with 30 attempts for each person. Similarly, 90 backward fall attempts were performed. As shown in Table 1, 79 forward fall attempts were correctly detected with 11 unsuccessful results for the forward fall. For the backward falls, 81 out of 90 falls were successfully detected with 9 unsuccessful readings. The detection rates of the forward and backward falls were 88% and 90%, respectively.

Table 1. Fall Detection Rate

<table>
<thead>
<tr>
<th>FALLS</th>
<th>NUMBER OF FALLS DETECTED</th>
<th>NUMBER OF FALLS NOT DETECTED</th>
<th>DETECTION RATE(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORWARD FALL</td>
<td>79</td>
<td>11</td>
<td>88%</td>
</tr>
<tr>
<td>BACKWARD FALL</td>
<td>81</td>
<td>9</td>
<td>90%</td>
</tr>
</tbody>
</table>
CONCLUSION

In this paper, a fall detection system that can measure 3-dimentional acceleration values is developed. The measured values are transmitted to the server computer through the USN. The main focus of this paper is to differentiate the forward and backward fall by measuring the slope of the Y and Z direction measured voltage values. The proposed algorithm showed 88% correct detection for the forward fall. Similarly, the algorithm showed 90% of detection probability for the backward fall. Since the developed system can measure the heart beat rate of the person, the heart beat rate can be used to collaborate with slope of the graph to improve the detection probability of the fall for the future research.

REFERENCES

AUTOMATIC PANNING SHOT GENERATOR USING TWO PHOTOS

Sehyun Park, Kyuman Jeong
School of Computer and Communication Engineering
Daegu University
Gyeongsan, Gyeongbuk, Korea

ABSTRACT

Panning in photography is known as the rotation of the camera horizontally. It refers to a technique whereby one follows a moving subject and takes a photo with a slow shutter speed. This creates a blurred background, while retaining sharpness in the subject. Panning shot is widely used in sports activities because it dramatically emphasizes the movement of the subject. However, it is not easy for amateur photographers to take panning shots. This paper represents a digital algorithm to automatically generate panning shots using two photographs taken consecutively in time. The presented algorithm makes even novice photographers take professional panning shots very easily.

INTRODUCTION

Sir John Herschel used the compound word "Photograph" for the first time in 1839 which is based on the Greek phos (light in English) and graph (drawing in English). So “photograph” means “drawing with light” and “photography” is known as the process of creating photographs (London et al, 2004). In (analog) photography, exposure is the amount of light per unit area reaching a film, as determined mainly by shutter speed and lens aperture.

Shutter speed is the amount of time which a camera’s shutter is open when taking a photograph. The more shutter speed a camera has, the more amount of light reaches the image sensor. In addition to the effect of exposure, the shutter speed affects the movement in photographs. Very short shutter speed freeze even fast moving objects at sporting events while relatively long shutter speeds blur moving subjects intentionally.

If a photographer freezes the movement of a fast moving subject using very short shutter speed, the subject can be considered stand still as shown in Fig. 1 (a). In contrast, a fast moving subject cannot be recognized at all because it is totally blurred if a very long shutter speed is used. Panning shot can be a solution to these problems.

Panning shot refers to a technique in photography to follow a moving subject and take a photograph with slow shutter speed. It can capture the main subject very sharply while the background blurry in the opposite direction of the subject’s movement as shown in Fig. 1 (b). As a result, panning shot dramatically emphasizes the subject’s movement, which is often used in sports photography such as car racing.

Even though panning shot is the optimal way of expressing a subject’s movement, it is not easy for amateur photographers to take plausible panning shot photographs. It is because it is difficult for amateur photographers to keep the subject in the same position of...
the frame during the relatively long exposure time.

In this paper, we present a digital algorithm for automatically generating panning shots from two photographs taken consecutively in time. Even novice photographers can take plausible panning shots very easily by the aid of the presented algorithm. The presented algorithm is consisted of three parts: object segmentation, motion estimation and motion blur.

OBJECT SEGMENTATION

The first step of the proposed algorithm is to segment main object from background image. Object segmentation is widely used to simplify the representation of an input image into a more segments. Users can extract relevant information of the given image.

Since we use two photographs taken consecutively in time as input images as shown in Fig. 2 (a) and (b), we can assume that the main object is moving temporally. Therefore, this problem can be formulated as the segmentation of moving objects. The segmentation of moving objects can be categorized as three approaches: the region-based techniques, the boundary-based techniques, and the combination of region-based and boundary-based techniques.

In this paper, we use a traditional region-based method based on background subtraction and frame difference techniques (Bai and Sapiro, 2009). The moving object can be discriminated from the background by applying the background and frame difference approaches. Even the traditional method works in our experimental setting because only the main object moves.

MOTION ESTIMATION

Motion estimation is used to determine motion vectors from adjacent images which describe the transformation from one 2D image to another. The motion vectors may relate to the whole or specific parts, such as rectangular blocks, arbitrary shaped patches or even per pixel. In this paper, we use optical flow to estimate motion (Beauchemin and Barron, 1995), where the vectors correspond to the perceived movement of pixels. We can estimate motion as shown in Fig. 2 (c).

MOTION BLUR

Motion blur in effect in this paper can be achieved by applying motion blur kernel only on background area. In theory, motion blurring can be formulated as a convolution of motion blur kernel (shown in Fig. 4) into an image. In panning shot, only background area must be blurred while the main object remains sharp as shown in Fig. 2 (d).

![Fig. 2 Result image of a moving car](image-url)
RESULTS AND DISCUSSIONS

In Fig. 2, we apply the proposed algorithm to a moving car which is almost invariant in shape between two input images. Fig. 3 shows another result of a man on a bicycle and the main object is relatively variant between two input photographs. Even though the main object is variant, Fig. 3 verifies that the result is still plausible. Since the proposed algorithm discriminates the main object from the background and applying blur kernel is simple in theory, the result of panning shot is very plausible theoretically and practically.

CONCLUSION

In this paper, we propose a new method for generating panning shot effect using two photographs taken consecutively in time. The proposed method consists of three parts: object segmentation, motion estimation and motion blur. Resulting panning shots are very plausible in theory and practice since regardless of the shape variance of the main object.

It is very difficult to take nice panning shots even for professional photographers. However, the proposed method enables even novice photographers to take plausible panning shots just by taking two photographs at the same spot.

This algorithm can be implemented on any open OS environments in the near future. In this case, users do not need to buy expensive DSLR cameras to take nice panning shots.

The proposed algorithm can be applied to the field of computational photography (Raskar et al, 2006)(Jeong, 2009). Further research must be done to extend the quality of results by applying recomposition (Banerjee and Evans, 2007), cropping (Chen et al, 2003)(Suh et al, 2003), and image quality evaluation (Tang et al, 2011).
ACKNOWLEDGEMENT

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2014.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2011-0010081).

REFERENCES


Raskar, R., Tumblin, J., Levoy, M., Nayer, S., 2006, “SIGGRAPH 2006 course notes on computational photography”, SIGGRAPH.

MALAYSIA’S HEALTH INFORMATION EXCHANGE (MYHIX):
DEPLOYMENT BARRIERS AND SOLUTIONS

Probir Kumar Banerjee*
(Corresponding Author)
Hamid Bagha*
*Swinburne University of Technology
Kuching, Sarawak, Malaysia

ABSTRACT
Policy makers, researchers, industry groups, and healthcare professionals identify health information exchange (HIE) as a solution to the problems of fragmented health information. The concept has gained traction and many countries have either implemented or are in the process of implementing HIE to transform health care delivery. Malaysia has already launched a HIE project called MyHix and plans to deploy it in the near future. However, deploying an HIE is not just a technology based initiative. There are many organisational and regulatory barriers, apart from technology, and many stakeholders whose needs must be planned for and addressed. We use an integrated perspective of the TOE, stakeholder theory and motivation/goal setting theories to identify the barriers to deployment of Malaysia’s HIE initiative (MyHix) and recommend strategies to address them.

INTRODUCTION
Quality healthcare needs collaborative efforts of teams of healthcare providers. For effective collaboration, teams need to share and have access to different kinds of healthcare related information. This becomes important since patients move from one location to another, and they may need treatment from different physicians and different healthcare facilities at different times. Thus a centralised location where teams (stakeholders) can upload their information and access it from all locations becomes important for quality health care. Inaccurate or incomplete patient related information could be dangerous for the patient (as well as the health care provider!!). The concept of an integrated health information exchange (HIE) has therefore gained traction and many countries have either implemented or plan to implement HIE. The objective of an HIE is to transform industrial age healthcare delivery (at physical locations with fragmented information) to digital age healthcare delivery (where the patient can get quality treatment, anywhere, anytime). Policy makers, researchers, industry groups, and healthcare professionals identify health information exchange (HIE) as a solution to the problems of fragmented healthcare information (Brailer 2005).

In Malaysia, the process of healthcare transformation started as early as 1997 with announcement of the telemedicine blueprint as follows:

“Telemedicine should be considered to have the broadest possible scope. It encompasses any health, health-support or governance service that can be provided via a multimedia network and a range of network-based or network-linked information and multimedia tools and technologies used by people and health professionals to access, manage and deliver healthcare.”

The Malaysian government firmly believes that “The future healthcare system will be supported and strengthened by telemedicine…. Through the seamless and ubiquitous availability of information and other services, telemedicine will dramatically reshape the delivery of healthcare. Information and other services will become more virtual, more distributed and more integrated, resulting in better, more timely and more efficient healthcare delivery.”

In keeping with its vision of healthcare transformation, the Malaysian Ministry of Health has initiated a series of measures. It proposed a health ICT framework to enable the transformation (refer to figure 1). In 2006, the Ministry of Health announced a plan for supporting health care decisions of caregivers through implementation of an integrated information service that would enable a secure and private lifetime health record (LHR) of each Malaysian to be stored and accessed. Concepts of a personalised lifetime health plan (PHLP) data, a health data warehouse, and a health information exchange portal (MyHIX) have been developed. Funding for MyHIX project was obtained in 2008 from MDec (A Malaysian body incorporated under the Companies Act of Malaysia, and owned and funded by the Government that promotes ICT usage in Malaysia). Standards for data exchange have also been established. Newspaper reports indicate that the Malaysian government planned to complete the development of the
technology infrastructure for linking all Malaysian hospitals and clinics by May 2014 (Azam Ahmad, April 17, 2013 publication in Star online) to facilitate deployment of Malaysia’s MyHIx.

However, while HIE is a technology based initiative, there are many organisational and regulatory barriers, apart from technology, that could affect deployment. While technical blue prints and ICT infrastructure and systems are important for development of an HIE, researchers mention that there are other organisational and environmental barriers to their adoption and usage (for example Jha et al., 2008; Vest and Gamm, 2010; Rossa et al., 2010). Environmental factors such as industry competition (Vest and Gamm, 2010) and lack of appropriate legislations, acts and arbitration procedures for dispute resolution (Banerjee and Bagha, 2014) also deter adoption of EHR by physicians. In addition, since there are many stakeholders in HIE, their needs must be addressed for their participation in HIE. Prior research on HIE deployment has, however, focused only on low adoption of electronic health records as a barrier. A holistic assessment of barriers to HIE deployment has not yet not been made. Therefore, we aim to answer the following research questions:

What are the different barriers to deployment of HIE and how do they influence motivation of different stakeholders to participate in Malaysia’s HIE?

What steps can be taken to address the barriers?

We use an integrated perspective of the technology, organisation and environmental (TOE) framework (Tornatzky and Fleischer 1990), and the stakeholder theory (Freeman, 1984, 2008) to answer the research questions.

In the next section, we justify the relevance of the above frameworks and theories for this study. Based on an integrated perspective of these theories and framework, we categorize the different barriers with evidence from past research and discuss how these barriers could affect Malaysia’s HIE implementation. We conclude with discussion of the contributions of this research and offer suggestions for future research.

LITERATURE REVIEW

The technology, organisation and environment (TOE) framework (Tornatzky and Fleischer 1990) has been used by researchers to study implementation of technological innovations in organizations such as open systems adoption (Chau and Tam 1997) and web-services adoption (Bolloju and Turban, 2007). The framework suggest that in order to understand adoption and diffusion of a technology based initiative, technological factors such as existing technology in an organization and new available technology available in the market, organisational factors such as resources, competence, experience, motivation, and environmental factors such as industry competition in the industry to which the organisation belongs, rules and regulation for using a new technology targeted by the organization, would all exert combined systematic influence on the organisation’s intent to adopt a new technology and to use it effectively. HIE being a new technology based initiative for health care service delivery, we believe tenets of the framework would help to provide a holistic perspective of categories of barriers and their underlying causes, and facilitate identification of measures that can be taken to address the barriers. Thus there is justification for using the RTOE framework for our inquiry.

Since implementation and usage of HIE involves several stakeholders, it is also important to analyse how the technology, organisational and regulatory barriers are perceived by different stakeholders and how the perceptions affect participation in HIE. Prior research mentions that perceptions of barriers are likely to differ depending on stakeholder goals and motivations (Fontaine et al, 2010; Banerjee 2014). An appropriate theory is therefore needed to determine what factors would lead to participation by stakeholders. We therefore use the stakeholder theory for insights. The stakeholder theory suggests that persons or groups associated with an organisation participate in organizational endeavours with a motive of deriving individual benefits while working for organisational goals. This theoretical premise is complemented by the goal setting theory (Locke and Bryan 1969) of motivation which mentions that individual goals prompt employee’s attention towards goal-relevant activities and away from goal-irrelevant activities (Shalley 1995; Latham et al 2002), and Vroom’s expectancy theory (Vroom 1964) of motivation which suggests that attitude (tendency to act in a specific way) depends on expectation of the act leading to a desired outcome and on the attractiveness of that outcome to the individual.

We therefore argue that motivation for stakeholder participation in HIE may be better explained and addressed with an integrated perspective of the TOE and the stakeholder theory. Based on this integrated theoretical lens, in the next section we review prior literature and identify the three types of barriers -
technology, organisation and regulatory— in the perspective of some of the key stakeholders—such as physicians, patients, suppliers of HIE related software and technology infrastructure, and the government.

**BARRIER TYPES AND STAKEHOLDER PERCEPTIONS OF BARRIERS**

**TECHNICAL BARRIERS**

*Physicians:* Physicians are not sure whether EHR systems connected to a HIE are secure enough for protection of their patient’s data. They are also concerned about security of the networking infrastructure for HIE. Since the fundamental building block of HIE is EHR, non-adoptions of EHR by physicians in small clinics and hospitals could derail any HIE initiative. Privacy issue remains one of the main concerns of the physicians which deter adoption of EHR systems (Farsi et al. 2006; Jha et al., 2009; Ludwick and Doucette 2009; Rao et al., 2011). Physicians also feel concerned about the lack of customisability of EHR software to suit their practices. They feel vanilla EHR systems implemented as it is would increase their work load. The government, on the other hand, feels that a tested EHR system, a secure HIE infrastructure, and appropriate access rights to data would be adequate to protect data on the HIE.

In the Malaysian context, there is evidence that EHR systems pose learning barriers to a majority of physicians (for example Banerjee and Bagha, 2014). They feel more comfortable with manual processes and in some cases, only use technology that they can manage and control. Thus a fundamental barrier to deployment of Malaysia’s HIE is the HIE technology non-adoptions by physicians.

*Patients:* If the patients are allowed access to their data on the HIE, it adds to the complexity. Older patients are not technically savvy. They would be difficult to train and they may not have the technical awareness and capability to understand the implications of their health data residing on a shared HIE. Even technically savvy younger generation may not fully understand the security features and the ability of the HIE infrastructure to keep their data secure. Thus the primary beneficiary of the HIE, the patient, may have privacy concerns that may not be addressed by training. They may choose to opt out of the system. They may also not be comfortable with physician’s who enter their health data on EHR systems if such data becomes sharable on the HIE.

*Suppliers of HIE technology:* Providers of EHR systems may have their own agenda for participating in HIE. They may not necessarily have the same level of concern with their systems as physicians who become directly responsible if the EHR systems and HIE interfaces fail to protect patient data. Physicians in small clinics and hospitals are also concerned about perceived ease of use and interoperability of the EHR with HIE (Vest and Gamm 2010). They mention lack of technical support by vendors of EHR system, training and system maintenance that prevents them from using these systems (Castillo et al. 2010). Supplier provided technical support may also not be sufficient to facilitate usage of the systems. Physicians feel concerned that training provided to them would only be geared towards the product functionality; it would not be geared towards the physician’s perceptions of support from the system - necessary to facilitate and develop clinical best practices. There is also evidence that support personnel are not completely aware of the healthcare context, which in turn causes difficulty in using the system in an efficient and effective manner (Ludwick and Doucette 2009). This also affects the system supplier’s ability to provide support for customisation of EHR systems, and do complementary changes that are necessary, especially for small to medium size care delivery organization which do not have sufficient skill and knowledge to plan and implement these complementary changes (Jha et al., 2009; Miller and Sim 2004).

In the Malaysian context, unlike the US and other developed countries, EHR systems are still not available in the market. Thus third-party supplier familiarity with EHR systems and their ability to support physicians is non-existent, which could thwart Malaysia’s effort towards HIE deployment.

*Government:* The government’s inability to provide the right telecommunication infrastructure for integrating small and private clinics with hospitals and the centralised HIE exchange, with sufficient bandwidth, could be a deterrent to HIE implementation. Telecommunication infrastructure in Malaysia, as yet, is not at par with the infrastructures in developed countries. Unless concerted efforts are made by the government, it could deter HIE deployment.

**ORGANISATIONAL BARRIERS**

*Physicians:* At the small clinic level, limited time, technical expertise, IT experience, and lack of funds are barriers in adoption of EHR by physicians. Increased workload and duplication of work is often mentioned as a deterrent that small clinics face in regard to EHR adoption (Randeree, 2007). Working with both old and new system for a certain period of time can increase the workload of physicians in small clinics. It forces them to track manual records of existing patients
and at the same time use the new system when consulting with new patients. According to Meade et al. (2009) lack of time is a significant barrier for small clinics in adopting EHR systems (Meade et al., 2009). Physicians mention they would have to spend more time for each patient consultation when using EHR, which would reduce the number of patients visited per day or increases the workday’s duration or both (Miller and Sim 2004; Jha et al., 2009). Since many physicians in small clinics have very little administrative support, they are occupied with their daily routine business processes and therefore do not have time to procure appropriate EHR system and analyse the consequences of transferring their business process to HER. Such initial time and associated cost thwart EHR adoption and utilization by small clinics, thereby hindering the deployment of HIE (Vest and Gamm 2009). These barriers are found to exist in the Malaysian context as well (Banerjee and Bagha, 2014).

Physicians in hospitals do not want to assume responsibility for privacy of patient data residing in the hospital server, and more so if it is on a shared HIE. This could lead to lax IT governance mechanisms that could be a negative outcome of HIE. In private hospitals, the interest of IT department, physicians and hospital management and promoters, if not balanced, could lead to problems in physician participation in HIE in the hospitals. Physicians may not feel comfortable assuming full responsibility of data security on the HIE when their hospital data servers are connected to the HIE. The first author conducted interviews in October 2013 with some physicians in a hospital in Kuala Lumpur and found corroboration to this effect.

Other organisational barriers such as financial returns also exist. For example, private sector hospitals feel that they may not get a good return on investment (ROI) from investing in the EHR technology and integrating it with the HIE (Vest and Gamm, 2010). Lack of organised training and support for stakeholders so that they appreciate the benefits of participating in an HIE, and lack of trust amongst stakeholders are also stated as major deterrents (Rossa et al, 2010).

Suppliers of HIE technology: Concerns about adequate technical support by vendors of EHR system, training and system maintenance are raised by physicians (Castillo, Martinez-Garcia and Pulido 2010). Thus software suppliers may not be able to address the concerns of physicians in regard to perceived ease of use and interoperability of the EHR with HIE (Vest and Gamm 2010). Physicians interviewed in Malaysia mentioned expressed similar concern; they were not sure about quality of support from EHR software providers. Thus efficacy of EHR software suppliers to provide adequate support for EHR systems is perceived as a barrier by physicians, which demotivates them from adopting and using EHR adoption (Banerjee and Bagha, 2014).

Software houses in Malaysia also do not see the benefit of developing their organizational infrastructure and putting in resources for developing EHR systems for Malaysian health care industry. They have limited resources and are keen on maintaining their ROI from existing products and services. This statement is based on evidence gathered from field interviews conducted by the first author in October 2013 with software providers in Malaysia. Thus there is no financial incentive for them to participate in a national level technical initiative such as MyHIX.

**Government:** At the government level, the move towards HIE entails restructuring of government departments involved in health care and redefining their roles and responsibilities. Governments across the world have entrusted responsibility to appropriate government departments for monitoring HIE deployment initiatives. In Malaysia, the government realises that existing organisational infrastructure may not be adequate for HIE. It has therefore been proactive in formulating a government level organisational infrastructure that would be conducive to implementation of an HIE. The proposed organisational infrastructure is shown in figure 1. However, whether the re-organised structure would ensure smooth co-ordination of all departments for realising the common objective of MyHIX deployment is debatable, because the departments may have conflicting agendas that may thwart time-bound deployment. According to a 2012 study cited by the Bipartisan Policy Center (BPC), the absence of efficient, uniform care co-ordination of stakeholders in the U.S. has cost the nation between $148 and $226 billion annually (McCann, 2012).

**REGULATORY BARRIERS**

**Patients:** Patients express concern about possible recourse to data privacy breaches in a networked HIE environment as more stakeholders gain access to patient data stored on HIE. One patient interviewed in Malaysia by the first author tied this issue to fraudulent use by insurance companies to deny genuine claim payments. There is anecdotal support for this concern. One patient was denied claim for tuberculosis treatment on the argument that tuberculosis comes from pre-existing health conditions which the patient should have reported when getting insurance cover. Though the patient was not aware of pre-existing conditions, the claim was refused on the grounds that
the patient’s health records indicated that he had been prescribed medication that indicated suspicions of tuberculosis. This information was obtained during interview with a physician in a hospital in Kuala Lumpur conducted by the first author in October 2013.

Data privacy issues are still not fully addressed by legislations. Cyberlaws are still not well framed. It can be seen therefore that without proper legislations that protect the rights of patients in a networked HIE environment, patient approval of sharable data on HIE may not be forthcoming.

Interestingly, there is also a positive effect from access to HIE by insurance providers. Patients would not be able to their previous medical conditions from health insurance providers for low premium medical insurance cover. They would also be deterred from lodging false medical claims. This would be also helpful for employers of patients if their insurance is covered by their employers. Thus there may be different underlying causes for motivations (or lack thereof) for participating in HIE.

Physicians: Whether physicians can upload their patient’s complete medical history data on HIE without patient’s consent is an issue that is not yet fully settled. Absence of patient’s complete medical history would make it difficult for other physicians in hospitals or speciality clinics to provide care when cases are referred to them, thus defeating the objectives of HIE. Physicians express concern about the implied patient-physician contract and the violation of such contractual terms if patient data is made sharable by the consulting physician. This holds true in the Malaysian context as well (Banerjee and Bagha, 2014). Absence of regulations and legislations complicate these issues.

Government: Even when regulations are attempted by government, there are delays in framing laws and further delays in their enforcement (Qing, 2013). Existing laws may also be interpreted to better serve the interest of professional bodies and may come in conflict with the provisions of some laws. For example, Malaysia’s Personal Data Protection Bill was first drafted in 2001 but a long time elapsed before Malaysia’s Personal Data Protection Act 2010 was announced. It was due to take effect on January 1, 2013, but the law is still not in force due to legal formalities. It was finally brought into force on 15 November 2013 and three months grace period has been given to organisations for compliance till end on 25 February 2014. However, companies and organizations are said to be still lacking in compliance while consumers doubt it would be strongly enforced. Many Malaysian companies are still not prepared for the eventual implementation of the law, (Qing, 2013).

The above review indicates that for successful deployment of Malaysia’s HIE, steps must be taken to resolve stakeholder concerns and lack of motivation in regard to technical, organizational and environmental barriers. In the next section we suggest some strategies for minimizing these identified barriers.

ADDRESSING BARRIERS TO DEPLOYMENT OF MYHIX

The following strategies may be adopted for addressing the barriers to deployment of MyHIX.

ADDRESSING TECHNICAL BARRIERS

Unlike the US, EU and other developed countries, in Malaysia, EHR systems are still not available in the market. We argue that not having an EHR system gives opportunity to the Malaysian government to co-ordinate and consult with Malaysian software developers and users of EHR, and develop a standardised system with joint consultation. The back-end data-exchange standards definition already put in place by the Malaysian government should facilitate this process.

Preliminary work on MyHIX done by the Malaysian government would further facilitate joint collaboration for technical infrastructure and EHR software development. For example, in Feb 2013, the national research and development (R&D) centre in information and communication technology, MIMOS Berhad signed an MoU with the Ministry of Health (MoH) for deployment of Healthcare Information Technology (Healthcare IT). In terms of the MoU, MoH would collaborate with the R&D centre in the development, implementation and proliferation of the Ministry’s Healthcare IT framework; which would comprise MIMOS-developed technology components such as knowledge management system, cloud computing, intelligent informatics platform, intelligent service delivery platform and network-based information security system. Such an opportunity exists for starting straightforward with a cloud based EHR system developed by MIMOS in consultation with physicians in hospitals and small clinics. It could be deployed on a secure private cloud and accessed by small clinics and hospitals on a secure server with secured network access. Such a system would reduce training time. The system could be government certified (US government mandates certified EHR use by clinics). This certification could be provided by the Malaysian Software Testing Board (MSTB). Certification from a
national Malaysian body would contribute to allaying fears of software security and other interoperability concerns of physicians and patients. Such technology planning would remove the barrier of low EHR adoption seen in other countries, and pave the way for Malaysia to overtake other countries in HIE deployment and provide quality health care service delivery.

A crucial technical enabler is also the provision of a secure telecommunication network infrastructure. To this end, MIMOS can steer collaborative moves with Malaysia’s local private and government telecommunication providers such as Digi, TM, Celcom, Maxis etc. Bandwidth of current networks may not be adequate, particularly if large digital data files have to be accessed from HIE for critical care. This can be facilitated by the government by discussions with the telecommunication providers and MIMOS (responsible for the EHR systems).

ADDRESSING ORGANISATIONAL BARRIERS

The Malaysian government has re-structured the MoH and the functional responsibilities in the re-organised structure (Refer to figure 1). This is likely to facilitate governance mechanisms necessary for successful deployment of MyHIx. However, effectiveness of this revised organisational structure would depend to a large extent on the ability of the different departments to co-ordinate and work towards the common goal of transforming health care with MyHIx. Power equations and rivalry in government departments are known to thwart government initiatives, which hinders trust formation in collaborative moves such as HIE (Rossa et al, 2010). A central steering committee composed of important stakeholders in the MyHIx, drawn from the industry and government departments, with powers to arbitrate conflicts and resolve issues that may derail the MyHIx project may be needed.

Stakeholder understanding of the medical context is also essential for stakeholder participation in MyHIx. Patients, insurance providers, software vendors/developers of health information systems, suppliers/providers of medical equipment including wearable medical devices used by patients, must be trained in the medical context and also on the objectives of MyHIx. This would generate shared understanding and remove barriers of trust, thereby fostering participation by these stakeholders.

ADDRESSING REGULATORY BARRIERS

Institutional mechanism such as government incentives for participation and penalties for not participating in HIE initiative could be an enabler. The US government has enacted a legislation for meaningful use of EHR. (Refer to figure 3 in appendix for aspects of the meaningful use criteria). The Malaysian government can enact such legislations for meaningful use so that private clinics and hospitals who are not directly under the control of MoH may participate meaningfully in EHR for successful deployment of MyHIx.

The primary reason for low adoption of EHR, which is a significant barrier to successful deployment of HIE as discussed in the literature review, appears to be security issues. Physicians prefer to store their data locally on their site rather than HIE. A program called D2D to develop a national HIE has been conducted in Germany but less than 2% (2300) of physician in private section joined this program (Jha et al, 2008). Thus the Malaysian government must be able to enact legislations and enforce them quickly to address security concerns of physicians in regard to EHR usage and participating in HIE. Mere technology security may not be sufficient to build this trust in physicians. Delays in framing legislations and enforcing them could be detrimental to MYHIx deployment.

Issues to be addressed through legislations would clarify who owns data on the HIE and who would grant access to patient data on the HIE exchange. For example, in Australia, the patient decides whether access rights can be given to other physicians and stakeholders. Can the same mechanisms be adopted and applied in the Malaysian context? For example, under the PDPA Act 2010, all doctors are required to register as data users if they are licensed under the Private health care facilities and services act for a private clinic, chain clinic or dental clinic. The Malaysian Medical Association feels this would expose doctors to undue and unwarranted risks and it has therefore asked the government to exempt doctors from having to register as ‘data users’ under the PDPA Act 2010 (Fong, 2014). Unless the issue is resolved, it can be seen that participation by physicians in the HIE initiative may not occur voluntarily.

CONCLUSION

By using an integrated perspective of the TOE and stakeholder theory, we have identified the technology, organisation and environmental barriers that could negatively affect Malaysia’s move towards a national HIE and how different stakeholders perceive these barriers. Our research demonstrates that deployment of HIE is more than a technology initiative. While there are issues of technology, there are equally important organisational and regulatory barriers that
must be addressed. Stakeholder perceptions of these barriers may differ and conflict of interest may lead to different levels of motivation for participation in Malaysia’s MyHIX. Stakeholder conflicts of interest may be resolved through government interventions, specifically by framing appropriate legislations and acts. Financial incentives, especially for physicians small clinics, to hospitals, and software suppliers may also help. Malaysian government can create facilitating conditions through regulatory measures, financial support and training of stakeholders like primary care providers (small clinics), tertiary care providers (hospitals), software and technology infrastructure providers, and medical insurance providers.

However, there is evidence that inspite of best efforts and incentives from government, adoption and meaningful use of EHR in advanced countries such as US and Europe is still low. In Canada, HIE implementation has slowed down due to lack of EHR systems adoption by clinics and hospitals; the national Infoway program aimed at increasing EHR adoption amongst clinics and hospitals in Canada has not produced the desired results. In Germany the effort to develop a national HIE by software companies has not been successful so far (Jha et al., 2008)

The US government now uses a ‘carrot and stick’ approach – adopt and move up the meaningful use ladder and get incentives, or get penalised for not doing so. Whether the Malaysian government can avoid this situation by concerted effort for training physicians upfront prior to mandating EHR adoption, must be explored. Crucial to such training is government’s development and approval of a standard and secure EHR software that integrates with hospital systems and the MyHIX platform. Thus a staged approach for removal of the identified barriers may be necessary.

While government led initiatives would lead to a governmental ‘Push’ for Malaysia’s MyHIX, there is also a need for medical associations, management personnel in hospitals, insurance providers to proactively create conducive conditions that can generate a ‘Pull’ effect for MyHIX in the minds of the physician and patients, the primary stakeholders.

**Practical and Theoretical Contribution**

This research has significant practical contribution in terms of informing Malaysian policy makers and other stakeholders associated with health care delivery about the barriers to deployment of MyHIX and how they can be addressed by their active participation. The findings may be extended by development of testable hypotheses and validating the hypotheses through appropriately designed large scale surveys.

The theoretical contribution of this research is in terms of highlighting the usefulness of integrated perspectives from appropriate frameworks and theories to conduct exploratory analysis of problems related to a phenomenon and prescribing theoretically grounded solutions. The Malaysian government is keen on initiatives to foster adoption of IT by small and medium firms to help them grow and sustain in a competitive landscape and is funding research, for example exploring theoretical perspectives that can help explain adoption of cloud computing by SMEs (Banerjee etal, 2011). This research complements such initiatives.

**REFERENCES**


<table>
<thead>
<tr>
<th>Primary care</th>
<th>Australia</th>
<th>Canada</th>
<th>Germany</th>
<th>Netherlands</th>
<th>New Zealand</th>
<th>United Kingdom</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHR (%)</td>
<td>79-90</td>
<td>20-23</td>
<td>42-90</td>
<td>95-98</td>
<td>92-98</td>
<td>89-99</td>
<td>24-28</td>
</tr>
<tr>
<td>CPOE (%)</td>
<td>75-81</td>
<td>5-11</td>
<td>59</td>
<td>85</td>
<td>90</td>
<td>&gt;90</td>
<td>9</td>
</tr>
<tr>
<td>Hospital care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EHR (%)</td>
<td>&lt;10</td>
<td>&lt;10</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>&lt;10</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>CPOE (%)</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;5</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: EHR Participation in Some Developed Countries

Table 1: Meaningful Use of EHR Criteria

<table>
<thead>
<tr>
<th>Meaningful Use Objectives</th>
<th>Corresponding EHR Software Features</th>
<th>Meaningful Use Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use computer provider order entry (CPOE)</td>
<td>Enable a user to electronically record, store, retrieve, and manage, at a minimum, the following order types: Medications; Laboratory; Radiology/imaging; Provider referrals; Blood bank; Physical therapy; Occupational therapy; Respiratory therapy; Rehabilitation therapy; Dialysis; Provider consults; and Discharge and transfer.</td>
<td>CPOE is used for at least 80% of all orders; 10% for hospitals</td>
</tr>
<tr>
<td>Implement drug/allergy checks</td>
<td>(1) Real-time, alerts at the point of care for drug-drug and drug-allergy contraindications; (2) Electronically check if drugs are in a formulary or preferred drug list; (3) Provide certain users rights to deactivate, modify, and add rules for drug-drug and drug-allergy checking; (4) Track number of alerts users respond to</td>
<td>Function is enabled</td>
</tr>
<tr>
<td>Maintain an up-to-date problem list of current and active diagnoses based on ICD-9-CM or SNOMED CT</td>
<td>Electronically record, modify, and retrieve a patient’s problem list over multiple visits</td>
<td>At least 80% of all unique patients have at least one entry or an indication of none recorded.</td>
</tr>
<tr>
<td>E-prescribing (EP only)</td>
<td>Electronically transmit prescriptions</td>
<td>At least 75% of all permissible prescriptions written by the EP are transmitted electronically</td>
</tr>
<tr>
<td>Maintain active medication/allergy list</td>
<td>Electronically record, modify, and retrieve a patient’s active medication/allergy list</td>
<td>At least 80% of all unique patients have at least one entry or an indication of “none”</td>
</tr>
<tr>
<td>Record demographics</td>
<td>Electronically record, modify, and retrieve patient demographic data</td>
<td>At least 80% of all unique patients have demographics recorded</td>
</tr>
<tr>
<td>Record and chart changes in vital signs</td>
<td>(1) Enable a user to electronically record, modify, and retrieve a patient’s vital signs; (2) Automatically calculate and display body mass index (BMI); (3) Plot and electronically display, upon request, growth charts for patients 2-20 years old.</td>
<td>For at least 80 percent of all unique patients age 2 and over seen by the EP or admitted to the eligible hospital, record blood pressure and BMI, additionally, plot growth chart for children age 2 to 20</td>
</tr>
<tr>
<td>Record smoking status for patients 13 years old or older</td>
<td>Electronically record, modify, and retrieve the smoking status of a patient</td>
<td>At least 80% of all unique patients 13 years old or older have “smoking status” recorded</td>
</tr>
</tbody>
</table>
Table 1 (continued): Meaningful Use of EHR Criteria

<table>
<thead>
<tr>
<th>Incorporate clinical lab test results into EHR as structured data</th>
<th>(1) Electronically receive clinical laboratory test results and display such results in human readable format; (2) Electronically display in human readable format any clinical laboratory tests that have been received with LOINC codes; (3) Electronically display all the information for a test report; as structured data: (4) Electronically update a patient's record based upon received laboratory test results</th>
<th>At least 50% of all clinical lab tests results are incorporated</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Generate lists of patients by specific conditions</td>
<td>Electronically select, sort, retrieve, and output a list of patients and patients' clinical information</td>
<td>Generate at least one report listing patients with a specific condition</td>
<td></td>
</tr>
<tr>
<td>Report ambulatory quality measures to CMS or the States (EP only)</td>
<td>(1) Calculate and electronically display quality measure results as specified by CMS or states; (2) Electronically submit calculated quality measures</td>
<td>For 2011, an EP/hospital would attest this has been done</td>
<td></td>
</tr>
<tr>
<td>Send reminders to patients for preventive/follow-up care</td>
<td>Electronically generate a patient reminder list for preventive or follow-up care</td>
<td>Reminders sent to at least 50% of all unique patients that are 50 and over</td>
<td></td>
</tr>
<tr>
<td>Implement five clinical decision support rules relevant to specialty or high clinical priority</td>
<td>(1) Implement automated, electronic clinical decision support rules according to specialty or clinical priorities; (2) Automatically and electronically generate real-time alerts and care suggestions based upon clinical decision support rules and evidence grade; (3) Automatically and electronically track, record, and generate reports on the number of alerts responded to by a user</td>
<td>Implement five clinical decision support rules relevant to the clinical quality metrics the EP/Eligible Hospital is responsible for</td>
<td></td>
</tr>
<tr>
<td>Check insurance eligibility electronically</td>
<td>Electronically record and display patients’ insurance eligibility, and submit insurance eligibility queries</td>
<td>Insurance eligibility checked electronically for at least 80% of all unique patients</td>
<td></td>
</tr>
<tr>
<td>Submit claims electronically to public and private payers.</td>
<td>Electronically submit claims</td>
<td>At least 80% of all claims filed electronically</td>
<td></td>
</tr>
<tr>
<td>Provide patients with an electronic copy of their health information upon request</td>
<td>Enable a user to create an electronic copy of a patient's clinical information and provide to a patient on electronic media, or through some other electronic means</td>
<td>At least 80% of all patients who request an electronic copy of their health information are provided it within 48 hours</td>
<td></td>
</tr>
<tr>
<td>Provide patients with an electronic copy of their discharge instructions and procedures at time of discharge, upon request (Hospital only)</td>
<td>Enable a user to create an electronic copy of the discharge instructions and procedures for a patient, in human readable format, at the time of discharge to provide to a patient on electronic media, or through some other electronic means</td>
<td>At least 80% of all patients who are discharged from an eligible hospital and who request an electronic copy of their discharge instructions and procedures are provided it</td>
<td></td>
</tr>
</tbody>
</table>

Acknowledgment:

This research is funded by the Ministry of Higher Education, Malaysia under the Fundamental Research Grants Scheme; Award No: FRGS 1/2012/SS05/Swin/02/1
UTILIZATION OF WOMEN RESOURCES WITH HIGHER EDUCATION IN KOREA

Choonhee Yang
Department of Business Administration
Namseoul University
Chenan, Chungcheongnam-do, S.KOREA

ABSTRACT

Korea has achieved immense economic growth in a relatively short period of time owing to the society’s dedication to education and abundant human resources. However, its subsequent low birthrate and aging society have resulted in the decrease of the population of key economic actors and the increase of the elderly population.

Korea has a high college entrance rate that exceeds 80%. Since 2009 the ratio of men to women attending college has been fairly equal and even overturned at times. And while the economy activity rate of Korean women has been on a steady increase, there exists a serious problem of career discontinuity due to childbirth and childcare in women.

Korea has the lowest rate of female managers among OECD countries. Furthermore, the issue of the glass-ceiling, which arises at the intermediary managerial level, is more severe in Korea than in other countries.

The purpose of this paper is to analyze the various statistics on women resources in Korea; to find out the current status of the utilization of Korean women resources; and to suggest policies that will aid the development of women resources with higher education.

INTRODUCTION:

Korea is suffering from a labor shortage due to its aging society and low birth rate. Statistics show that the nation’s prime-aged labor force peaked in 2011 with 24,318 thousand, then dropped to 24,307 thousand in the following year. It is now expected that the population will drop by 2.22 million in 2025.

Korea will not be able to continue its economic growth in the long term unless it finds a substitute to its currently dwindling labor force. One method used by developed countries to supplement its lack of economically active people is to promote immigration. However, immigration has had a largely insignificant impact. Thus, taking Korea's economic environment into consideration, the utilization of the abundant women resources with higher education can be said to be a major strategy for Korea.

A good indicator of Korea's women resources with higher education is the college entrance rate. While more number of male high school students than female high school students entered university in 2000, the ratio of female students entering college continued to rise to become almost equal to that of men at 84% in 2008, until it was finally overturned in 2009.

However, the potential for these highly educated women in the labor market remain largely unrealized, with Korea having one of the lowest rates of women in the labor market among the 34 OECD countries.

Women with higher education encounter many obstacles when entering the labor market. Those that have succeeded then face career discontinuation due to personal matters such as marriage, pregnancy, childbirth, and childcare. Many women who return to work after that are unable to get jobs that require continued career development.

This paper aims to find out the current state of women in the Korean labor market via statistics, and to analyze the utilization of women resources according to different sectors.

ACTUAL CONDITION OF WOMEN RESOURCES

Career Discontinuity of Women Resources

Burden of children care causes 20-30 aged women resources to give up career development. This leads to career discontinuity of labor market, and finally it makes women resources with higher education not to use their possibility, called, M-curve.

Among women with work experience, the ratio of women who quit their job because of family file is 58%. The proportion of women who stopped their job of family file reason and comeback is 66.3%. (Kim, M, A, 2014)
Female Resources by age in Canada and Sweden which can be developed countries shows up reverse shape U where as the proportion of economic activity rate of women Resources aged 20’s to 30’s in Korea appears ‘M shape curve’. It means that women resources might be suffering discontinuity of labor market.

It is so difficult for women resources to take a job again after raising their children in the labor market. They already forgot the sense of conducting the job. If they had been at work, they could have been expert at their field at the prime age. Even though they can enter the labor market, their job cannot be core job any more, low-ranking occupation or simple contract status. (Kim, T, H, et al., 2013)

Women’s Economic Activity

In Korea, a point of sight 2013, the rate of male economic activity rate shows 77.9%, where as that of female turns up 56.1%. The gap is over 20%. (Ministry of Women and Family, 2013)

Let’s compare with other countries. Among comparison target, OECD leading countries, women’s economic activity rate of Korea is very low. (See Table 1)

A point of sight 2012, women’s economic activity rate of Korea is 55.2%, compared with OECD mean of activity rate is 62.3%, the gap is 7%. Among OECD 34 countries, the rate of Korea women’s economic activity is ranked 30. Among OECD countries, the rate of Korea is very lower than the rate of Iceland, ranked first place, 83.3%. There is an enormous discrepancy between those rates.

UTILIZATION OF FEMALE RESOURCES WITH HIGHER EDUCATION

The Employed by Education Level, and by Gender

There is a difference between education level in terms of utilization of human resources. Employment population ratio of female college graduate marked 58.4% in 2000 is increasing 62.4% in 2007. Since then the percentage is somewhat decreased in 2008 global financial crisis, but the percent of employment population ratio of female college graduate marked 61.65% in which is that of higher than female with lower education group. (Kim, T, H, 2013)

But in gender comparison at the same college graduate, employment population ratio of female with college graduate is 61.65%, where as that of male is 86.8%, in the time of 2012. The Percentage of female college graduate is very lower than that of male in Korea. When we compared with main OECD countries’ rate, 80.4%, the gap is about 20% that shows a big difference in the rate of employment of female college graduate.

A High College Entrance Rate of Female in Korea

College entrance rate of female in 2001 was less 70%, but it is over 80% in 2013. The gap between male and female is few. In 2009, the female college entrance rate surpassed that of male. Table 2 depicts a high college entrance rate of female is 80.5% in 2010, all the more the rate of male is less 77.6%. In recent, there are similar ratios by gender. In spite of college entrance rate of female, we can make an interpretation that there is too little job for female. (See Table 2)

Economic Activity Rate of Women with Higher Education

According to recent survey, the percentage of female employment of just fresh out of college graduate marked 71.2%, where as that of female marked 66.5 in 2008. But in 2012, the gap is getting diminished between male and female. (See Table 3)

In order to separate out college graduate activity rate from total female activity, I compare women resources with higher education in Korea with those of leading OECD countries.

Table 4 shows female economic activity rate of total female is 62.1% that shows a big difference in the rate of OECD average, 82.6%. The gap is above 20%. The rate of economic activity of college graduate in Korea is ranked last in the OECD countries. (See Table 4)

Current State of Female Management Staff

As mentioned before, the employment-population ratio of women with a college diploma was increasing consistently 58.4% in 2000 to 62.4%, since global crisis in 2008 it was decreasing a little, recently is on the comeback trail.

The ratio of female administrative position in Korea is on the low rank among the OECD 34 countries. The ratio of female administrative position of OECD average 28.3%, where as that of Korea is only 9%. (OECD, 2009)

Working women in the Private enterprise are mainly sub-ordinary Position. Moreover, the proportion of female management staff is very lower than that of other sectors. According to the statistics, the ratio of female executives in the private large company was increasing 6.5% in 2008 to 6.6% in 2009. In other countries’ instance, female executive ratio at the large company in Korea is very lower than that of 41% in Norway, 27% in Sweden, 13% in Germany. (Kim T., H., et al., 2013)

Most of women go through an ordeal to promote a higher position, called ‘Glass Ceiling’ especially from
section manager. Like this, women with college graduate diploma and with high standard competence experience frustration entering the middle management position.

CONCLUSION

Korea is aging at a faster rate than the world’s major developed nations. Previous research shows that the utilization of women resources in companies led to improved management results. Nevertheless, Korea is not fully utilizing women with higher education as a labor source as it should. This is a major reason that is reducing Korea's national competitiveness.

While the rate of Korean women entering college is equally high as that of men, their employment is significantly lower than that of their counterparts. Furthermore, women in the workforce suffer career discontinuation due to matters such as childbirth and childcare. When they return to the labor market this discontinuation poses an obstacle in their attaining jobs at the core of the business. Consequently, women are given peripheral duties.

One of the ways to prevent the career discontinuation in highly educated women is governmental support. For example, policies that provide childcare facilities will address the physical aspects, or the “hardware” of childcare for the mothers. At the same time, companies altering their management practices to allow for a better balance between work and personal time will provide the “software”.

Korean women hit a glass ceiling during promotion sat the intermediary managerial level. This leads to a lack of women in the workplace who can act as role models and networking partners for female college graduates. It might be worthwhile considering the Northern European model of promotion quotas for women as a method of turning this vicious cycle into a virtuous one. Bold actions to support women by the government are necessary to have an impact on the highly male-oriented Korean business culture.

It is a national waste to have highly educated women lag behind men in terms of organization capacity, leadership capacity, and work capacity. Based on these findings, it can be concluded that Korea’s answer to supplementing the currently decreasing labor force is developing and utilizing the women resources available at hand.

REFERENCES


Korea National Statistics Office, Each Year Statistics.


Korean women’s Development Institute, “Annual Report 2013”.


Ministry of Gender Equality & Family, 2013, “Major Statistics Data”.


OECD, Employment Outlook 2013

FIGURES AND TABLES

Table 1   Leading Countries among OECD Women’s Economic Activity

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2010</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea</td>
<td>55.2%</td>
<td>54.9%</td>
<td>54.5%</td>
</tr>
<tr>
<td>Japan</td>
<td>63.4%</td>
<td>63.0%</td>
<td>63.2%</td>
</tr>
<tr>
<td>U.S.A</td>
<td>76.7%</td>
<td>67.8%</td>
<td>68.4%</td>
</tr>
<tr>
<td>OECD</td>
<td>62.3%</td>
<td>61.8%</td>
<td>61.8%</td>
</tr>
</tbody>
</table>

source : OECD, Employment Outlook, 2013
Table 2  College Entrance Rate by Gender

<table>
<thead>
<tr>
<th>year</th>
<th>High School -&gt; College</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (%)</td>
</tr>
<tr>
<td>2001</td>
<td>70.5</td>
</tr>
<tr>
<td>2002</td>
<td>74.2</td>
</tr>
<tr>
<td>2003</td>
<td>79.7</td>
</tr>
<tr>
<td>2004</td>
<td>81.3</td>
</tr>
<tr>
<td>2005</td>
<td>82.1</td>
</tr>
<tr>
<td>2006</td>
<td>82.1</td>
</tr>
<tr>
<td>2007</td>
<td>82.8</td>
</tr>
<tr>
<td>2008</td>
<td>83.8</td>
</tr>
<tr>
<td>2009</td>
<td>81.9</td>
</tr>
</tbody>
</table>

Source: Ministry of Education, science and Technology; Korea Education Development Institute, Education Statistical Year Book, 2013

Table 3  Employment Figures of College Graduate by Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>71.2</td>
<td>70.2</td>
<td>55.7</td>
<td>57.3</td>
<td>55.9</td>
</tr>
<tr>
<td>Female</td>
<td>66.5</td>
<td>66.1</td>
<td>47.9</td>
<td>50.0</td>
<td>51.7</td>
</tr>
</tbody>
</table>

Source: Ministry of Gender Equality & Family, Major Statistics Data, 2013

Table 4  OECD Leading Countries Economic Activity Rate (at the age of 25 to 64) unit; %, 2010

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea</td>
<td>92.2</td>
<td>62.1</td>
</tr>
<tr>
<td>OECD</td>
<td>91.9</td>
<td>82.6</td>
</tr>
<tr>
<td>Japan</td>
<td>95.5</td>
<td>69.4</td>
</tr>
<tr>
<td>U.S.A</td>
<td>89.8</td>
<td>80.0</td>
</tr>
<tr>
<td>U.K</td>
<td>91.9</td>
<td>82.8</td>
</tr>
<tr>
<td>Australia</td>
<td>92.3</td>
<td>81.6</td>
</tr>
<tr>
<td>Canada</td>
<td>89.7</td>
<td>82.8</td>
</tr>
<tr>
<td>France</td>
<td>91.2</td>
<td>85.2</td>
</tr>
<tr>
<td>Sweden</td>
<td>93.6</td>
<td>90.9</td>
</tr>
<tr>
<td>Norway</td>
<td>93.0</td>
<td>90.9</td>
</tr>
</tbody>
</table>

Source: OECD employment outlook 2012
A STUDY ON THE QUALITY FACTOR OF
CULTURE MARBLE BATH-TUB & WASHBOWL

Hwang Kyoo-il, Artima
Department of Business Administration
Namseoul University
91 Daehakro, Seonghwa-eup, Seobuk-gu, Cheonan-city, Choongnam, 331-707
KOREA
artima@nsu.ac.kr

ABSTRACT

Bathtubs and Washbowls over the world are mainly produced by small and medium-sized companies and there are different manufacturing methods according to the companies and their nations. They are regulated mainly by their national industrial standards. South Korea, for example has Korean Industrial Standards.

This study aims to produce high-quality products by suggesting factors for successful quality management in regards to Korea’s small and medium-sized companies manufacturing marble bathtubs and washbowls. To be specific, these factors under manufacturing environment are: Temperature, Viscosity, Line, and Surface. Especially for the temperature factor, the researcher has put in much time and cost in setting the adequate temperature, duration and conditions for each process. The researcher aims to go public his experiences and promote overall developments in the related industry.

While studies on marble lacks over the world, this study expects to present new standards for quality management in related fields.

II. FEATURES OF MARBLE BATHTUBS AND WASHBOWLS

1. Features of Culture Marble
   - Culture Marble is manufactured by naturally earned stone flour and is environment-friendly.
   - By adding various types of stone flours and pigments, it can make diverse colors, such as texture of natural stone, white, or pink.
   - It can be produced in various sizes from very small accessories to large bathtubs or wash basins.
   - It can be used semi-permanently thanks to its excellent material property and durability.
   - Flaws or cracks that have been made after or while installing the product can be simply fixed and reused again.
   - It can be produced in a short period of time and is cost-friendly.
   - There is no burden of expense for molding. The manufacturer can produce various molds which lowers the burden of fixed investment cost.

2. Source Materials for Cultured Marble Bathtubs and Washbowls

In producing cultured marble product,
unsaturated polyester resin—an adhesive, and stone flour—which works as fillers are main source materials. Judged by the appearance, it has almost the same texture and shape with natural stones such as marble, wave pattern, granite and gem stone. Details are shown in Table 1.

3. Main Products

Main products of culture marble are bathtubs and washbowls. Also, there are other products widely used in our daily lives, such as dressing tables incorporating basins, wall materials, flooring materials, sink materials, sink counter, door-sills, material segregators, plaques, toilet partitions, chairs, agricultural pipes and dolls.

4. Korea’s National Standards for Product testing (KSF-4806)

Details are shown in Table 2.

III. KEY FACTORS FOR SUCCESSFUL QUALITY MANAGEMENT

1. Temperature

Add 1~2% of hardener into the resin content of the compound, and when the certain time pasts after mixing them, the mixture will harden through chemical reaction. If any artificial heat is applied during this process, it will boost the reaction and the mixture will combine more finely. The artificial heating is necessary because the resin is thermosetting resin and needs heat for adhesion. Maintaining the artificial heat from the source material to product forming is a crucial factor to improve the quality. It is because the heat makes high-quality products resistant from external conditions for a long time. The products get features such as durability, fouling resistance, weatherability, and rigidity. The appropriate temperature for each process is in Table 3.

Maintaining the right temperature leads to excessive investment on machine equipment and use of fuel. For this reason, many companies neglect the importance of keeping the right temperature blaming the cost increase and low productivity. Some companies that fail to invest in curing facilities, add 5~6% of hardener to the resin content which is higher than the regulation requires-2%, to evade temperature conditions. In such cases excessive amount of hardener rather may have adverse effects on the product, making flaws. Also, it may cause sharp exothermic reaction leading to cracks or modifications of the products, and hamper the chemical combination reaction, preventing the adhesion of stone flour.

2. Viscosity

As people mix sand, pebbles and water in the proper proportion when mixing cement mortar, adequate proportion of stone flour and resin is also required in producing imitation stones. Compared with the cement mortar, the stone flour is added instead of sand and pebbles, and resin, instead of water. Thus, in order to make a proper mixture of stone flour and resin, there are various mixing ratios according to the source materials. In most cases the proper ratio should be 66~80 percent of stone flour and 20~34 percent of resin. Maintaining the adequate temperature between 20~25°C is crucial. If the temperature is too low, the viscosity thickens relatively more, making it difficult to fill the compound in to the mould. Also, it may produce faulty goods as it lowers the adhesion. The proportion of raw materials is shown in Table 4.

Stone flour is used in various thicknesses—from #80 to 320 mesh, so it can maintain the adequate viscosity for production. Generally, when the stone flour is thin, more resin is needed than the thicker ones. It is because the thinner mesh takes up more surface area than the thicker mesh in same volume.

When the content of the resin is low, it can be a reason for low durability, causing cracks or swelling. This is mainly because the resin is very costly and the manufacturer reduces the amount of the resin, producing faulty goods. On the other hand, too much resin makes too much bending and unnecessary cost increase.

3. Line

Bathtubs and Washbowls are finishing material in building construction. Moreover, especially as they are connected with tiles in bathrooms, they need to be horizontal and vertical.

Stone flour does not shrink, however resin shows around 7 percent of contraction. Therefore, when these two are mixed together, 1~3 percent of shrinkage will be witnessed.

In order to meet the horizontal and vertical conditions, a level, a mass of iron and jig are used. During the forming process the product is placed on the level, and by using the mass of iron or jig, it prevents any bending until the product cures completely.

Among the marble products, bathtubs, washbasins, shelves, and thresholds for bathroom doors should keep horizontal and vertical lines. To be specific, bathtubs, washbasins, and shelves should be horizontal with the mirror or tiles. The door
thresholds also should be horizontal and vertical with the part where the doorframe and the door meet. Example is shown in Figure 1.

4. Surface

Bathtubs and washbowls are used daily and are within people’s reach. Thus, the surface of such products has to be very smooth and sleek. Also, it should be kept clean even after use.

First of all, the surface of the mould needs to maintain its delicate and smooth surface. It depends on the surface process of the mould and main framemaking which can only be done by skilled workers.

Second, after covering the product with mould releasing agent it needs to be cleaned well. Before coating the product with gel coat-to take the product off the mold easily, the mold releasing agent needs to be applied, wiped and polished well.

Third, the quality, thickness of the gel coat depends on the skill of the worker. The gel coat shows sensitive chemical reactions and without thorough quality control, it cannot reduce the differences in the chemical reaction. The following are conditions required for using gel coat.

i. The gel coat needs to maintain adequate viscosity and shows fine workability.
ii. It should be weatherproof, not showing discoloration.
iii. The surface needs wear resistance.

IV. Research Variables

This study has set temperature, viscosity, line, and surface as key factors for quality management, which are derived from prior studies. Key factors are shown in Table 5.

V. Analysis Methods

1. Variable Selection

This study has used variables in analyzing the key factors for manufacture. In terms of statistics, age, educational background, service career, assigned task, company scale-number of workers, sales figures have been the standards for the sample. Respondents were to choose one response per question. The survey requests answers on an interval scale so it can be used in regression analysis, factor analysis, correlation analysis and variance analysis (ANOVA). For analyzing the relevance, it practiced ANOVA. Variables used in analysis are shown in Table 6.

VI. Empirical Analysis

1. Statistical Analysis

In order to collect the data, the researcher has sent surveys to 310 manufacturing companies, and 275 companies have answered. Among them, 202 questionnaires have been used for this research. The result of the survey is achieved through frequency analysis and is shown in Table 7.

2. Reliability and Validity Analysis

In reliability measurement, there are test-retest method, split-half method, parallel-forms reliability technique and internal consistency. By constantly asking one same concept in series of questions the researcher estimates the internal consistency, and it uses Cronbach’s alpha in analyzing the reliability.

The range of Cronbach α is between 0 and 1, and if Cronbach’s alpha is higher than 0.8, it shows good reliability, and acceptable when it is over 0.6. The result for this study is shown in Table 8.

Based on the analysis, in order to increase the reliability of the questions that resulted very low in the first analysis, the researcher did the second analysis by excluding questions that are relatively less relevant from each range. As a result, all of the questions recorded higher than 0.6, making the study reliable.

For factor extractions, the study used method of generalized least squares, and for the rotation method, it applied Varimax. Also, to find the common factor, the study did the factor analysis using three factors. The result is shown in Table 9.

VII. Conclusion

While studies on marble lacks over the world, this study expects to present new standards for quality management in related fields.

Bathtubs and washbowls are mainly produced by small and medium-sized companies. Each nation has its own national standard. It sets quality standards and its companies produce their bathtubs and washbowls according to these standards. As these nations and companies have different manufacturing methods, this study aimed at suggesting key factors deciding the quality of the products and providing standard quality factors.

To be specific, for the key factors which affect the quality of products in manufacturing environment are: Temperature, Viscosity, Line, and Surface. Especially, this study is based on the researcher’s own experience in setting the temperature and duration for each process, making good suggestions to the related industries.

The researcher aims to make public his know-hows and thirty-year experience in a Korean small and medium-sized manufacturing company. The
A researcher hopes that bathtub manufacturing companies of the world have quality standards and take their full social responsibility by manufacturing high-quality products.

**REFERENCE**


ANSI Z124.3, for plastic lavatories, 2010


Go Young Sik, 2003, ‘*Research on Material Properties of Unsaturated Polyester Resin and Imitation Marble*’, Samhwa Paint Institute of Technology.

JIS A 5704 (Japanese Industrial Standards), Culture Marble Bathtubs, 2010

KS F 4810 (Korean Industrial Standards), Culture Marble Bathtubs, 2011

**FIGURES AND TABLES**

### Table 1 Raw Materials

<table>
<thead>
<tr>
<th>Material</th>
<th>Adhesive</th>
<th>Filler</th>
<th>Stiffener</th>
<th>Addition</th>
<th>Surface</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>unsaturated polyester resin</td>
<td>A(OH)$_3$</td>
<td>CaCO$_3$</td>
<td>SiO$_2$</td>
<td>Glass Fite</td>
<td>Glass Chop 1.5-6 mm</td>
</tr>
<tr>
<td>Ratio</td>
<td>20-35%</td>
<td>40-65%</td>
<td>1.5-6%</td>
<td>Suitable amount</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 2 Performance Requirement Standards of Bathtubs

<table>
<thead>
<tr>
<th>Test Details</th>
<th>Performance Requirements</th>
<th>Measurement Unit</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of Malt Modification of Apron</td>
<td>The degree of the horizontal external force the apron bears</td>
<td>mm</td>
<td>Modification of the central part should be within 10mm.</td>
</tr>
<tr>
<td>Modification when filled with water</td>
<td>The degree of the modification when filled with water</td>
<td>mm</td>
<td>The bending of the drain and four central parts of the upper boarder should be within 2mm.</td>
</tr>
<tr>
<td>Sandbag Shocks</td>
<td>Changes on the surface when any impacts are given by human body</td>
<td>-</td>
<td>There should be no cracks on the surface nor any harmful flaws in using the bathtubs</td>
</tr>
</tbody>
</table>

### Table 3 Adequate Temperature and Duration of Each Process

<table>
<thead>
<tr>
<th>Status</th>
<th>Process</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>Store the source materials-resin, stone flour</td>
<td>20 ~ 25°C</td>
</tr>
<tr>
<td>While Mixing</td>
<td>Keand resin and stone flour</td>
<td>20 ~ 25°C</td>
</tr>
<tr>
<td>During the Production Process</td>
<td>Coating with gelcoat</td>
<td>50°C for 15min.</td>
</tr>
<tr>
<td>Promote semi-products</td>
<td>50~60°C for 1hr.</td>
<td></td>
</tr>
<tr>
<td>After the Production</td>
<td>Hardening finished products</td>
<td>80°C for 1 hr.</td>
</tr>
</tbody>
</table>

### Table 4 Proportion of Materials

<table>
<thead>
<tr>
<th>Material</th>
<th>Proportion(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stone Flour</td>
<td>66 ~ 80</td>
</tr>
<tr>
<td>Resin</td>
<td>34 ~ 20</td>
</tr>
</tbody>
</table>
Table 5 Research Model

<table>
<thead>
<tr>
<th>Key factors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td></td>
</tr>
<tr>
<td>Viscosity</td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td></td>
</tr>
<tr>
<td>Surface</td>
<td></td>
</tr>
</tbody>
</table>

Table 6 Variables Used in Analysis

<table>
<thead>
<tr>
<th>Factor</th>
<th>Details</th>
<th>Number of questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Factors(18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature-related</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Viscosity-related</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Lines of the products</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Surface of the products</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 7 Statistical Results for Key Factors

<table>
<thead>
<tr>
<th>Variable</th>
<th>Item</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Proportion of the hardener</td>
<td>4.33</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Use of Heating Facilities</td>
<td>3.97</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>Management of humidity</td>
<td>4.31</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Heating the Compound</td>
<td>3.97</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Use of Hardener</td>
<td>4.35</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>Testing the Viscosity</td>
<td>3.74</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Use of Heater</td>
<td>3.79</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>Use of Standard Mixture Proportion</td>
<td>4.49</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Temperature Measurement of the Compound</td>
<td>3.99</td>
<td>.09</td>
</tr>
<tr>
<td>Viscosity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8 Reliability Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>1st Analysis</th>
<th>2nd Analysis</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cronbach’s alpha</td>
<td>Num. of Questions</td>
<td>Cronbach’s alpha</td>
</tr>
<tr>
<td>Temperature</td>
<td>.8063</td>
<td>6</td>
<td>.824</td>
</tr>
<tr>
<td>Viscosity</td>
<td>.8370</td>
<td>4</td>
<td>.885</td>
</tr>
<tr>
<td>Line</td>
<td>.7732</td>
<td>4</td>
<td>.779</td>
</tr>
<tr>
<td>Surface</td>
<td>.7799</td>
<td>4</td>
<td>.779</td>
</tr>
</tbody>
</table>

Table 9 Factors Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor 1 (Temperature and Viscosity)</th>
<th>Factor 2 (Lines)</th>
<th>Factor 3 (Surface)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Heater</td>
<td>.945</td>
<td>.087</td>
<td>.048</td>
</tr>
<tr>
<td>Testing the Viscosity</td>
<td>.814</td>
<td>.216</td>
<td>.103</td>
</tr>
<tr>
<td>Heating the Compound</td>
<td>.813</td>
<td>.153</td>
<td>.162</td>
</tr>
<tr>
<td>Use of Heating Facilities</td>
<td>.789</td>
<td>.055</td>
<td>.226</td>
</tr>
<tr>
<td>Temperature Measurement of the Compound</td>
<td>.718</td>
<td>.256</td>
<td>.141</td>
</tr>
<tr>
<td>Use of Hardener</td>
<td>.607</td>
<td>.012</td>
<td>.289</td>
</tr>
<tr>
<td>Setting Standards for Coating</td>
<td>.475</td>
<td>.287</td>
<td>.388</td>
</tr>
<tr>
<td>Adjusting Horizontality</td>
<td>.042</td>
<td>.876</td>
<td>.121</td>
</tr>
<tr>
<td>Curing for Horizontality</td>
<td>.125</td>
<td>.738</td>
<td>.150</td>
</tr>
<tr>
<td>Setting Standards for Bending</td>
<td>.197</td>
<td>.554</td>
<td>.305</td>
</tr>
<tr>
<td>Natural Cooling after Curing</td>
<td>.177</td>
<td>.512</td>
<td>.184</td>
</tr>
<tr>
<td>Use of Mould</td>
<td>.128</td>
<td>.148</td>
<td>.789</td>
</tr>
<tr>
<td>Use of Fine Material</td>
<td>.189</td>
<td>.286</td>
<td>.643</td>
</tr>
<tr>
<td>Setting Standards for Waxing</td>
<td>.281</td>
<td>.479</td>
<td>.554</td>
</tr>
</tbody>
</table>

Figure 1 Example for Horizontal and Vertical Lines
THE EFFECTS OF EXCHANGE RATE ON THE KOREAN REGIONAL TRADE BALANCE

Dong-Su Yoo
Department of Business Administration
Namseoul University
Cheonan, Chungcheongnam-do, S.KOREA

ABSTRACT

In this paper, we used the GARCH model to analyze the effect of exchange rate on the trade balance of a province in Korea. We used monthly data of real effective Won/Dollar exchange rate, Korea's index of industrial production, and US’s index of industrial production. To analyze change of trade balance in a province of Korea in accordance with change of Won/Dollar exchange rate, this paper used monthly data of trade balance rate (TBR), REER, KIPI and UIPI, from January 2001 to December 2012. The findings of this paper are as follows. First, as Won/Dollar exchange rate rose by 1 unit, trade balance of a province in Korea increased meaningfully by 0.62 unit. Second, the effect of exchange rate and exchange rate volatility on trade balance was different in each industry. Third, In the industries of electronic & electric products, chemical products and machinery, as Won/Dollar exchange rate rise, trade balance increases meaningfully. Therefore, analysis shows that Korean provinces have to devise strategies in preparation for the fall of exchange rate.

INTRODUCTION

Korea abolished daily exchange rate fluctuating limitation in 1997 during the Asian foreign currency crisis and practically adopted freely fluctuating exchange rate system. After introduction of the freely fluctuating exchange rate system, fluctuation of exchange rate for Won currency has been largely expanded and there have been increasing number of studies on the relation between change of exchange rate and export. Those studies do not give same results but can be divided into studies that show negative effect of change of exchange rate on export, studies that show positive effect of change of exchange rate on export, and studies that show impossibility of judging the effect of change of exchange rate on export.

Considering this circumstance, this paper uses trade data of a certain region in Korea to analyze the effect of exchange rate on trade balance before and after the recent global financial crisis and derive implications for establishment of trade policies depending on change of exchange rate.

Akhtar and Hilton (1984) discovered using real exchange rate and nominal exchange rate that exchange rate volatility reduced export in the manufacturing industry, and Kenen and Rodrik (1986) discovered that exchange rate volatility reduced the amount of trade in some countries while Japan and Sweden showed larger reactions than other countries. De Grauwe (1987) showed that long term volatility in exchange rate reduced the amount of trade in these countries, and Cushman (1988) revealed that volatility of real exchange rate affected negatively two nations trade of USA. Lastrapes and Koray (1990) discovered the negative relation between the US real exchange rate and the amount of trade using the VAR model, and Kroner and Lastrapes (1993) analyzed the effect of volatility of nominal exchange rate on export and import of 6 advanced countries using the multivariate GARCH-M model. McKenzie and Brooks (1997) analyzed the effect of exchange rate volatility on trade between USA and Germany. After deriving exchange rate volatility using the ARCH model, they analyzed the effect of exchange rate volatility on German export to and import from USA. Their analysis showed that exchange rate volatility positively affected trade between these two countries.

Chit et al. (2010) discovered in the analysis of Asian 5 countries that exchange rate volatility affected export of these countries negatively. Warner and Kreinin (1983) used OLS for analysis but found no clear effect of exchange rate volatility on export, and Medhora (1990) used OLS for analysis of African new countries and could not find meaningful effect of exchange rate volatility on export. Belanger et al. (1992) could not find any meaningful effect of exchange rate volatility on export in the analysis of quarterly data of advanced countries from 1975 to 1987. Hondroyiannis et al. (2008) reported no discovery of meaningful negative effect of exchange rate volatility on export in the study of 12 advanced countries, and showed that exchange rate volatility did not reduce international trade. Ozturk and Kalyoncu (2009) used ECM for analysis and could not find any effect of exchange rate volatility on export. Boug and Fagereng (2010) used CVAR for analysis and could not
find any meaningful effect between exchange rate volatility and export. Song, Uk Heon(1997) used multi-variable GARCH-M model for analysis and concluded that judging the effect of exchange rate volatility on export was difficult. Analysis using IGARCH-M model could not find any significant result that exchange rate volatility gave to export.

In Korea, after foreign currency crisis, as volatility is largely expanded in the Won currency exchange rate due to transition to freely fluctuating exchange rate system and rapid progress of liberalization of capital movements, the relation between exchange rate volatility and export attracts much interest. In a highly foreign dependent economy as in Korea, nation's trade balance is an important factor of economy policy along with economic growth. Trade balance is such that as domestic income increases demand increases and consequently import increases, and that as the overseas relative price of imported products increases import decreases. Hence, the overseas relative price of imported products is in the negative relation to import volume, and if exchange rate rises the price of imported products increases and consequently import decreases.

If exchange rate rises, import increases due to difference of relative price while import decreases, and consequently trade balance can be improved. Also, when domestic economy rises, import of capital goods and raw materials grows and growth of domestic income due to domestic economy growth causes import growth of consumer goods, and consequently trade balance may be worsened.

**DATA AND EMPIRICAL METHOD**

To analyze change of trade balance in a province of Korea in accordance with change of Won/Dollar exchange rate, this thesis used monthly data of trade balance rate (TBR), REER, KIPI and UIPI, in this province from January 2001 to December 2012. And, trade balance rate data for each of 10 industries according to export & import items classification (MTI) was used to analyze trade balance of the 10 industries depending on change of Won/Dollar exchange rate, and the data was obtained from Korea Foreign Trade Association and FnGuide database. Logarithmic return of each rate was taken and used as each variable of this thesis, and to remove seasonal properties, each variable was adjusted over seasons by Census X12 (Eviews).

In this paper, to analyze the effect of exchange rate on the trade balance of a province in Korea, the GARCH model was used. Rose and Yellen(1989) proposed exchange rate, domestic income, and foreign income as variables affecting trade balance. This paper used monthly data of real effective Won/Dollar exchange rate, Korea's index of industrial production, and US’s index of industrial production. The function to investigate if exchange rate affects Korea's trade balance can be described as

\[
RTB = f(\text{REER, KIPI, UIPI}) \tag{1}
\]

where \(RTB\) is real trade balance of a province in Korea, REER is real Won/Dollar exchange rate, KIPI is Korea's index of industrial production, and UIPI is US's index of industrial production. The model adopted for empirical analysis made use of logarithmic return of variables adjusted over seasons. Statistical trade balance data of a region of Korea for each product item, which is required for analysis, was obtained from Korea Foreign Trade Association and FnGuide database.

This paper uses the GARCH (1,1) model, and mean equation and variance equation in the GARCH (1,1) model which considers exogenous variables. In this paper, to investigate the effect of exchange rate on trade balance of Korea, unit root test, autocorrelation test, and analysis by GARCH model were performed.

GARCH type models (Engle, 1982; Bollerslev, 1986) enable conditional modeling of dispersion with respect to past dispersion and errors unlike OLS. GARCH\((p,q)\) model has \(p\)-th order auto regression (ARCH term) and \(q\)-th order moving average (GARCH term). Engle (2001) and Hansen and Lunde (2001) claimed that GARCH\((1,1)\) model is the simplest and strongest volatility model and, therefore, this paper uses GARCH\((1,1)\) model.

Effects of Won/Dollar real exchange rate and Korea's and US index of industrial production on trade balance rate of Korean provinces is expressed in terms of coefficients of the GARCH\((1,1)\) model.

**RESULTS AND DISCUSSIONS**

The effect of change of Won/Dollar exchange rate on trade balance of a Korean province and trade balance of each industry are investigated in Table 1. Rise of Won/Dollar exchange rate turned out to increase meaningfully trade balance of the Korean province. The reason is considered to be that export of the Korean province increased much more than import and consequently export is larger than import. That is, it is considered that rise of exchange rate increases export and decreases import to improve trade balance.

The occupancy of each industry in the export from the Korean province is in the order of electronic & electric products, chemical products, machinery, mineral products, and steel & metal products. And, electronic & electric products occupies about 2/3 of the export from the Korean province, and the combined occupancy of electronic & electric products, chemical products and machinery in the export from the Korean province is as high as almost 90%. In Table 1, all 5 major export industries from Korean provinces, except mineral products, showed meaningfully positive coefficients in the period after financial crisis (2008-2012). That is, in the industries of electronic & electric products, chemical products, machinery, and steel
& metal products, as Won/Dollar exchange rate rises trade balance increases meaningfully. On the contrary, this means that as Won/Dollar exchange rate drops trade balance of the industries of electronic & electric products, chemical products, machinery, mineral products, and steel & metal products which occupy the majority of export from Korean provinces is worsened, and consequently trade balance of the Korean province may be worsened. Therefore, our result shows that the Korean province have to devise strategies in preparation for the fall of exchange rate.

CONCLUSION

This paper made an analysis of the effect of change of exchange rate on trade balance of a province in Korea under the circumstance of increasing exchange rate volatility due to introduction of freely fluctuating exchange rate system. This paper raised efficiency of analysis by applying the GARCH model which can analyze not only the effect of change of exchange rate on trade balance but also the effect of exchange rate volatility. Analysis and implications of this paper are as follows:

First, as Won/Dollar exchange rate rose by 1 unit trade balance of a korean province increased meaningfully by 0.63 unit. If the total period is divided into sub-periods, in the sub-period 1(2001-2007) before the financial crisis, as exchange rate rose by 1 unit, trade balance rate decreased by 0.66 unit and this was not meaningful statistically. But, in the sub-period 2(2008-2012) after the financial crisis, as exchange rate rose by 1 unit, trade balance rate increased statistically meaningfully by 0.49 unit. Hence, the effect of Won/Dollar real exchange rate on trade balance has increased more meaningfully recently after financial crisis. And, the effect of exchange rate volatility on trade balance of a Korean province is such that as exchange rate volatility rises by 1 unit trade balance rate increases by 6.45%, which reveals positive effect of exchange rate volatility on trade balance. The reason that rise of Won/Dollar exchange rate meaningfully increases trade balance of a Korean province is considered to be that export of Korea increases much more than import. That is, rise of exchange rate is considered to improve trade balance because increase of export is larger than the decrease of import due to fall of exchange rate.

Second, the effect of exchange rate and exchange rate volatility on trade balance was different in each industry. Only in the period after financial crisis, change of exchange rate meaningfully affected industries of electronic & electric products, chemical products, machinery, steel & metal products, and other products.

Third, electronic & electric products occupies about 2/3 of the export from the Korean province, and the combined occupancy of electronic & electric products, chemical products and machinery in the export from the Korean province is as high as almost 90%. In the industries of electronic & electric products, chemical products and machinery, as Won/Dollar exchange rate rise, trade balance increases meaningfully. On the contrary, this means that as Won/Dollar exchange rate drops trade balance of the industries of electronic & electric products, chemical products, machinery, mineral products, and steel & metal products which occupy the majority of export from Korean provinces is worsened, and consequently trade balance of Korean provinces may be worsened. Therefore, analysis shows that Korean provinces have to devise strategies in preparation for the fall of exchange rate.

REFERENCES


Arize, A. C., Osang, T. and Slottje, D. J., 2005, Exchange-rate volatility in Latin America and its impact on foreign trade."working paper.


### FIGURES AND TABLES

#### TABLE 1  EFFECTS OF EXCHANGE ON TRADE

<table>
<thead>
<tr>
<th></th>
<th>Total period</th>
<th>2001-2007</th>
<th>2008-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade balance</td>
<td>0.63**</td>
<td>-0.66</td>
<td>0.49*</td>
</tr>
<tr>
<td>Agricultural Products</td>
<td>0.26</td>
<td>1.86</td>
<td>-0.53</td>
</tr>
<tr>
<td>Mineral products</td>
<td>-0.51</td>
<td>0.32</td>
<td>-0.23</td>
</tr>
<tr>
<td>Chemical products</td>
<td>1.04***</td>
<td>0.02</td>
<td>1.40**</td>
</tr>
<tr>
<td>Plastic products</td>
<td>-0.19</td>
<td>-0.97</td>
<td>0.66</td>
</tr>
<tr>
<td>Textile products</td>
<td>-0.47</td>
<td>0.68</td>
<td>-0.64</td>
</tr>
<tr>
<td>Daily products</td>
<td>1.09</td>
<td>-0.43</td>
<td>0.67</td>
</tr>
<tr>
<td>Steel products</td>
<td>0.48</td>
<td>1.01</td>
<td>1.20*</td>
</tr>
<tr>
<td>Machinery</td>
<td>0.88</td>
<td>-1.39</td>
<td>2.01**</td>
</tr>
<tr>
<td>Electronic goods</td>
<td>-0.01</td>
<td>0.08</td>
<td>0.54*</td>
</tr>
<tr>
<td>Other goods</td>
<td>-5.54**</td>
<td>-0.71</td>
<td>-6.73*</td>
</tr>
</tbody>
</table>

*, **, *** Significant at the 10%, 5%, and 1% level, respectively.
THE STOCK MARKET INTEGRATION AMONG US, JAPAN, AND KOREA

Gu-Hyung Jung and Bu-Yeon Jo
Department of Business Administration
Namseoul University
Dallas, Texas 79409-1021, USA

ABSTRACT

In this paper, we compare and reinvestigate the efficiency of Korean stock market regarding information transmission from the US stock market and Japanese market by analyzing according to detailed time differences of impacts of the US and Japanese stock markets on Korean stock market based on TICK Data. We review whether investors can practically outperform in actual stock trading by utilizing information about stock price changes of the US and Japan. We tried to analyze integration of Korean, the US and Japanese stock. We executed Granger causality analysis to analyze short term information transmission mechanism with Korean stock market, Japanese stock market and the US stock market. We found out synchronization phenomena of Korean stock market with the US market and Japanese market became more deepened after global financial crisis. We found that if Korean investors want to earn abnormal returns by using stock returns changes of S&P500 index futures, it is only possible when they take positions as useful investment strategy just between about five minutes to 10 minutes.

INTRODUCTION

Researches on short term information transmission mechanism between international capital markets had become a momentum to implement empirical researches on the occasion of collapse of the US stock market in 1987. International capital markets related existing researches until now have attempted analysis of actual proof of information transmission mechanism between each market by using weekly or daily closing price return(close to close return) data of developed markets such as the USA, Japan and England, etc. Methods to analyze integration between stock markets can be classified into the following three methods. First, there is a research method to find actual proof evidences about short term information transmission mechanism and mutual relationships of each country’s stock market by using vector auto regression model (VAR) suggested in researches of Levy and Sarnat (1970), Eun and Shim (1989) and Jeon and VonFurstner. Second, there is a method which estimates error correction model (ECM) after confirming existence or not of common factors which moves stock markets of object countries of interest through integration verification. Its typical example is the research of Arshansapalli and Doukas(1993), Campbell and Hamao(1992), Bekaert & Harvey(1997). Third, there is a method which analyzes relationship between stock markets by analyzing phenomena, that changes of each country’s stock price are transferred among countries, by using ARCH suggested in researches of King and Sathweni(1990), Becker et al. (1990), Hamao et al.(1990), Lin et al. (1994) and GARCH model of Bollerslev which have generalized GARCH model.

In this research, we compare and reinvestigate the efficiency of Korean stock market regarding information transmission from the US stock market and Japanese market by analyzing according to detailed time differences of impacts of the US and Japanese stock markets on Korean stock market based on TICK Data of the US S&P500 index futures, Korean KOSPI200 index and index futures, Japanese NIKKEI225 index and index futures, which are all intraday stock return data and also we review whether investors can practically utilize in actual trade by utilizing information about stock price changes of the US and Japan. We tried to analyze integration of Korean, the US and Japanese stock market.

EMPIRICAL METHODOLOGY

We used intraday stock returns with same time period and intervals of each 5 minutes. Intraday stock return data used in this research were S&P500 index futures, KOSPI200 index and index futures and NIKKEI225 index and index futures which were obtained from Tick data Inc. The period is from January 2, 2009 to August 30, 2012. We have analyzed how the US stock market and Japanese stock market provide impacts on stock price changes of Korean stock market. For analysis of actual proof about short term information transmission mechanism between the US stock market, Japanese stock market and Korean stock market, we have analyzed through E-views 8 which is a time series analysis program. First of all, through ADF (Augmented Dicky-Fuller) unit root test, we have
analyzed whether data is stable. In addition, Granger causality analysis and variance decomposition analysis were conducted through VAR(vector autoregressive model) to find out whether a specific stock price index has preceding nature and predictive ability for other stock index.

VAR model is a method which makes models of over two time series concurrently by expanding a single time series model. This can be expressed as follows.

\[ KO_t = a + \sum_{t=-r}^{r} \beta_i S_{P_i} + \varepsilon_t \]  

(1)

Where \( KO_t \) means current day stock returns of KOSPI 200 and KOSPI 200 Index futures of Korean stock market, \( S_{P_i} \) means the stock returns of S&P500 Index futures of the US stock market and \( \varepsilon_t \) means errors term. If one or some parts among \( \beta_i \) are statistically meaningful, this means that S&P500 Index futures leads KOSPI200 Index.

Composition of VAR model used KOSPI 200 Index and KOSPI200 Index futures of Korean stock market, NIKKEI225 Index and Index futures of Japan and S&P500 Index futures of the USA as variables. As a null hypothesis regarding whether values of specific variables in VAR model are zero (0) or not, we used F-test statistics.

Therefore, in this research, we carried out dynamic analysis through VAR(vector auto regression model) model which includes past data of other variables in explanation variables as well as its own past data of explanation variable.

RESULTS AND DISCUSSIONS

In order to analyze short term information transmission mechanism with Korean stock market, Japanese stock market and the US stock market (information transmission mechanism), we have executed Granger causality analysis as follow by reducing frequencies from the first to the fifth. First, we have analyzed whether S&P500 index futures of the US stock market has predictive ability regarding Korean stock market as well as Japanese stock market. Second, we have analyzed whether the US stock market has predictive ability regarding Japanese stock market. Third, we have analyzed whether Japanese stock market has predictive ability about Korean stock market.

First of all, results of analysis of Granger causality with stock index return data of KOSPI200 and KOSPI200 index futures of Korean stock market and S&P500 index futures of the US stock market are shown in Table 1. F statistics was used to verify Granger causality relationship and it was found out that a null hypothesis (SPFU\( \neq \)KM) that “S&P 500 Index futures does not Granger cause to KOSPI 200 Index” (does not Granger cause) was each rejected at meaningful level with F statistics value of 1% against the null hypothesis consistently throughout all lags. Table 1 shows Granger causality relationship between S&P 500 Index futures and KOSPI 200 Index (futures).

In Table 2, results of analyses about Granger causality relationship between Japanese NIKKEI225 index and index futures and KOSPI200 index and index futures are shown. Looking into the results, we can know that the null hypothesis that “NIKKEI225 index and index futures does not Granger cause KOSPI200 index and index futures” is rejected at meaningful level of 1%. In Table 3, analyses results of Granger causality relationships between the S&P500 index futures and NIKKEI225 index and index futures are shown.

In order to find out specifically how much do the US S&P500 index futures and Japanese NIKKEI225 index and Index futures have explanation ability toward Korean KOSPI200 index and index futures with regard to distribution of forecast errors generating when all variables within VAR model estimate future values, we have analyzed through variance decomposition analysis. In Table 4, results of ten-period ahead forecasts are shown.

As results of verification analysis of Granger causality relationship using VAR model, in analyses of Japanese and Korean stock index returns using S&P 500 Index futures, it was found out that S&P500 index futures, KOSPI200 index and index futures and NIKKEI225 index and index futures have mutual feedbacks of Granger causality relationships. This can be understood that it was due to information transmission mechanism of efficiency between stock markets traded at the same time period.

In variance decomposition results, when we look at analyses of KOSPI200 index and index futures as well as NIKKEI225 index and index futures using S&P500 index futures, it can be seen that the effect of one country to another country have become larger. It is judged that analysis using intraday stock index returns instead of existing analysis using daily stock index returns is more useful in analyzing efficiency of markets. This can be also judged as results that the efficiency of information transmission mechanism of each stock market has been increased.

CONCLUSION

Summarizing above results of analyses, we can find out as results of existing researches that synchronization phenomena of Korean stock market with the US market and Japanese market became more deepened after global financial crisis and we can see that stock index returns changes of the US stock market on (t-1) day are providing impacts on stock index returns changes of Korean and Japanese stock market. In addition, we can see that there are mutual impact relationships between US, Japanese and Korean stock markets after the global financial crisis. In this paper, we tried to solve problems of non-synchronous trade using S&P500 index futures on t day as we judged it.
is difficult to judge efficiency of markets and utilization possibility of trade information since it is impossible to analyze transfer effect of stock prices between stock markets as we use data of daily stock index returns of t day and t-1 day in existing researches. Therefore, we tried to solve this issue some different from existing researches. Above all things, we could find out using S&P500 index futures that there is only about five minutes to 10 minutes impact on KOSPI200 index and index futures and NIKKEI225 index and index futures from the US S&P500 index futures. In other words, stock index returns of the US S&P500 index futures does not sufficiently impact on current KOSPI200 index and index futures, NIKKEI225 index and index futures in the same time period but there is a delayed effects about 5 or 10 minutes. Therefore, we could find out that if Korean investors want to earn excess returns by using S&P500 index futures returns, it is only possible when they take positions as useful investment strategy just between about 5 minutes to 10 minutes.

REFERENCES


Ng, Angela, 2000, Volatility spillover effects from Japan and the US to the Pacific-Basin, Journal of International Money and Finance 19, pp. 207-233.

FIGURES AND TABLES

TABLE 1 GRANGER CAUSALITY BETWEEN S&P500 INDEX FUTURES AND KOSPI200 INDEX (FUTURES)

<table>
<thead>
<tr>
<th></th>
<th>SPF→KS</th>
<th>KS→SPF</th>
<th>SPF→KSF</th>
<th>KSF→SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>lag1</td>
<td>42.172**</td>
<td>126.76**</td>
<td>31.77**</td>
<td>152.95**</td>
</tr>
<tr>
<td>lag2</td>
<td>26.783**</td>
<td>109.91**</td>
<td>23.56**</td>
<td>118.60**</td>
</tr>
<tr>
<td>lag3</td>
<td>21.296**</td>
<td>81.73**</td>
<td>20.69**</td>
<td>87.99**</td>
</tr>
<tr>
<td>lag4</td>
<td>17.296**</td>
<td>69.69**</td>
<td>15.93**</td>
<td>73.90**</td>
</tr>
<tr>
<td>lag5</td>
<td>16.038**</td>
<td>57.30**</td>
<td>15.73**</td>
<td>61.20**</td>
</tr>
</tbody>
</table>

*, ** Significant at the 5% and 1% level, respectively

TABLE 2 GRANGER CAUSALITY BETWEEN NIKKEI225 INDEX(FUTURES) AND KOSPI200 INDEX(FUTURES)

<table>
<thead>
<tr>
<th></th>
<th>NK→KS</th>
<th>KS→NK</th>
<th>NKF→KS</th>
<th>KS→NKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>lag1</td>
<td>172.39**</td>
<td>52.70**</td>
<td>501.24**</td>
<td>52.91**</td>
</tr>
<tr>
<td>lag2</td>
<td>94.22**</td>
<td>29.68**</td>
<td>262.28**</td>
<td>37.26**</td>
</tr>
<tr>
<td>lag3</td>
<td>64.17**</td>
<td>24.67**</td>
<td>177.20**</td>
<td>28.99**</td>
</tr>
<tr>
<td>lag4</td>
<td>50.48**</td>
<td>18.81**</td>
<td>135.26**</td>
<td>24.06**</td>
</tr>
<tr>
<td>lag5</td>
<td>43.94**</td>
<td>15.97**</td>
<td>110.20**</td>
<td>19.98**</td>
</tr>
</tbody>
</table>

*, ** Significant at the 5% and 1% level, respectively
### TABLE 3  GRANGER CAUSALITY BETWEEN S&P500 INDEX FUTURES AND NIKKEI225 INDEX (FUTURES)

<table>
<thead>
<tr>
<th></th>
<th>SPF→NK</th>
<th>NK→SPF</th>
<th>SPF→NKF</th>
<th>NKF→SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag1</td>
<td>2.57</td>
<td>227.66**</td>
<td>14.08**</td>
<td>225.19**</td>
</tr>
<tr>
<td>lag2</td>
<td>2.64</td>
<td>193.27**</td>
<td>11.83**</td>
<td>163.65**</td>
</tr>
<tr>
<td>lag3</td>
<td>8.76**</td>
<td>138.62**</td>
<td>15.57**</td>
<td>131.08**</td>
</tr>
<tr>
<td>lag4</td>
<td>5.89**</td>
<td>110.39**</td>
<td>11.30**</td>
<td>104.76**</td>
</tr>
<tr>
<td>lag5</td>
<td>4.46**</td>
<td>90.20**</td>
<td>8.65**</td>
<td>88.84**</td>
</tr>
</tbody>
</table>

*, ** Significant at the 5% and 1% level, respectively

### TABLE 4  FORECAST ERROR VARIANCE DECOMPOSITION

<table>
<thead>
<tr>
<th></th>
<th>KS</th>
<th>KSF</th>
<th>NK</th>
<th>NKF</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
<td>40.16</td>
<td>55.17</td>
<td>0.16</td>
<td>4.16</td>
<td>0.35</td>
</tr>
<tr>
<td>KSF</td>
<td>0.01</td>
<td>93.23</td>
<td>0.02</td>
<td>6.32</td>
<td>0.41</td>
</tr>
<tr>
<td>NK</td>
<td>0.02</td>
<td>0.17</td>
<td>54.27</td>
<td>43.94</td>
<td>0.79</td>
</tr>
<tr>
<td>NKF</td>
<td>0.01</td>
<td>0.22</td>
<td>0.07</td>
<td>99.10</td>
<td>0.60</td>
</tr>
<tr>
<td>SPF</td>
<td>0.02</td>
<td>0.21</td>
<td>0.14</td>
<td>0.52</td>
<td>99.11</td>
</tr>
</tbody>
</table>
COMOVMENTS BETWEEN ASIAN EQUITY MARKETS

Heon-Yong Jung
Department of Business Administration
Namseoul University
Cheonan, Chungcheongnam-do, S.KOREA

ABSTRACT
Stock market comovements between Asian stock markets are of great importance for the financial decisions of international investors. We investigated stock market comovements between 9 Asian stock markets in the period from 2000 to 2013. We studied comovements between Asian equity markets with VAR model, and we also examined how global financial crisis affected the comovements between Asian stock markets. The results of the analysis show that most Asian stock markets have significant influence on the other Asian stock markets. The findings show that Asian stock markets have significant lead/lag linkages. Granger causality and Variance decomposition analysis of predicted error show that comovements between Asian countries were rather strengthened. At the after global financial crisis period Asian stock markets were more interconnected to each other. The global financial crisis has decreased the diversification benefits available from country based equity market indices in Asian equity markets.

INTRODUCTION
In this paper, we study comovements between Asian equity markets with VAR model with a sample of nine Asian equity markets during the January 1, 2000-December 30, 2013 period. We divide the 14-year period into two sub-periods(January 1, 2000 – August 6, 2007 and September 8, 2008 – December 30, 2013) and we analyze the significant changes in the comovements patterns of Asian stock markets over time.

Studies on comovements between stock markets in the world began after the stock market crash in US on October 10, 1987, and the number of such studies has increased since the 2007 global financial crisis. Many studies have been performed focusing on stock markets of developed countries, and the majority of them showed that stock market globalization has been realized as stock market connection between developed countries has grown. However, these studies focused on relations between markets of developed countries. Recently, a small number of studies were performed on comovements between Asian emerging markets and major developed markets as capital liberalization was expanded. Global financial crisis in 2007 was a turning point for studies on comovements between stock markets. Scope of study on stock market comovements which had focused on developed countries was expanded to emerging stock markets.

The purpose of this paper is to analyze how comovements between Asian stock markets were changed before and after the 2007 global financial crisis. This study intends to analyze if geographically close Asian stock markets have diversified investment effect, and performs time series analysis using various VAR models in the aspect of return spillover effect proposed by previous studies on comovements. It also analyzes interdependence between Asian countries and, based on the result, derives the economic meaning on the effect of international diversified investment. For this purpose, this study analyzes comovements between nine Asian countries.

For analysis of comovements between stock markets, there is the method that vector auto regression (VAR) is used to investigate how correlation of each stock market reacts, and the method that error correction model is estimated after verifying whether or not common factors moving a target stock market exists through cointegration test. And, there is also the method that connectivity between stock markets is understood by analyzing transfer of market volatility between countries using ARCH and GARCH model. Longin and Solnik (1995), Tse and Tsui (2002), Engle (2002) showed that correlation between asset returns was changing dynamically with time. As volatility grows, comovements between markets grow, and from the viewpoint of investors or risk managers, relation between volatility and correlation coefficient is a very important factor in diversified investment or determining for risk management.

Eun and Shim (1989) made an analysis focusing on how far a country's stock price fluctuation can be explained by another country's stock price fluctuation and how quickly it is propagated to other countries. They suggested that US market affected other developed countries. Hamao, Masulis and ng(1990) studied short-term interdependence of price and price volatility in the major stock markets, and they observed the evidence supporting existence of transfer effect of

349
price volatility in the period prior to October 1987 from New York to Tokyo, from London to Tokyo, and from New York to London, and no transfer effect of price volatility in the other directions.

The paper is organized as follows. In the next section, we explain our data and methodology. We present our VAR results in section 3, and we summarize our findings and present our conclusion in section 4.

DATA AND METHODOLOGY

This study covers the following nine stock markets: Japan, South Korea, Hong Kong, Singapore, Taiwan, Malaysia, Indonesia, Thailand, and the Philippines. The daily stock market data were downloaded from FnGuide database. The daily stock returns were computed as the log difference in the index(\(\text{ln}(I_T) - \text{ln}(I_{T-1})\)).

VAR model is the method that expands the single time series model and uses simultaneous modeling of more than two time series. It is appropriate for modeling interdependence of main economic variables and particularly useful for flexibly modeling autocorrelative structure of endogenous variables. VAR model is an expansion of traditional autoregressive moving average (ARMA) model to a multivariate model, and estimates ARMA model with 2 or more variables instead of 1 variable. But the moving average part is omitted because it can make estimate of nonlinear function inevitable, and only the autoregression part is considered. If past data of other variables has explanation power for the present dependent variable, this means that independent variable precedes dependent variable. VAR model is composed of \(n\) linear regression equations, and in each equation the current measured value of each variable is set up as a dependent variable and the past measured values of itself and other variables are set up as explanatory variables. That is, VAR model considers the current measured values of all variables in the model as endogenous variables, and considers all lagged variables as exogenous variables.

In the VAR model construction, stock price indexes, such as Japan's NIKKEI225, Korea's KOSPI, Malaysia's KLSE, Indonesia's JKSE, Thailand's SET, Philippines' PSEI, Singapore's STI, Hong Kong's Hang Seng and Taiwan's TAIEX were used as variables. F-test statistics was used for null hypothesis on whether or not values of certain variables in the VAR model are all 0. Hence, in this study, dynamic analysis was performed through the VAR model in which explanatory variables contain past data of other variables and itself.

In this study, analysis period is from January 3, 2000 to December 30, 2013, and the total period was divided into two sub-periods of before and after the financial crisis. There are several opinions for the period of global financial crisis, and this study determined the period of 2007. 8. 7 – 2008. 9. 7 as the period of global financial crisis. Excluding this period, the sample period was divided to before and after the crisis. The problem of data disagreement occurs because of holidays of each country and differences of weekday stock market operation hour, and in the case of no trading in any country, data of the previous day was extended and used.

Prior to VAR model analysis, correlation analysis and unit root test were performed for each market to investigate properties and relation between markets. Stock price index in each country is an unstable process, and testing stationarity of variables prior to analysis of time series data is important. Then, result containing constant term was derived, and result containing constant term and trend simultaneously was also derived. As a result of unit root test, null hypothesis was rejected at the 1% significance level and it was verified that time series variables were stable.

Test method to verify causal relation of the functional relation in the case that cause and result are not clear is the Granger causality analysis. In this study, to analyze information delivery mechanism using stock market return data of 9 Asian stock markets, Granger causality analysis was performed at lag 1. First of all, Granger causality analysis was performed in the total sample period, and was also performed in the two sub-periods of before and after crisis.

RESULTS AND DISCUSSIONS

The result for the total period, in Table 1, shows that Japan affected Singapore only, and Korea affected Malaysia, Indonesia and Thailand. Hong Kong affected Singapore, Taiwan, Indonesia and the Philippines, Singapore affected Hong Kong and Thailand. Taiwan affected Japan, Hong Kong, Singapore, Malaysia and Thailand. Malaysia affected Hong Kong, Singapore, Indonesia and Thailand, Indonesia affected Hong Kong, Singapore, Malaysia and Thailand. Thailand affected Hong Kong, Singapore and Malaysia, the Philippines affected Japan, Hong Kong, Singapore, Malaysia, Indonesia and Thailand significantly. Test result for the period before global financial crisis, in Table 2, shows Indonesia and the Philippines, Taiwan affected Singapore and Malaysia. Malaysia affected Hong Kong and Singapore, Indonesia affected Hong Kong, Singapore, Malaysia, Thailand and the Philippines. Thailand affected Hong Kong, Singapore and Malaysia, the Philippines affected Hong Kong, Singapore, Taiwan, Malaysia and Thailand significantly. Test result for the period after global financial crisis, in Table 3, shows that Japan affected Singapore only and Hong Kong affected all Asian countries significantly. Singapore
affected Hong Kong only, and Taiwan affected Japan, Hong Kong, Singapore, Malaysia and Thailand significantly. Malaysia affected Hong Kong, Singapore, Taiwan and Indonesia significantly, Indonesia affected Hong Kong, Singapore and Thailand significantly. Thailand affected Singapore only.

As regards dispersion of predicted error that occurred when future values of variables in the VAR model were predicted, variance decomposition analysis was performed to find how much the impact to each variable can explain other variables. Table 4 displays result of variance decomposition analysis of predicted error in the twenty period ahead forecasts.

In Japan, 13.3% was affected by Hong Kong in the total period, 3.27% was affected by Korea, and 1.14% was affected by Indonesia. Before the crisis, 4.14% was affected by Hong Kong, 1.32% was affected by Korea, and 1.06% was affected by Indonesia. After the crisis, 6.60% was affected by Hong Kong, 6.09% was affected by Korea, 5.63% was affected by Singapore, 4.30% was affected by Philippines, 4.21% was affected by Indonesia, and 3.28% was affected by Malaysia. In Korea, 0.91% was affected by Hong Kong in the total period, and 1.06% was affected by Hong Kong before the crisis while after the crisis 1.30% was affected by Hong Kong, 1.34% was affected by Philippines, and 1.24% was affected by Indonesia.

In Hong Kong, 3.90% was affected by Korea in the total period, and 2.06% was affected by the Philippines, and 1.89% was affected by Korea, 1.56% was affected by Philippines, 1.22% was affected by Japan before the crisis. While after the crisis 12.4% was affected by Singapore, 8.52% was affected by Korea, 7.44% was affected by the Philippines, and 6.24% was affected by Indonesia. In Singapore, 18.1% was affected by Hong Kong in the total period, 3.87% was affected by Korea, and 3.14% was affected by Malaysia. Before the crisis, 5.95% was affected by Hong Kong, 5.29% was affected by Malaysia, and 1.78% was affected by Indonesia. After the crisis, 10.1% was affected by Hong Kong, 8.71% was affected by the Philippines, 6.20% was affected by Indonesia, 5.20% was affected by Malaysia.

The findings show that Asian stock markets have significant lead/lag linkages. Granger causality and Variance decomposition analysis of predicted error show that comovements between Asian countries were rather strengthened.

CONCLUSION

In this study, we study the comovements between nine Asian equity markets for the January 1, 2000-December 30, 2013 period. We analyzed the total sample period and two sub-periods of before and after the global financial crisis. Granger causality analysis using VAR model shows that Japan affected Singapore only while Korea affected Malaysia, Indonesia and Thailand. Korea's influence over Japan, and Asian countries grew after the crisis compared to before the crisis. Variance decomposition analysis of predicted error shows that influence of Japan over Asian countries decreased after the global financial crisis compared to before the crisis, and that comovements between Asian countries were rather strengthened. The findings show that Asian stock markets have significant lead/lag linkages. Our results indicate that after the crisis the mutual influences between Asian countries increased compared to before the crisis. We found that global financial crisis in the observed period led increases in Asian stock market comovements. At the after global financial crisis period Asian stock markets were more interconnected to each other. The global financial crisis has decreased the diversification benefits available from country based equity market indices in Asian equity markets.

REFERENCES


Jeon, B. N. and Von, F. G. M., 1990, Growing international co-movement in stock price indexes markets,
FIGURES AND TABLES

**TABLE 1 GRANGER CAUSALITY BETWEEN 9 STOCK MARKETS(TOTAL PERIOD)**

<table>
<thead>
<tr>
<th>Independent Variable – F Statistic</th>
<th>JN</th>
<th>SK</th>
<th>HK</th>
<th>SI</th>
<th>TW</th>
<th>MA</th>
<th>ID</th>
<th>TH</th>
<th>PH</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN</td>
<td>0.5</td>
<td>1.4</td>
<td>0.5</td>
<td>7.4</td>
<td>2.4</td>
<td>0.9</td>
<td>0.2</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>0.1</td>
<td>0.8</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>0.3</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HK</td>
<td>1.4</td>
<td>0.7</td>
<td>7.7</td>
<td>8.3</td>
<td>9.2</td>
<td>9.2</td>
<td>5.8</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>9.2</td>
<td>1.9</td>
<td>9.9</td>
<td>9.1</td>
<td>9.7</td>
<td>9.7</td>
<td>8.4</td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>0.1</td>
<td>1.9</td>
<td>7.2</td>
<td>2.3</td>
<td>0.1</td>
<td>2.3</td>
<td>3.0</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>1.7</td>
<td>8.7</td>
<td>0.2</td>
<td>4.8</td>
<td>6.0</td>
<td>6.1</td>
<td>5.3</td>
<td>8.2</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>2.3</td>
<td>7.9</td>
<td>7.5</td>
<td>2.0</td>
<td>2.8</td>
<td>5.6</td>
<td>2.6</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>TH</td>
<td>0.2</td>
<td>7.6</td>
<td>1.3</td>
<td>5.9</td>
<td>6.2</td>
<td>5.2</td>
<td>7.4</td>
<td>7.6</td>
<td></td>
</tr>
<tr>
<td>PH</td>
<td>3.0</td>
<td>2.6</td>
<td>7.8</td>
<td>0.0</td>
<td>0.3</td>
<td>3.0</td>
<td>3.7</td>
<td>4.4</td>
<td></td>
</tr>
</tbody>
</table>

JN: Japan, SK: South Korea, HK: Hong Kong, SI: Singapore, TW: Taiwan, MA: Malaysia, ID: Indonesia, TH: Thailand, PH: Philippines

Bold characters: Significant at the 1% level.

**TABLE 2 GRANGER CAUSALITY BETWEEN 9 STOCK MARKETS(BEFORE CRISIS PERIOD)**

<table>
<thead>
<tr>
<th>Independent Variable – F Statistic</th>
<th>JN</th>
<th>SK</th>
<th>HK</th>
<th>SI</th>
<th>TW</th>
<th>MA</th>
<th>ID</th>
<th>TH</th>
<th>PH</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN</td>
<td>0.6</td>
<td>1.9</td>
<td>0.4</td>
<td>3.4</td>
<td>0.3</td>
<td>2.3</td>
<td>0.0</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
<td>5.2</td>
<td>1.2</td>
<td>0.8</td>
<td>0.0</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>HK</td>
<td>3.9</td>
<td>1.0</td>
<td></td>
<td>6.1</td>
<td>4.2</td>
<td>4.6</td>
<td>7.6</td>
<td>4.7</td>
<td>5.3</td>
</tr>
<tr>
<td>SI</td>
<td>3.9</td>
<td>0.2</td>
<td>5.4</td>
<td>5.4</td>
<td>6.8</td>
<td>7.4</td>
<td>5.0</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>0.1</td>
<td>2.9</td>
<td>4.9</td>
<td>1.4</td>
<td>1.7</td>
<td>0.4</td>
<td>1.3</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>0.1</td>
<td>2.5</td>
<td>4.8</td>
<td>0.0</td>
<td>4.7</td>
<td>4.9</td>
<td>4.4</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>4.3</td>
<td>2.2</td>
<td>5.9</td>
<td>3.7</td>
<td>0.0</td>
<td>2.4</td>
<td>0.1</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>TH</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>3.2</td>
<td>1.8</td>
<td>5.0</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>PH</td>
<td>1.0</td>
<td>0.1</td>
<td>5.2</td>
<td>0.3</td>
<td>0.7</td>
<td>0.2</td>
<td>5.2</td>
<td>0.7</td>
<td></td>
</tr>
</tbody>
</table>

JN: Japan, SK: South Korea, HK: Hong Kong, SI: Singapore, TW: Taiwan, MA: Malaysia, ID: Indonesia, TH: Thailand, PH: Philippines

Bold characters: Significant at the 1% level.

**TABLE 3 GRANGER CAUSALITY BETWEEN 9 STOCK MARKETS(AFTER CRISIS PERIOD)**

<table>
<thead>
<tr>
<th>Independent Variable – F Statistic</th>
<th>JN</th>
<th>SK</th>
<th>HK</th>
<th>SI</th>
<th>TW</th>
<th>MA</th>
<th>ID</th>
<th>TH</th>
<th>PH</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN</td>
<td>1.3</td>
<td>5.1</td>
<td>0.8</td>
<td>4.8</td>
<td>0.6</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>0.4</td>
<td>4.4</td>
<td>0.9</td>
<td>0.6</td>
<td>0.4</td>
<td>0.6</td>
<td>0.3</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>HK</td>
<td>4.9</td>
<td>3.6</td>
<td></td>
<td>5.7</td>
<td>5.1</td>
<td>4.3</td>
<td>4.3</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>9.5</td>
<td>0.1</td>
<td>7.5</td>
<td></td>
<td>4.3</td>
<td>4.1</td>
<td>4.2</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>0.0</td>
<td>0.2</td>
<td>7.4</td>
<td>0.5</td>
<td></td>
<td>4.0</td>
<td>0.6</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>2.7</td>
<td>3.8</td>
<td>4.6</td>
<td>2.9</td>
<td>4.8</td>
<td></td>
<td>1.3</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>0.1</td>
<td>4.1</td>
<td>5.3</td>
<td>0.6</td>
<td>1.9</td>
<td>4.0</td>
<td></td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>TH</td>
<td>0.1</td>
<td>3.8</td>
<td>4.8</td>
<td>4.6</td>
<td>4.7</td>
<td>2.6</td>
<td>5.4</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>PH</td>
<td>1.9</td>
<td>0.4</td>
<td>5.1</td>
<td>0.3</td>
<td>1.6</td>
<td>2.7</td>
<td>0.2</td>
<td>1.8</td>
<td></td>
</tr>
</tbody>
</table>

JN: Japan, SK: South Korea, HK: Hong Kong, SI: Singapore, TW: Taiwan, MA: Malaysia, ID: Indonesia, TH: Thailand, PH: Philippines

Bold characters: Significant at the 1% level.

**TABLE 4 FORECAST ERROR VARIANCE DECOMPOSITION(TOTAL PERIOD)**

<table>
<thead>
<tr>
<th>Forecast error variance decomposition</th>
<th>JN</th>
<th>SK</th>
<th>HK</th>
<th>SI</th>
<th>TW</th>
<th>MA</th>
<th>ID</th>
<th>TH</th>
<th>PH</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN</td>
<td>79.7</td>
<td>3.27</td>
<td>13.3</td>
<td>0.25</td>
<td>0.50</td>
<td>0.61</td>
<td>1.14</td>
<td>0.33</td>
<td>0.68</td>
</tr>
<tr>
<td>SK</td>
<td>0.22</td>
<td>96.2</td>
<td>0.91</td>
<td>0.27</td>
<td>0.28</td>
<td>0.68</td>
<td>0.72</td>
<td>0.21</td>
<td>0.25</td>
</tr>
<tr>
<td>HK</td>
<td>0.22</td>
<td>3.90</td>
<td>91.6</td>
<td>0.22</td>
<td>0.36</td>
<td>0.42</td>
<td>0.72</td>
<td>0.20</td>
<td>0.26</td>
</tr>
<tr>
<td>SI</td>
<td>0.81</td>
<td>3.87</td>
<td>18.1</td>
<td>67.5</td>
<td>0.43</td>
<td>3.14</td>
<td>2.25</td>
<td>1.10</td>
<td>2.39</td>
</tr>
<tr>
<td>TW</td>
<td>1.72</td>
<td>2.49</td>
<td>5.60</td>
<td>0.47</td>
<td>85.5</td>
<td>0.73</td>
<td>1.18</td>
<td>0.56</td>
<td>1.17</td>
</tr>
<tr>
<td>MA</td>
<td>0.12</td>
<td>2.08</td>
<td>11.1</td>
<td>0.19</td>
<td>0.50</td>
<td>80.4</td>
<td>3.14</td>
<td>0.19</td>
<td>2.00</td>
</tr>
<tr>
<td>ID</td>
<td>0.25</td>
<td>2.29</td>
<td>8.69</td>
<td>0.15</td>
<td>0.28</td>
<td>0.74</td>
<td>85.4</td>
<td>0.43</td>
<td>1.51</td>
</tr>
<tr>
<td>TH</td>
<td>0.46</td>
<td>4.27</td>
<td>3.68</td>
<td>0.12</td>
<td>0.42</td>
<td>3.42</td>
<td>3.79</td>
<td>76.0</td>
<td>2.65</td>
</tr>
<tr>
<td>PH</td>
<td>0.33</td>
<td>1.87</td>
<td>6.70</td>
<td>0.24</td>
<td>0.28</td>
<td>0.89</td>
<td>4.19</td>
<td>0.17</td>
<td>85.0</td>
</tr>
</tbody>
</table>

JN: Japan, SK: South Korea, HK: Hong Kong, SI: Singapore, TW: Taiwan, MA: Malaysia, ID: Indonesia, TH: Thailand, PH: Philippines

**TABLE 5 FORECAST ERROR VARIANCE DECOMPOSITION(BEFORE CRISIS PERIOD)**

<table>
<thead>
<tr>
<th>Forecast error variance decomposition</th>
<th>JN</th>
<th>SK</th>
<th>HK</th>
<th>SI</th>
<th>TW</th>
<th>MA</th>
<th>ID</th>
<th>TH</th>
<th>PH</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN</td>
<td>88.7</td>
<td>1.32</td>
<td>4.14</td>
<td>0.64</td>
<td>0.65</td>
<td>0.90</td>
<td>1.06</td>
<td>0.84</td>
<td>0.96</td>
</tr>
<tr>
<td>SK</td>
<td>0.82</td>
<td>91.9</td>
<td>1.06</td>
<td>0.53</td>
<td>1.03</td>
<td>1.09</td>
<td>1.07</td>
<td>0.83</td>
<td>0.77</td>
</tr>
<tr>
<td>HK</td>
<td>1.22</td>
<td>1.89</td>
<td>89.3</td>
<td>0.40</td>
<td>0.78</td>
<td>0.85</td>
<td>2.45</td>
<td>0.50</td>
<td>1.56</td>
</tr>
<tr>
<td>SI</td>
<td>0.59</td>
<td>0.93</td>
<td>5.95</td>
<td>82.3</td>
<td>0.60</td>
<td>5.29</td>
<td>1.78</td>
<td>0.50</td>
<td>1.07</td>
</tr>
<tr>
<td>TW</td>
<td>1.27</td>
<td>1.42</td>
<td>1.85</td>
<td>86.4</td>
<td>1.64</td>
<td>1.58</td>
<td>1.20</td>
<td>2.80</td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>1.30</td>
<td>1.36</td>
<td>14.0</td>
<td>0.31</td>
<td>0.62</td>
<td>71.9</td>
<td>6.34</td>
<td>1.43</td>
<td>1.73</td>
</tr>
<tr>
<td>ID</td>
<td>1.52</td>
<td>1.20</td>
<td>4.19</td>
<td>1.10</td>
<td>0.73</td>
<td>2.08</td>
<td>86.3</td>
<td>0.63</td>
<td>1.27</td>
</tr>
<tr>
<td>TH</td>
<td>1.12</td>
<td>1.89</td>
<td>5.38</td>
<td>0.29</td>
<td>1.62</td>
<td>4.22</td>
<td>3.40</td>
<td>78.9</td>
<td>1.72</td>
</tr>
<tr>
<td>PH</td>
<td>1.18</td>
<td>1.31</td>
<td>4.06</td>
<td>0.36</td>
<td>1.32</td>
<td>2.07</td>
<td>7.78</td>
<td>1.07</td>
<td>79.9</td>
</tr>
</tbody>
</table>

JN: Japan, SK: South Korea, HK: Hong Kong, SI: Singapore, TW: Taiwan, MA: Malaysia, ID: Indonesia, TH: Thailand, PH: Philippines
<table>
<thead>
<tr>
<th></th>
<th>JN</th>
<th>SK</th>
<th>HK</th>
<th>SI</th>
<th>TW</th>
<th>MA</th>
<th>ID</th>
<th>TH</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN</td>
<td>65.8</td>
<td>6.09</td>
<td>6.60</td>
<td>5.63</td>
<td>2.19</td>
<td>3.28</td>
<td>4.21</td>
<td>1.40</td>
</tr>
<tr>
<td>SK</td>
<td>0.53</td>
<td>91.5</td>
<td>1.30</td>
<td>0.80</td>
<td>1.14</td>
<td>1.17</td>
<td>1.24</td>
<td>0.47</td>
</tr>
<tr>
<td>HK</td>
<td>0.32</td>
<td>8.52</td>
<td>56.2</td>
<td>12.4</td>
<td>0.64</td>
<td>4.08</td>
<td>6.24</td>
<td>3.68</td>
</tr>
<tr>
<td>SI</td>
<td>0.80</td>
<td>0.51</td>
<td>10.1</td>
<td>62.6</td>
<td>0.83</td>
<td>5.20</td>
<td>6.20</td>
<td>4.32</td>
</tr>
<tr>
<td>TW</td>
<td>0.32</td>
<td>4.40</td>
<td>2.69</td>
<td>3.55</td>
<td>79.0</td>
<td>1.69</td>
<td>2.36</td>
<td>1.61</td>
</tr>
<tr>
<td>MA</td>
<td>0.44</td>
<td>2.63</td>
<td>1.29</td>
<td>0.40</td>
<td>1.04</td>
<td>84.6</td>
<td>3.97</td>
<td>0.47</td>
</tr>
<tr>
<td>ID</td>
<td>0.28</td>
<td>2.68</td>
<td>0.67</td>
<td>0.68</td>
<td>0.53</td>
<td>0.89</td>
<td>85.7</td>
<td>0.83</td>
</tr>
<tr>
<td>TH</td>
<td>0.58</td>
<td>6.51</td>
<td>0.84</td>
<td>0.33</td>
<td>0.36</td>
<td>4.24</td>
<td>5.55</td>
<td>73.9</td>
</tr>
<tr>
<td>PH</td>
<td>0.32</td>
<td>3.82</td>
<td>1.06</td>
<td>0.87</td>
<td>0.32</td>
<td>0.78</td>
<td>0.59</td>
<td>0.23</td>
</tr>
</tbody>
</table>

JN: Japan, SK: South Korea, HK: Hong Kong, SI: Singapore, TW: Taiwan, MA: Malaysia, ID: Indonesia, TH: Thailand, PH: Philippines
AN EXPERIMENTAL STUDY OF WALL-PLATE U-BRACKET CONNECTION

A. L. Y. Ng, S. L. Tsok, W. H. Hii
Research Centre for Sustainable Technologies
Swinburne University of Technology
Kuching, Sarawak, Malaysia.

ABSTRACT

This project focuses on the capacity of wall-plate connection under uplift force. The connection is usually made up of U-bracket connected to the lipped C-channel at the flanges and bolted down to its support at its web. Apart from connecting the wall-plate to its support, these U-brackets are also used to level the beams so that the roof trusses that sit on the wall-plate would have consistent height. In order to achieve this, a gap was formed in between the channel and its supports. Although this type of connection has been widely used, the effect of this gap to the capacity of the connection is not known. Thus, it is the aim of this research to study the effect of the gap to the capacity of the connection through experiments. It was found that the U-brackets failed with the formation of two yield lines along the web of the U-bracket outside the bolt head. However, the presence of the gap does not affect the capacity of the U-bracket connections under uplift force. Besides, it was also found that the capacity of the U-brackets increases as the thickness of the plate increases. The estimated capacity of the U-bracket using the theoretical plastic bending capacity is conservative.

NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>Width of U-bracket, (mm)</td>
</tr>
<tr>
<td>d</td>
<td>Lever arm distance, (mm)</td>
</tr>
<tr>
<td>h</td>
<td>Thickness of U-bracket, (mm)</td>
</tr>
<tr>
<td>M_p</td>
<td>Plastic bending capacity, (Nmm)</td>
</tr>
<tr>
<td>f_y</td>
<td>Yield strength, (MPa)</td>
</tr>
</tbody>
</table>

INTRODUCTION

Functions of U-brackets

Pre-fabricated steel roof trusses are gaining popularity in the construction industry for residential or commercial structures. This is due to the fact that steel roof trusses offer numerous advantages (Mahmood et al. 2005). Steel roof trusses are made of multiple sets of geometrically stable arrangement using lipped C-channels as its members. These trusses normally sit on wall-plates which are connected to the roof beams using U-brackets. Due to the inconsistent heights of roof beams, U-brackets are also used to provide a consistent support level to the roof trusses. In cases where beams are not levelled, wall-plates are connected to the U-brackets at different heights creating gaps in between the wall-plate and its supports. Although this is common practice in the local construction industry, the effect of these gaps to the capacity of the U-bracket connections is not known.

Numerous related studies had been made before the execution of this project. These studies were not less than the buckling modes of cold-formed steel structures, behaviour of connections in shear and stripped screws behaviour. The only literature associated to this project is Ng and Mei (2010). However, these are only researches made on the behaviour of cold-formed steel anchor plate subjected to pure axial tension force. No researches had been done on the capacity of U-bracket related to the use as connectors in the installation of steel roof trusses.

Aim and Objectives

Hence, the aim of this research is to study the behaviour of cold-formed steel U-bracket connections acting as connectors in lightweight roof truss system. The scope of this research is limited to the behaviour of the U-bracket connections under uplift force. The following objectives have been developed to achieve this: (1) to design a laboratory test setup for the testing of U-bracket connections (2) to study the effect of clearance (the gaps between the wall-plate and its supports) onto the capacity of the connections (3) to study the effect of the thickness of the U-bracket plate onto the capacity of the connections (4) to compare the experimental and theoretical capacity of U-bracket using the principle of plastic bending.

EXPERIMENTAL INVESTIGATION

Experimental Setup

A test setup has been designed to study the capacity of the U-bracket connection. The schematic diagram of the test setup is as shown in Figure 1. A universal testing machine (UTM) was used to apply a point load at the midpoint of the channel, which was 500 mm away from
each end of the supports. The U-brackets were connected to the lipped C-channel using 2 self-drilling screws. They were then fastened to the support rigs using a Grade 8.8 M16 bolt. Figure 2 shows the actual experimental test setup in the laboratory.

Test Specimens

Two sets of test specimens were fabricated for this project; one set with U-brackets with clearance and one set with U-brackets without clearance. The U-brackets were cold-formed from 1.2 mm and 1.6 mm thick steel plates. The configurations of the test specimens are shown in Table 1 and are labelled accordingly. The letter W and WO refers to With Clearance and Without Clearance respectively and the number 1.2 and 1.6 refers to the thickness of the U-brackets.

Two self-drilling screws were used to connect the U-brackets to the wall-plate. The pitch, end and edge distances of the screw connection were designed according to AS4600 (Australian Standard; 2005). The arrangements of the screws on the U-bracket with and without clearance are as shown in Figure 3 and 4 respectively. The fabricated test specimens are as shown in Figure 5.

Material Properties

The material properties of the test specimens were determined from tensile coupon tests according to AS1391: 2007 (Australian Standard; 2007). Three tensile coupons were fabricated from the lipped C-channel; one from the web and one from each of the flanges. The strength of the material is taken as the average capacity of the three coupons. The yield and tensile strength of the test specimens are shown in Table 2.

TEST RIG AND OPERATION

The test setup as shown in Figure 2 was designed for the laboratory tests. The U-brackets were connected to the lipped C-channel at 1m span using self-drilling screws and bolted down to the support rigs. Four 5 mm thick Teflon pads were installed at each side of the loading rig to minimise friction between the lipped C-channel and the loading rig during the experiment. The UTM loading rate was set at 5cm per min to achieve static loading to prevent dynamic effect which might crush the lipped C-channel instantly. A linear vertical displacement transducer (LVDT) was installed at each end of the lipped C-channel to measure the vertical displacement of the U-bracket connections. A 100 mm long wood block was inserted to the lipped C-channel at the point of loading to avoid inward buckling of the channel. A bearing plate was also used to spread the load to prevent local buckling of the lipped C-channel at the point of loading.

THEORETICAL PLASTIC BENDING CAPACITY

Two yield lines were formed in the U-bracket as shown in Figure 6. Thus, the capacity of the U-bracket was estimated using the plastic bending capacity as follows:

\[
M_p = \frac{f_y}{2} \times \left( \frac{bh^2}{4} \right)
\]

\[
M_p = \frac{50 \times 1.2^2}{4} \times 585 = 10330 \text{ Nmm}
\]

The plastic bending capacity for the 1.6mm thick U-bracket is:

\[
f_y = 511.94 \text{ MPa}
\]

\[
M_p = \frac{50 \times 1.6^2}{4} \times 511.94 = 16380 \text{ Nmm}
\]

As the yield lines were formed outside the bolt head, the length of the lever arm measured was 35mm for the width of the brackets was 100 mm and the diameter of the bolt head was 30mm. The load that the U-bracket can sustain is calculated below and is tabulated in Table 3:

\[
\text{Force} = \frac{\text{Moment}}{d} \times 2 \text{ yieldlines}
\]

U-bracket 1.2mm, \(F_{1.2} = \frac{10530}{35} \times 2 = 601.71 \text{ N} \)
U - bracket 1.6mm, $F_{u} = \frac{16380}{35} \times 2 = 936.00 \text{ N}$

**TEST RESULTS**

**Experimental Observation**

It was observed that both sets of specimens behaved similarly under load. Failure started with the yield in the U-brackets followed by the deformation in the lipped C-channel as load increases. Figure 7 shows the test setup before the commencement of the test. When the U-bracket reached its ultimate strength, the C-channel was pulled down. Two vivid yield lines appeared along the web of the U-brackets. This happened as the U-brackets were more flexible than the lipped C-channel. The deformation of the U-bracket is as shown in Figure 8.

**Experimental Results**

The force that was applied onto the specimen was obtained from the UTM where else the displacement of the specimens were obtained from the external data acquisition system. The force which was applied at the midpoint of the specimen was divided by 2 to determine the value of applied load on a single U-bracket. The results were tabulated into a table and used to produce a load versus displacement graph for the U-brackets. The U-bracket was assumed to have reached its ultimate strength when there is a change in the stiffness of the graph. Two tangents were drawn on each section of constant gradient and intercepted to find the change in stiffness as shown in Figure 9. All the graphs generated from different specimens were superposed to study the relationship of the different configurations. Table 4 shows the ultimate capacity of the system and the capacity of the U-brackets obtained from the experiments conducted. The capacity shown in the table is the average capacity taken from repeated tests done on the same configuration. The difference in the test results shall not be more than 10%. Otherwise, the test will be repeated.

**DISCUSSION AND ANALYSIS**

A laboratory test setup has been designed to test the capacity of U-bracket connection in lightweight roof truss system under uplift force.

**a) Effect of Clearance on the Capacity of U-bracket**

It was found that the failure modes for both the U-bracket with and without clearance were similar. Two yield lines formed along the web of the U-bracket outside the bolt heads before the lipped C-channel was deformed. Results as shown in Table 5 demonstrate that the gap between the channel and its supports did not affect the capacity of the connection under uplift force.

**b) Effect of U-bracket Plate Thickness**

The load displacement behaviour of the test specimens is shown in Figure 10. A comparison was done between specimens W1.2, W1.6, W01.2 and W01.6. It can be seen that as the thickness of the U-bracket increases, the steeper the graph becomes. This indicates that the thicker the U-bracket, the stiffer is the connection of the system. Besides, it was also found that the total deformation of the system minimises as the thickness of the U-bracket increases. All the U-brackets failed in yielding before the failure of the channel. The ultimate load of the system was indicated by the peak of the graph.

**c) Theoretical vs. Experimental Results**

By comparing the theoretical capacity with the actual experimental capacity of the U-brackets, it was found that there was a big difference in the capacity of the U-brackets, with the actual experimental load being 30 to 40 percent higher than the theoretical load. The difference may be due to the fact that the U-bracket was assumed to have failed only when there is a change in the stiffness of the graph. Besides, the interaction between the U-bracket and the lipped C-channel was not taken into consideration. Therefore, the theoretical estimation of the capacity of the U-bracket using the principle of plastic bending capacity is conservative. Comparison between the experimental and theoretical capacity of the U-bracket is shown in Table 5.

**CONCLUSION**

The experiment had been successfully conducted using the design test setup. The failure mode of the U-brackets was the formation of yield lines in the web of the U-brackets parallel to the C-channel. Results show that the provision of clearance did not affect the capacity of U-brackets connection under uplift force. Therefore, it is not a concern that the provision of clearance due to inconsistent height of the beam might alter the capacity of the connections. However, a different mode of failure might occur under gravity load. Thus, it is suggested that further studies be conducted on the U-bracket connections under gravity load. The capacity of U-brackets increases with the thickness of the U-brackets and the estimated capacity of the U-bracket using the theoretical plastic bending capacity is conservative. The experimental capacity is higher than theoretical capacity. Hence, theoretical calculation can be used to estimate the capacity of the U-bracket.

**REFERENCES**


Ng, ALY & Mei, CC 2010, ‘The behaviour of cold-formed steel anchor plate subjected to pure axial tension force’, The fourth international conference on steel & composite structures (ICSCS’10), Sydney, Australia.

FIGURES AND TABLES

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Thickness of U-brackets (mm)</th>
<th>Thickness of C-channel (mm)</th>
<th>Provision of Clearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>W01.2</td>
<td>1.2</td>
<td>1.2</td>
<td>No</td>
</tr>
<tr>
<td>W01.6</td>
<td>1.6</td>
<td>1.2</td>
<td>No</td>
</tr>
<tr>
<td>W1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>Yes</td>
</tr>
<tr>
<td>W1.6</td>
<td>1.6</td>
<td>1.2</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 2 Yield and tensile strength of test specimens according to thickness

<table>
<thead>
<tr>
<th>Thickness (mm)</th>
<th>Yield Strength, $f_y$ (MPa)</th>
<th>Tensile Strength, $f_u$ (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>585.00</td>
<td>587.28</td>
</tr>
<tr>
<td>1.6</td>
<td>511.94</td>
<td>529.08</td>
</tr>
</tbody>
</table>

Fig. 6 Yield lines and lever arm

Table 3 Theoretical capacity

<table>
<thead>
<tr>
<th>Specimen</th>
<th>U-bracket Capacity (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WO1.2</td>
<td>601.71</td>
</tr>
<tr>
<td>WO1.6</td>
<td>936.00</td>
</tr>
<tr>
<td>W1.2</td>
<td>601.71</td>
</tr>
<tr>
<td>W1.6</td>
<td>936.00</td>
</tr>
</tbody>
</table>

Fig. 7 Test setup before load

Fig. 8 Deformed U-bracket

Fig. 9 Change in stiffness in the specimen under load

Table 4 Change in stiffness for each experiment

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Average Capacity (N)</th>
<th>Capacity of a U-bracket (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WO1.2</td>
<td>2137.50</td>
<td>1068.75</td>
</tr>
<tr>
<td>WO1.6</td>
<td>2830.00</td>
<td>1415.00</td>
</tr>
<tr>
<td>W1.2</td>
<td>2050.00</td>
<td>1025.00</td>
</tr>
<tr>
<td>W1.6</td>
<td>3000.00</td>
<td>1500.00</td>
</tr>
</tbody>
</table>

Fig. 10 Comparison of load versus displacement behaviour of test specimens of different configurations
Table 5 Comparison between the theoretical and experimental capacity of U-bracket

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Theoretical Value (N)</th>
<th>Experimental Value (N)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WO 1.2</td>
<td>601.71</td>
<td>1068.75</td>
<td>44</td>
</tr>
<tr>
<td>WO 1.6</td>
<td>936.00</td>
<td>1415.00</td>
<td>34</td>
</tr>
<tr>
<td>W 1.2</td>
<td>601.71</td>
<td>1025.00</td>
<td>41</td>
</tr>
<tr>
<td>W 1.6</td>
<td>936.00</td>
<td>1500.00</td>
<td>38</td>
</tr>
</tbody>
</table>
AN ASSET MANAGEMENT CASE STUDY IN A MALAYSIAN PUBLIC AGENCY –
MOVING TOWARDS THE ACCRUAL BASIS OF ACCOUNTING
Sharifah Sabrina bt Syed Ali
Sharon Cheuk Choy Sheung
Department of Accounting and Finance
Universiti Malaysia Sarawak, Malaysia

ABSTRACT

As part of the strategic reform of Malaysian public services under the Government Transformation Program (GTP), accrual accounting will be fully adopted in public sector financial reporting commencing 1 January 2015, in order to ensure alignment with the global accounting standards. This study examined the asset accounting system of a Malaysian public agency to evaluate the extent of compliance with MPSAS 17, Property, Plant and Equipment (PPE), IPSAS 26, Impairment of Cash-Generating Assets and IPSAS 21, Impairment of Non-Cash-Generating Assets. Findings include the following: MPSAS 17 has not been strictly adhered to and software is used to monitor the assets; however the disposal of assets is a manual process and is not automated. The study also discussed any weaknesses pertaining to the said asset accounting system, and suggested recommendations for improvement thereon.

INTRODUCTION

The rationale for accrual accounting

Accrual accounting is a basis of accounting that is most commonly used by organisations around the world. Under accrual accounting, revenues are recorded when they are earned and expenses are recorded when they are incurred. In accrual accounting there is the term called accrued revenues which is a category of an asset and accrued expenses which is a category of liability.

This is in contrast to the cash basis of accounting, whereby revenues and expenses are recognised when they are received and paid in cash, respectively. Clearly, the accrual basis of accounting is more comprehensive in capturing transactions, as not all transactions are cash-based; when adopted, it better reflects an organisation’s financial performance and position.

Under International Financial Reporting Standards this difference is summarized by IAS 37 which states:

"(a) trade payables are liabilities to pay for goods or services that have been received or supplied and have been invoiced or formally agreed with the supplier; and

"(b) accruals are liabilities to pay for goods or services that have been received or supplied but have not been paid, invoiced or formally agreed with the supplier, including amounts due to employees (for example, amounts relating to accrued vacation pay). Although it is sometimes necessary to estimate the amount or timing of accruals, the uncertainty is generally much less than for provisions.

"Accruals are often reported as part of trade and other payables, whereas provisions are reported separately."

Accrual accounting in the Malaysian public sector

The move to accrual accounting is part of the process adopting private sector accounting under the International Financial Reporting Standards. It is aimed to allow comparisons of the current cost of providing the service to the other public sector, not-for-profit, or private sectors. It is also part of the strategic reform of Malaysian public services under the Government Transformation Program (GTP), a policy measure in the New Economic Model. The New Economic Model is an economic plan that was unveiled in Malaysia in 2010, with was intended to more than double the national per capita income by 2020. According to the roadmap drawn by the Malaysian Institute of Accountants and the Malaysian Accounting Standards Board, accrual accounting will be fully adopted in public sector financial reporting commencing 1 January 2015, in order to ensure alignment with the International Public Sector Accounting Standards (IPSAS), issued by the International Federation of Accountants (IFAC), which are internationally accepted for current and future adoption. IPSAS have generally been based on the International Financial Reporting Standards (IFRS). However, IPSAS may have standards not the obtainable in IFRS and IPSAS may use different terminologies to reflect the nature of public sectors entities. Implementation of the said IPSAS will involve the review of relevant Acts including the Federal Constitution, Financial Procedures Act 1957,

The Federal Government of Malaysia’s financial statements will be prepared in accordance with the Malaysian Public Sector Accounting Standards (MPSAS), which have been adapted from IPSAS, by the year 2015. MPSASs are approved by the Government Accounting Advisory Committee and Accrual Accounting Steering Committee (AASC) under the Accountant General’s Department. As of now, eleven MPSASs have been approved by AASC as follows: MPSAS 1 – Presentation of Financial Statements, MPSAS 2 – Cash Flow Statements, MPSAS 3 – Accounting Policies, Changes in Accounting Estimates and Errors, MPSAS 4 – The Effect of Changes in Foreign Exchange Rates, MPSAS 9 – Revenue from Exchange Transactions, MPSAS 12 – Inventories, MPSAS 13- Leases, MPSAS 16 – Investment Property, MPSAS 17- Property, Plant and Equipment (PPE), MPSAS – 23 – Revenue from Non-Exchange Transactions and MPSAS 24 – Presentation of Budget Information in Financial Statements.

The adoption of accrual accounting will align the financial practices of the Malaysian public sector with that of countries such as the US, UK, New Zealand, Australia, Canada, Colombia, France, Singapore and Japan, which have adopted the full accrual basis of accounting (Abdul Khan & Mayes, 2009). In a PricewaterhouseCoopers survey published in 2013, of the 32 Organisation for Economic Co-operation and Economic Development (OECD) countries, 59% currently used accrual accounting for their central government accounts, whereas just 10% of the 68 non-OECD countries surveyed said they currently used accrual accounting, but 60% planned to do so within the next five years (Mann, 2013).

Applying accrual accounting and public sector accounting standards to property, plant and equipment

The primary objective of most public sector bodies is to deliver services to the public rather than to make profits and make a return on equity to investors. In public sector, assets are held commonly to provide services, such as military assets, transportation infrastructure, national parks and mineral reserves. Under the cash basis of accounting, fixed assets are recorded only when they are paid for (JANM, 2002). Thus, a large capital acquisition will distort expenditure upward in the first year but the usage of that asset will not be recognised in following years (Graham, 2007). Concentration on cash payments alone will sometimes result in an unnoticed deterioration in fixed assets (Graham, 2007). This will lead to failure to accurately represent the amount of resource usage. Other problems include the undervaluation of assets, as no depreciation is provided for (JANM, 2002), which then leads to a lack of an effective balance sheet to reflect the true worth (or net debt) of the organisation (Graham, 2007). Under the cash basis, not only the fixed assets are not depreciated, they are also not periodically revalued to take into account changing prices, which further compound the asset undervaluation problem. The cash basis of accounting for assets is also linked to frequent limited information about assets (Graham, 2007), leading to difficulty in effecting accountability over the said assets. A systematic approach for identifying, keeping track of, and valuing all assets and liabilities would encourage the development of systems (such as asset registers) and procedures for planning and management of assets and liabilities (Abdul Khan & Mayes, 2009). Thus, the introduction of accrual accounting can promote a general improvement in the management of assets, as well as a heightened awareness of the cost of holding and deploying assets.

The public sector accounting standards relevant to fixed assets are MPSAS 17, Property, Plant and Equipment (PPE) (Accountant General’s Department Malaysia, 2013), IPSAS 26, Impairment of Cash-Generating Assets (IFAC, 2011) and IPSAS 21, Impairment of Non-Cash-Generating Assets (IFAC, 2011). They are discussed in detail in the following paragraphs in this section.

MPSAS 17 covers the issues of asset recognition, measurement at recognition, cost or revaluation models, depreciation and depreciation methods, derecognition, disclosure and transitional provisions. Recognition of PPE should take place if future economic benefits or service potential will flow to the entity, and the cost or fair value of the item can be reliably measured reliably. PPE should be measured at cost, which comprises the net purchase price plus purchase-related expenses, including installation costs. Thereafter, PPE should be carried at cost less any accumulated depreciation and any accumulated impairment losses. If a PPE’s fair value can be reliably measured, it should be carried at a revalued amount, being its fair value at the date of the revaluation, less any subsequent accumulated depreciation, and subsequent accumulated impairment losses. Revaluations should be made with sufficient regularity to ensure that the carrying amount does not differ materially from that which would be determined using fair value at the reporting date.

In terms of depreciation, each significant PPE item should be depreciated separately, and the residual value and useful life of an asset, together with the depreciation
method, to be reviewed at least at each annual reporting date and, if expectations differ from previous estimates, the change(s) to be accounted for as a change in an accounting estimate. A PPE item is to be derecognised upon disposal or when no future economic benefit or service potential is expected from its use.

For each class of PPE, the following is to be disclosed: the measurement bases used for determining the gross carrying amount, depreciation methods used, useful lives or the depreciation rates used, gross carrying amount and the accumulated depreciation (aggregated with accumulated impairment losses) at the beginning and end of the period, and a reconciliation of the carrying amount at the beginning and end of the period showing additions, disposals, acquisitions through entity combinations, revaluation increases or decreases, impairment losses recognised or reversed, depreciation, foreign translation differences and other changes. Other required disclosures, usually in the notes to the financial statements, include the existence and amounts of restrictions on title, and PPE pledged as securities for liabilities, the amount of expenditures recognised in the carrying amount of a PPE item in the course of its construction, the amount of contractual commitments for the acquisition of PPE, and the amount of compensation from third parties for PPE items that were impaired, lost or given up.

If a class of PPE is stated at revalued amounts, the following need to be disclosed: the effective date of the revaluation, whether an independent valuer was involved, the methods and significant assumptions applied in estimating the assets’ fair values, the extent to which the assets’ fair values were determined directly by reference to observable prices in an active market or recent market transactions on arm’s length terms, or were estimated using other valuation techniques, the revaluation surplus, indicating the change for the period and any restrictions on the distribution of the balance to shareholders or other equity holders, the sum of all revaluation surpluses for individual PPE items within that class, and the sum of all revaluation deficits for individual PPE items within that class.

Transitional provisions are available for entities which are in the initial stage of adopting IPSAS accrual accounting. For instance, entities are not required to recognise PPE for reporting periods commencing within 5 years following the date of first adoption of IPSAS accrual accounting. PPE can be initially recognised at cost or fair value.

**IPSAS 21 and 26** refer to the recognition of impairment losses for PPE, whereby IPSAS 21 relate to non-cash-generating assets, while IPSAS 26 relate to cash-generating assets. Impairment is defined as a loss in the future economic benefits or service potential of an asset, over and above what has been deducted as depreciation. An asset is impaired when its carrying amount exceeds its recoverable amount or recoverable service amount; any potential impairment is assessed at each reporting date. Recoverable amount is the higher of an asset’s fair value less costs to sell and its value in use. For cash-generating assets, the value in use can be obtained by estimating the present value of future cash inflows and outflows to be derived from the continuing use of the asset and from its ultimate disposal; for non-cash-generating assets, the value in use is the present value of the remaining service potential of the asset, which can be determined using either the depreciated replacement cost approach, restoration cost approach or service units approach. An impairment loss is recognised if the recoverable amount is less than the carrying amount. Cash-generating units should be identified consistently from period to period for the same asset or types of assets, unless a change is justified. The redesignation of assets from cash-generating assets to non-cash-generating assets or vice versa would need to be performed when appropriate.

**STUDY OBJECTIVES AND RESEARCH METHODOLOGY**

The objective of this study is to examine the asset management system in a Malaysian public agency, to ascertain the extent of its compliance with MPSAS 17, IPSAS 21 and IPSAS 26. Any weaknesses are highlighted and recommendations proposed thereon.

This study adopted the qualitative case study approach due to the investigative nature of the research. Support for this approach has been noted from Miles and Huberman (1994), who used alternative analytic techniques of analysis within the case study design, and Yin (1994), who designed specific protocols for conducting the case study. Both Miles and Huberman (1994) and Yin (1994) opined that the case study approach added confidence to findings and increased study reliability. Other researchers using this approach included Meredith (1998) and Dubois and Araujo (2007); Deaconu et. al (2011) also used this approach within the public sector accounting context.

A preliminary study was conducted via interviews, following a set interview protocol which includes general instructions and a list of questions to be asked (Perry, 1998) and documents obtained prior to conducting the study were following: (a) asset management procedures; (b) process flowchart for the disposal of assets; (c) process flowchart for the write-off of movable assets; and
(d) process flowchart for loss and write-offs of assets in general.

BACKGROUND OF THE CASE STUDY

The accounting of assets is administered by the Asset Unit, under the purview of the Bursar’s office, which is staffed by 3 personnel. This division was established in 2008 under the advice of the Jabatan Audit Negara; previously, the said unit was part of the Accounts Department.

Assets are generally categorised into “General Assets” and “Movable Assets”. “General Assets” refer to property owned, in the possession or under the control of the agency, purchased, received through donations or gifts or obtained through the legislative process. “Movable Assets” refer to assets that can be transferred from one place to another, including assets that are supplied or installed together with the building. Such mobile assets are divided into two groups, namely “Capital Assets” and “Inventory”. “Capital Assets” are those with original acquisition price of RM3,000 and above per unit and with a shelf life of longer than one year, or those that require regular maintenance regardless of the original acquisition price. “Inventory” refer to movable assets with original acquisition price of less than RM3,000 per unit, with a shelf life of longer than one year and which does not require regular maintenance. “Inventory” also includes furniture and fixtures, carpeting, furnishings, curtains, tableware sets and knowledge resources regardless of the original acquisition price.

The asset management system (AMS) covers acquisition/acceptance, registration, use/storage/inspection and maintenance. In acquisition/acceptance, assets received are ensured to meet specifications, that the actual quantity and quality of the assets acquired agree with that on the acquisition order, and that the assets were received in good condition and were safe to use.

Under registration, AMS creates a database with complete, accurate and up-to-date information to facilitate detection and monitoring, know the condition of the assets and to facilitate maintenance, disposal and replacement. Under use/storage/inspection, assets are ensured that they are not used in a wasteful manner, abused or lost. Lastly under maintenance, assets are ensured to be in good condition and function smoothly.

Software is used to monitor the assets. The unit utilised Integrated Financial Accounting the System (IFAS) PeopleSoft, which was used to generate journals and the general ledger and for checking payments. The unit has recommended to the Internal Audit Department to adopt the government-recommended system Sistem Pengurusan Aset Kerajaan (SPAK), which will allow for easier monitoring of assets.

RESULTS AND FINDINGS

It was noted that every Responsibility Centre (RC) is not able to login to directly enter new assets into the system; currently, details of assets are only inputted into the Asset Management System (AMS) after payment is made. The point of recognition of an asset, therefore, is the payment date; where assets are concerned, the cash basis of accounting and not the accrual basis is followed. Hence, from the asset recognition perspective, MPSAS 17 has not been strictly adhered to. However, this is partially mitigated in the year end closing process, where any payments after the year end, in respect of asset acquisitions before the year end, are backdated so that the asset purchase is captured in that year. During this adjusting period, the IT department is requested to block the system to prevent any RC from entering next year asset acquisitions.

Depreciation is charged straight line at rates ranging from 2% to 33.3% per annum, depending on asset category, and the depreciation methods are approved by the Jabatan Audit Negara. Freehold land is not depreciated, whereas leasehold land is depreciated over the number of years of the lease. Fixed buildings are depreciated at 2% while temporary buildings are at 10%. Infrastructure is depreciated at 5%; while office equipment, fixtures and fittings, motor vehicles and machinery, teaching and research tools, sports, arts and leisure equipment, books and library equipment, telephone and communication tools, photographic and filmography tools and convocation tools are all charged at 20%. Computer and related tools are depreciated at 33.3%. Construction in progress is valued at cost and depreciation is not charged. Upon completion, construction in progress is transferred to the relevant PPE category.

As depreciation is charged, this is in line with IPSAS 27. However, PPE is not measured at fair value and impairment loss is not considered. Furthermore, revaluation was not regularly conducted, although we were informed that one valuation exercise was done recently as the result of an audit recommendation.

We were also informed that the depreciation charge for every asset is separately calculated every year and manually entered into the PeopleSoft software, as it does not have a function which automatically calculates depreciation using a given depreciation method and rate.
This may give rise to calculation errors and cause the final asset values to be under or overstated.

In terms of asset additions, we noted that very few were made in 2013 – only 7 for the year as at the date of our fieldwork. These include research-related assets such as machines and laboratory equipment. We were informed that all purchases were made in accordance with the budgets set by the respective RCs.

Disposal of assets is governed by set procedures. If the disposal relates to an ICT item, it will be handled by the IT Department; for non-ICT items, the Asset Management Department will handle it. Disposals can be effected in the following ways: tender, auction, waste sales, barter-trade, trade-in, cannibalisation, transfer, gift or destruction. Generally, disposal of assets is a manual process and not automated, and can be a long process from beginning to end.

Assets are disposed off according to the following procedures. First assets to be disposed off are identified, and their details (such as description, quantity, date of purchase, original acquisition value, and its present value) are noted on a form. The assets are then inspected and the examiners sign on the same form, which is then submitted to a disposal secretariat. The secretariat will then ensure that relevant documents are attached to the form, such as justification for the recommended method of disposal, a photo of the assets (for ICT equipment), an accident report (for a motor vehicle involved in accident), (for Non-ICT), letter of application for the method of disposal. The Treasurer is then informed. The application is then submitted to the Treasurer for approval if original asset acquisition value or total does not exceed RM350,000; to a second higher authority if value/total does not exceed RM700,000; to the highest level of authority if value/total does not exceed RM1,000,000; and to the Finance Committee if asset/total is more than RM1,000,000. Approval decision is then informed to the applicant and disposal is then executed accordingly. The final document is the Certificate of Disposal which needs to be submitted to the Treasurer's Office together with a copy of the sales receipt for the disposal method, the acknowledgment letter for the disposal method or Certificate of Destruction for assets destroyed. Lastly, the asset register records are updated.

Asset write-offs are effected when assets are no longer in existence due to theft, accident, fire, natural disaster, loss, fraud or negligence of public officers. Write-off is a process to cancel the lost asset records. Asset write-offs are intended to control the losses incurred due to asset loss coordinate the asset records, to promote awareness and responsibility for the security of the assets, and to allow surcharges and/or disciplinary action to be imposed on negligent officials.

The Treasurer's Office is the secretariat for the loss and write-offs with the cooperation of the General Administration and Security. The secretariat first obtains information about the missing assets from the Capital Asset Register or Inventory, and obtains the Preliminary Report and a copy of the police report from the RC concerned. An Investigating Committee is set up to obtain the results from the respective police investigations, and the Committee will then produce the Final Report which will be submitted for review and recommendation by the Controlling Officer. The application documents are then submitted to the Approving Authority within two (2) months from the submission date of the Preliminary Report and the police investigation report. The Approving Authority’s decision is then informed to the RC for action updates to be made to the asset register records. Surcharge or disciplinary recommendations are also made to the Disciplinary Board for further action.

We were informed that assets were usually lost because they were borrowed but not returned, and the borrowing was not informed to the RC concerned. In most cases, assets lost were laptops. We also were informed that in 70% of the asset losses, write-offs were not effected and the said assets remained in the system. This would not be in compliance with IPSAS as a PPE item should be derecognised upon disposal or when no future economic benefit or service potential is expected from its use. As such, the total PPE would be overstated in view of the continued inclusion of the lost assets in the asset listing. To mitigate this, the cost of the lost asset, at current market value, would be charged to the officer responsible; however, enforcement has not been consistent.

The PeopleSoft software is not able to generate detailed asset listings, especially depreciation charge by individual asset. Therefore, auditors would not be able to check the correctness of the depreciation calculations.

It is also noticed that assets written off and assets disposed are compiled in the same list, and need to be manually highlighted to differentiate the two categories. For the purpose of effective asset monitoring, these categories should be automatically separated.

Other general weaknesses noted include the fact that due to regular staff reshuffling, the procedures and guidelines for asset management, and any audit recommendations, may not be properly passed on to the new staff-in-charge. Also, not all staff in the Asset Unit
were aware of the compulsory adoption of IPSAS by 2015.

CONCLUSION AND RECOMMENDATIONS

The objective of this study was to examine the asset management system in a Malaysian public agency, to ascertain the extent of its compliance with MPSAS 17, IPSAS 21 and IPSAS 26, as well as to highlight any weaknesses and propose recommendations thereon.

Based on our findings, the agency under study is not in full compliance with MPSAS 17 and IPSAS 21 and 26. However, in view of the fact that there are more than 6 months to go before 1 January 2015 (at the time of writing), there is still time to rectify the current system. Furthermore, as mentioned earlier, transitional provisions are available for entities which are in the initial stage of adopting IPSAS accrual accounting. For instance, entities are not required to recognise PPE for reporting periods commencing within 5 years following the date of first adoption of IPSAS accrual accounting. Therefore the agency can make use of the intervening period to make the necessary changes in the asset management system to ensure full compliance with the relevant IPSAS’s.

The main adjustment that is needed is related to the point of recognition of PPE acquisition. The standard accounting procedures must be changed so that once an asset item is purchased and ownership is transferred to the agency, input must be made into the accounting system accordingly. Such changes must be disseminated to all relevant personnel through formal instruction and via training/workshops, and the subsequent actual actions must be monitored for compliance and accuracy thereof accordingly.

Other recommendations are as follows:

- All asset losses should be written off and excluded from the asset listing once they are confirmed to be irrecoverable.
- Physical count on assets should be made by RCs at least yearly although the bar coding system is been implemented on asset purchased.
- Rectifications that should be made include regular review of PPE for impairment; PPE should also be regularly revalued to ensure that assets are more accurately stated at fair value.
- The PeopleSoft software needs to be replaced or customised accordingly to allow for automatic calculation of depreciation, and to generate detailed asset listings for review by auditors. The capability to generate separate listings for assets disposed or written off in a year is also necessary.
- In general, all relevant staff should be well-apprised of the requirements of IPSAS, via external training in government-led workshops. More senior personnel in the accounting field could also be utilised in an advisory role.

Acknowledgement

We would like to gratefully acknowledge UNIMAS for providing us with a grant to attend this conference.

REFERENCES


Mann, N., 2013, “Governments using accrual accounting set to soar,” Retrieved April 7, 2014 from


EXAMINING THE ROLE OF INTELLECTUAL CAPITAL IN ELECTRONICS SMES IN KUCHING, MALAYSIA

Dr. Muhammad Khalique
Department of Business Management,
Faculty of Economics and Business,
Universiti Malaysia Sarawak
94300, Kota Samarahan, Malaysia

Tang Yu Chou
Student, Department of Business Management,
Faculty of Economics and Business,
Universiti Malaysia Sarawak
94300, Kota Samarahan, Malaysia

ABSTRACT

The goal of this paper is to examine the association of the components of intellectual capital with the organizational performance of electronics SMEs operating in service sector in Kuching. To achieve the objective of this study, Integrated Intellectual Capital Model was used. A valid research questionnaire was applied to gather the data from the targeted respondents. A total of 400 individuals from electrical and electronics manufacturing sector were involved. Pearson correlation technique was used to test the proposed six research hypotheses. The findings of this study show that all the six research hypotheses were supported. The findings of this study will enhance the concept and applications of intellectual capital in SMEs operating in Kuching, Malaysia.

KEYWORDS: Intellectual capital, Integrated intellectual capital model, Electronics SMEs, Malaysia

INTRODUCTION

Intellectual capital is considered as a vital asset for the success of organizations in order to achieve a competitive edge and sustainable growth. Intellectual capital refers to the accumulation of all intangible assets of the organization that contribute to enhance the organizational performance. Many researchers such as (Daud and Yusoff, 2011; de Pablos, 2005; Gallato et al., 2012; Hormiga, Batista-Canino, and Sánchez-Medina, 2011; Kamukama, Ahiauzu, and Ntayi, 2010; Khalique, 2012; Khalique, Isa, and Shaari, 2013; William S and Jasper J, 2011) argued that intellectual capital plays a significant contribution to enhance the organizational performance of organizations. There is no doubt that in a knowledge based economy, intellectual capital has secured overwhelming response from the researchers, policy makers and academicians.

In developed economies the concept and applications of intellectual capital is at maturity stage for example (Edvinsson, 1997; Stewart, 1997; Sveiby, 1997) introduced the concept of intellectual capital and identify its contribution in business performance. Bontis, Keow, and Richardson (2000) introduced the concept and the role of intellectual capital in Malaysian organizations and they found that intellectual capital is a main source of organizations to get the competitive edge. However, in developing countries the concept and applications of intellectual capital is still at the initial stage. More specifically, in Malaysian context there is a need to promote and explore the concept of intellectual capital in SMEs.

Malaysia is a fast growing developing country. Malaysian economy is mainly based on small and medium enterprises (SMEs). Small and medium enterprises (SMEs) play significant role in national economy in various forms. Hashim, Ahmad, and Zakaria (2010) stated that SMEs contribute in different sectors such as enhancing income generation, export capabilities, providing support to large companies, and reduce the rate of unemployment or poverty. According to Malaysia SMEs annual report 2011/12, SMEs accounted for 645,136 or 97.3% of the total establishment of 662,939 in manufacturing, services and agricultural sectors respectively. In 2012, it had contributed 32% of Gross Domestic Product (GDP), 56% of total workforce and approximately 19% of exports (Wong and Teo, 2013). These facts clearly showed that SMEs play remarkable contribution in the development and growth of national economy.
DEFINITION OF INTELLECTUAL CAPITAL

The term of intellectual capital was first suggested by economist John Kenneth Galbraith in the year of 1969 who in a letter given to Michael Kalecki. He reported that intellectual capital is more that only knowledge or intellect as pure intellect and it has been viewed as an approach for creating value and assets (Lopez, Munoz, Cuartas and Hoyos, 2010). Nevertheless, the idea of intellectual capital was popularized by Tom Stewart in June 1991 through his article named “Brainpower: How intellectual capital is becoming America’s most valuable asset” (Hormiga, Batista-Canino and Sanchez-Medina, 2010). After publishing this article the term intellectual capital was started to be used in the business context.

Hormiga et al. (2010) stipulated that intellectual capital is a knowledge and skill created by the workers of enterprise, information systems, financial relations and social networks both inside and outside of the organizations. Khalique (2012) argued that intellectual capital is a combination of intangible assets or resources, such as knowledge, know-how, professional skills and expertise, customer relationships, information, databases, organizational structures, innovations, social values, faith and honesty. These can be used to create organizational value and provide a competitive edge to organizations. He proposed an Integrated Intellectual Capital Model which comprises on human capital, customer capital, structural capital, social capital, technological capital and spiritual capital. The effective management of an organization can utilize these resources to create value and competitive advantages and to achieve its stated goals.

Sub-Components of Intellectual Capital

There is no doubt that intellectual capital is a multidisciplinary nature subject and therefore it is difficult to quantify accurately and propose a universally accepted definition. However, many researchers proposed different modes of intellectual capital to understand this concept. For example Stewart (1997) proposed three components namely human capital, customer capital and structural capital. Ismail (2005) argued that intellectual capital is based on human capital, customer capital, structural capital and spiritual capital. After one year Bueno, Salmador, Rodriguez, and De Castro (2006) argued that intellectual capital is accumulation of human capital, organizational capital, social capital, business capital and technological capital. Khalique, Shaari, and Isa (2011) proposed an Integrated Intellectual Capital Model which comprises on human capital, customer capital, structural capital, social capital, technological capital and spiritual capital.

Human capital

Human capital refers the work-related knowledge, skills or expertise, experiences, creativity, professional competence, talent and abilities that carried by the workers into the value adding process (de Pablos, 2005; Kamukama, Ahiauzu and Ntayi, 2010).

Customer Capital

Customer capital is refers to the customer’s satisfaction, customer’s loyalty to the organization (Khalique, Shaari, Isa, and Samad, 2013). Customer capital is a crucial component of intellectual capital and its importance in small and medium enterprises is highly acknowledged.

Structural Capital

Structural capital consists of non-human storehouse of codified knowledge within the organization such as database, management processes, corporate culture, managerial best practices, information and network system (Bontis et al., 2000).

Social Capital

Social capital is a prime component of intellectual capital. Generally social capital is based on the combination of information and experience of different parties to form the knowledge. It acts like a glue of organization (Khalique, Shaari, and Mansor, 2014).

Technological Capital

Technological capital is a main source of innovation in an organization. Khalique et al. (2014) argued that it encompasses advance technological knowledge, process, intellectual property right and industrial property rights (Khalique, 2012; Khalique et al., 2014).

Spiritual Capital

Khalique et al. (2014) argued that it includes the faith, honesty, and ethical values of employees that will help to the organization to perform effectively and efficiently. Spiritual capital is another attribute of intellectual capital and it also has significant influence on business performance. The organization with high spiritual capital is able to achieve the goals or objectives effectively and have excellent sense of holism.
CONCEPTUAL FRAMEWORK
This study is based on the Integrated Intellectual Capital Model (IICM). IICM is based on six components of intellectual capital namely, human capital, customer capital, structural capital, social capital, technological capital and spiritual capital. The conceptual framework of this study is shown in Figure 1.

RESEARCH HYPOTHESES
On the basis of Integrated Intellectual Capital Model six research hypotheses were constructed to find the objective of the study.

1. Human capital has positive association on the organizational performance of electronics SMEs in Kuching, Sarawak
2. Customer capital has positive association on the organizational performance of electronics SMEs in Kuching, Sarawak
3. Structural capital has positive association on the organizational performance of electronics SMEs in Kuching, Sarawak
4. Social capital has positive association on the organizational performance of electronics SMEs in Kuching, Sarawak
5. Technological capital has positive association on the organizational performance of electronics SMEs in Kuching, Sarawak
6. Spiritual capital has positive association on the organizational performance of electronics SMEs in Kuching, Sarawak

RESEARCH METHODOLOGY
The data for this study was collected through a valid questionnaire from targeted respondents. The respondents were selected from SMEs operating in electronics service sector in Kuching. The convenience sampling technique was used to select the respondents. The population of SMEs in services sector in Sarawak is 40,608 and the sample size that had calculated is 381 (Sekaran, 1992). Therefore, the total 500 copies of questionnaires were distributed to employees, managers or top executives of electronics SMEs respectively in order to gather the accurate data. The instrument used in this study is designed to measure intellectual capital and organizational performance. The questionnaire items were adopted from Khalique (2012). The questionnaires were personally distributed to the respondents at the workplace and got feedback.

DATA CLEANING AND SCREENING
Data cleaning or screening was involved to check the mistake or inaccuracy in the data file (Pallant, 2011). Total of 500 questionnaires were distributed to the electronics SMEs in Kuching and 433 copies were returned which represented 87% of response rate. It is at the satisfactory and effective level. Finally, the researcher only utilize a total of 400 questionnaires to conduct this present study because 33 copies were discarded due to incomplete the questionnaire form.

RELIABILITY ANALYSIS
Reliability refers to its capability to create consistent measurements or measures which free of random error. The researcher utilized Cronbach’s Alpha to evaluate whether the data collected to calculate the attributes of intellectual capital were reliably measured as suggested by Cohen and Kaimenakis (2007). Reliability represents level of correctness and unreliability which means level of incorrectness. The reliability is depended on the coefficient between 0 and 1.00 (Einspruch, 2005). This indicates that the higher coefficient represents the more reliable of test. The results of reliability test as shown in Table 1 in which the coefficient of human capital is 0.924, followed by structural capital is 0.896, relational capital is 0.898, social capital is 0.911, technological capital is 0.930, spiritual capital is 0.852 and organizational performance is 0.927, showing that this device is adequately reliable to evaluate. Technological capital has shown the highest reliability and the spiritual capital has shown the lowest reliability. However, the researcher concludes that the reliability for each dependent and independent variables were adequate and highly accepted since all the scales were greater than the proposed value which is 0.50 (Hong et al., 2008; Khalique et al., 2011).

TESTING OF HYPOTHESES
Pearson correlation technique was used to test the hypotheses of this study. The association between intellectual capital elements and organizational performance was tested through the Pearson correlation. Correlation coefficient, $r$, range from -1 to +1 is used to measure the strength of association and the direction of association either positive or negative.

Table 2 showed that the findings from the correlation analysis illustrated that the six components of the intellectual capital has positive and significant association with the performance of SMEs operating in electronics service sector in Kuching. The results show that there is a significant positive association between human capital and organizational performance ($r = 0.70, P < 0.01$), structural capital and organizational performance ($r = 0.710, P < 0.01$), customer capital and organizational
performance ($r = 0.73$, $P < 0.01$), social capital and organizational performance ($r = 0.64$, $P < 0.01$), technological capital and organizational performance ($r = 0.52$, $P < 0.01$), and spiritual capital and organizational performance ($r = 0.62$, $P < 0.01$). Overall, the results showed that the components of intellectual capital have significant association with the performance of SMEs.

**CONCLUSION AND DISCUSSION**

The aim of this study was to find the association between the components of intellectual capital and the organizational performance of SMEs operating in Kuching. The findings of this study revealed that the six components of intellectual capital have significant association with the performance of SMEs. The findings of this study are corroborating with research conducted by prior researchers in intellectual capital field such as (Bontis et al., 2000; Hsu and Fang, 2009; Khalique, Isa, et al., 2013; Khalique, Shaari, et al., 2013; Ordonez de Pablos, 2004; Tovstiga and Tulugurova, 2007). The findings of this study affirmed that the components of intellectual capital have significant association with the performance of SMEs.

**ACKNOWLEDGEMENT**

The authors acknowledge the financial support from Universiti Malaysia Sarawak.

**REFERENCES**


**FIGURES AND TABLES**

![Conceptual Framework](image)

**Source:** Integrated Intellectual Capital Model (IICM2011)

Khalique et al., (2011)

**Figure 1 Conceptual Framework**


<table>
<thead>
<tr>
<th>Variables</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Capital</td>
<td>0.924</td>
</tr>
<tr>
<td>Structural Capital</td>
<td>0.896</td>
</tr>
<tr>
<td>Relational Capital</td>
<td>0.898</td>
</tr>
<tr>
<td>Social Capital</td>
<td>0.911</td>
</tr>
<tr>
<td>Technological Capital</td>
<td>0.930</td>
</tr>
<tr>
<td>Spiritual Capital</td>
<td>0.852</td>
</tr>
<tr>
<td>Organizational Performance</td>
<td>0.927</td>
</tr>
</tbody>
</table>

**Table 1 Reliability Test of Constructs**
<table>
<thead>
<tr>
<th>Variables</th>
<th>HUC</th>
<th>STC</th>
<th>REC</th>
<th>SOC</th>
<th>TEC</th>
<th>SPC</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Capital (HUC)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Capital (STC)</td>
<td>.72**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Capital (CUC)</td>
<td>.67**</td>
<td>.70**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Capital (SOC)</td>
<td>.66**</td>
<td>.66**</td>
<td>61**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technological Capital (TEC)</td>
<td>.45**</td>
<td>.56**</td>
<td>47**</td>
<td>.56**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spiritual Capital (SPC)</td>
<td>.58**</td>
<td>.54**</td>
<td>55**</td>
<td>.51**</td>
<td>39**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Organizational Performance (OP)</td>
<td>.70**</td>
<td>.71**</td>
<td>73**</td>
<td>.64**</td>
<td>.52**</td>
<td>.62**</td>
<td>1</td>
</tr>
</tbody>
</table>

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

**Table 2 Testing of Research Hypotheses**
ORGANIZATIONAL INNOVATION AND BUSINESS PERFORMANCE:
A MALAYSIAN SERVICE INDUSTRY PERSPECTIVE

Latip, H. A, Safian, N. S. A., & Wah, C. R. J

Faculty of Economics and Business,
Universiti Malaysia Sarawak (UNIMAS)
94300 Sarawak, MALAYSIA

ABSTRACT

This study examines the interrelationships among three components of organizational innovation (i.e., process, product, and administrative innovation) and business performance, using data collected from 150 business executives of Malaysian service industries. Organizational innovation is evaluated with respect to changes, improvement, or enhancement towards process, product, and administrative. The three components of organizational innovation are found to be related significantly to business performance. The findings of this paper help to enhance an understanding on organizational innovation in service industries. The findings could also assist managers to encourage more innovation in service organization for their strategic advantage. From an applied perspective, to ensure that service industries should consider whether sufficient attention is being paid at managerial and strategic levels, organizational innovation that enhances organizational capabilities are supported fully.

Keywords: Organizational Innovation; Organization Performance; Service Industry

INTRODUCTION

There are a number of studies on innovation and business performance (Agarwal, Erramilli, & Dev, 2003; Chapman, Soosay & Kandampully 2003; Love, Roper, & Hewitt-Dundas 2010; Victorino, Verma, Plaschka & Dev, 2005) that are related to service industry. This study determines the relationship of organizational innovation and business performance in Malaysian service industry. The main purpose of this study is to explore relationships between the independent variables of organizational innovation (i.e., process innovation, product innovation and administrative innovation) with the dependent variable (business performance).

1Corresponding author.alhamrila@feb.unimas.my

This paper consists of five sections. The next section presents a brief literature review on organizational innovation and business performance in service industry. This is followed by the data and methodology used. The fourth section discusses the results, and finally the last section concludes the findings of the study.

LITERATURE REVIEW

Service has gradually increased its important to economic development. The service sectors include the transportation, education, health care, social and personal services, retail and wholesale, hotels and restaurants, telecommunication and financial sectors. In the U.S., service industries had contributed around 70 percent of the gross domestic product in 2011 (Kim, Gilmore, & Jolliff, 2012). Whilst in the UK, service sector has contributed to the growth of GNP (Figure 1) for over the last ten years (Oke, 2007).

Innovation has been identified to be one of the potential factors through which firms may gain competitive advantage in the marketplace (Kamukama, Ahiauzu, & Ntayi. 2011; Sirmon, Hitt, & Ireland, 2007). Traditionally, innovation activities of firms were mostly perceived to involve product and process. However, it is argued that an innovation in firms should not be limited in developing and applying new technologies, but it also should adopt and reorganize business routines and the internal organization (Lam, 2004; Dosi, Faillo, & Marengo, 2008). Hence, a complete picture of the concept of innovation could be extended to non-technological innovation, including organizational innovation (Schmidt & Rammer, 2007; Sundbo, 1997; Chapman, Soosay & Kandampully, 2003).

Organizational Innovation

From the literature (Crossan & Apaydin, 2010; Jung, Chow, & Wu, 2003; Lam, 2004; Yang,
Ma, Su & Moulton 2011;), organizational innovation can be referred to changes and improvement on the existing arrangement or structures that lead to new ideas and ways of work to be organized, and accomplished within an organization to encourage competitive advantage. Organizational innovation encompasses how organizations and individuals specifically manage work processes in such areas as customer relationships, employee performance and retention, and also knowledge management.

In line with Daft (1978) and Damanpour (1991), Chang and Lee (2008) viewed organizational innovation in terms of administrative innovation and management technique. Organizational innovation includes organizational production or operating systems and associated administrative or managerial processes (Damanpour et al., 2009). This type of innovation is basically related to the organizational settings involving the network actors. Elaborating on value-creation and innovation, Johannessen and Olsen (2010) suggested that firms need to search for new ways of organizing (e.g., a front-line focus on those closest to the customers).

Organizational Innovation and Business Performance

As innovation is a principal factor triggering business growth, Tsai and Goshal (1998) contended that firms that can continually develop new products, systems, and services that meet the demands of customers, are more likely to achieve success. In the current global competitive business climate, firms must maintain their innovation activities in order to sustain business growth and take a leading position in the marketplace (Chang & Lee, 2008). Despite the business scale, no innovation means a very slim chance of survival (Chang & Lee, 2008, p. 6).

Based on previous studies, it shows that there have mixed findings as to which type of innovation has impact on business performance. Lin and Chen (2007) study has found that there is positive relationship of administrative innovation with performance, compared to technological innovation (process and product innovation). Similarly, results of empirical studies on technological innovations show differences in their findings.

Hypotheses

To explore further on the relationship between different types of innovation and business performance, the relationships are hypothesized as follows:

H1: Process innovation is positively related to business performance
H2: Product innovation is positively related to business performance
H3: Administrative innovation is positively related to business performance

DATA AND METHODOLOGY

The study is based on quantitative methodology that uses survey design method. 275 questionnaires were administered to business business managers/executives of Malaysian SMEs from service sectors in February, 2013. Over a five month period, 164 questionnaires were returned by mail, yielding a response rate of 59.6%. Of the 164 respondents, 14 respondents who indicated holding non-managerial positions were excluded. Pearson Correlation was used to examine the correlations between the independent variables and dependent variable.

RESULTS

Table 1 shows the correlations analysis result for all the variables which are process innovation, product innovation, administrative innovation, and business performance.

It is shown that process innovation \( (r=0.628, p <0.01) \), has strong or positive relationship results with dependent variable which is business performance. Thus, hypothesis (H1) is accepted. In this context of study, process innovation indicates the implementation of new or significantly-improved production or delivery method.

The second independent variable, i.e., product innovation is found to have a moderate positive relationship with business performance \( (r=0.314, p<0.01) \). Hence, hypothesis H2 is accepted. In this study, product innovation is related to by-product that offered to complete the services.

Finally, administrative innovation has been found to have strong positive relationship with dependent variable \( (r=0.683, p<0.01) \). From the result, we conclude that hypothesis (H3) is accepted. An acceptance of and implementation of new procedures, policies and organizational forms are indicated in this administrative innovation.

DISCUSSION

The results show that all three types of organizational innovation are positively related with business performance in service industry. However, the strength of relationship between
product innovation and business performance is lower \((r=0.314)\), compared to the other two independent variables (process innovation, \(r=0.628\) and administrative innovation, \(r=0.683\)) accordingly. It might be due to the nature of service industry itself, where product innovation not really important rather that administrative and process innovation.

These analytical results are consistent with Yang, Ma, Su and Moulton (2011); Chang and Lee (2008), Lin and Chen (2007) the notion that organizational innovation in service industry is more related to organizational structure and administrative processes. In line with Chang and Lee (2008) who identified the organizational innovation mechanisms as administrative innovation and management techniques, this study associates organizational innovation with the settings related to the production of services, the process structures, organizational settings as well as market approach.

CONCLUSION

The present study explores organizational innovation as factors related to services that lead to business performance. Within the context of a quantitative design involving 150 responses, this study investigates the relationships of three types of organizational innovation with business performance. The significant empirical findings derived from this study are significant positive relationships between process, product and administrative innovation, and business performance, demonstrating that the organizational capability view is appropriate platform for grounding investigations of business performance. Understanding how organizational innovation affects business performance helps to enhance our understanding of organizational capabilities.

The findings of this study have implications for managers and business executives in service sectors as they integrate business processes and management techniques as routines in organizational innovation that will eventually reinforce the culture of innovation which develop dynamism and value-creation throughout their service businesses. Future studies could identify whether the implementation of organizational innovation could be heterogeneous by effect of the extent of employees’ knowledge capacity (i.e., education level, business exposure) with different service sectoral types. In addition, further research could include qualitative studies to determine how service industry integrating organizational innovation components, and what determinants that specifically contribute to its implementation.

ACKNOWLEDGEMENT

This paper is related to the RAGS grant: RAGS/e(7)/906/2012(07) and supported by the Universiti Malaysia Sarawak (UNIMAS).

REFERENCES


Table 1. Pearson Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Process Innovation</th>
<th>Product Innovation</th>
<th>Administrative Innovation</th>
<th>Business Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Innovation</td>
<td>1</td>
<td>.362**</td>
<td>.635**</td>
<td>.628**</td>
</tr>
<tr>
<td>Product Innovation</td>
<td>.000</td>
<td>1</td>
<td>.090</td>
<td>.314**</td>
</tr>
<tr>
<td>Administrative Innovation</td>
<td>.635**</td>
<td>.090</td>
<td>1</td>
<td>.683**</td>
</tr>
<tr>
<td>Business Performance</td>
<td>.628**</td>
<td>.314**</td>
<td>.683**</td>
<td>1</td>
</tr>
</tbody>
</table>

N = 150 150 150 150

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

Table 2. Hypotheses Testing

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Process innovation is positively related to business performance</td>
<td>Accepted</td>
</tr>
<tr>
<td>H2: Product innovation is positively related to business performance</td>
<td>Accepted</td>
</tr>
<tr>
<td>H3: Administrative innovation is positively related to business performance</td>
<td>Accepted</td>
</tr>
</tbody>
</table>
ABSTRACT

Prior studies argue that information costs firm’s capital due to the information asymmetry, and most of those research papers investigated developed countries. Malaysia, as an emerging market, offers its unique characteristic in terms of financial reporting regulation and is hugely influence by export-oriented firms. Therefore, this research aims to investigate whether information disclosure may affect the cost of equity of firms. We investigate this hypothesis by using all Malaysian listed firms excluding the finance, services, and utilities companies over 3 years period of 2010-2012. We use robust panel regression where the values are based on White robust standard errors that control for heterocedasticity errors. Overall, our findings are consistent with previous research that higher level of disclosure might discount the firm’s cost of equity, suggesting that firms should disclose more information for better cost of capital. At the end of our research, we explain our findings using two perspectives which are: information cost and agency cost.

INTRODUCTION

With the sophistication in the business environment, information in corporate disclosure is becoming more important to business communities. The users of financial reports are more demanding and requesting better information of the company’s performance. These users focus on the quality and timeliness of the relevant information in corporate disclosure for better decision making. The purpose is plain and simple: to attract cheaper external financing. Yoshia (1995) found that firms may prefer bilateral to multilateral financing arrangements, in order to avoid disclosure which might leak to competitors. In the presence of a cost differential between these forms of financing, the higher quality firms (those with more to lose from disclosure) prefer bilateral financing. Francis (2005) concluded that firms with greater external financing needs have higher voluntary disclosure levels, and that an expanded disclosure policy for these firms leads to a lower cost of both debt and equity capital. Additionally, Hope et al (2011) found that firms with greater financial reporting credibility experience significantly lower perceived problems in gaining access to external finance. Further, the impact of financial credibility in reducing financing constraints in the presence of a controlling owner is more pronounced in countries with weaker creditor rights. In Malaysia context, large companies ponder the impacts of the disclosure to contingency liabilities and information asymmetry issue (Hashim and Salleh, 2007; Abdullah and Ismail, 2008), especially after the Asian financial crisis in 1997 (Ghazali and Weetman, 2006). Ho and Wong (2001, 2004) surmised that the extent of disclosure and transparency information has been highlighted to aid in turning the quality of investment and external financing decision, where corporate disclosures might reduce information asymmetries among firms and outside investors. This would lead companies to have cheaper financing.

The purpose of this study is to investigate the relationship between disclosure and cost of equity. Cost of equity is important for firms as it forms part of the total cost of capital and often considered in decision making process (Cotner and Fletcher, 2000) especially decision regarding investment. According to Beneda (2003), cost of equity is important because it forms a basis of comparison in evaluating investment opportunities. Therefore, it is imperative that firms maintain their cost of equity at a reasonable level because if it is too high, the firm has to lose its prospective investment. Meanwhile, disclosure shows the information of firms about items that is included in the financial statement through notes to financial statement (Shaw, 2003). Corporate disclosure emphasizes on financial data within the framework of generally accepted accounting principle.

The association between corporate disclosure and cost of equity is related to the information asymmetry. Botosan (1997) and Botosan and Plumlee (2002) found a
negative relationship between corporate disclosure and cost of equity because that the higher disclosure reduces information asymmetry leads to a reduction in transaction costs and reduce estimation cost. This implies that corporate disclosure is important to reduce the cost of capital of firms.

This study aims to contribute to the body of knowledge about the influence of the disclosure level on the cost of equity capital. The information disclosure in the annual reports may influence the outside investors in determining the share price. This study investigates how the information disclosure is capitalised and thus influence the cost equity capital. The second contribution of this study is to investigate the influence of the firm’s characteristics towards the cost of equity capital. The different characteristics of the firms may have a different impact on the cost of equity capital. The information gathered in the study may help the management in formulating suitable management strategies regarding the firms’ future development, planning and external equity financing.

Objective

The objective of this paper is to examine the relationship between corporate disclosure and firm’s cost of equity in Malaysia.

LITERATURE REVIEW

Corporate disclosure is critical for the functioning of an efficient capital market. It has been widely used by management to communicate firm information to the outside investors. Companies supply disclosure through regulated annual reports, consisting of corporate background, summary of historical performance, key non-financial analysis, projection, financial statements, footnotes, management discussion, and other regulatory filings. In addition, several firms provide extra information, such as management forecasts, analysts’ presentations, press releases, CSR reports, union workers activities, internet sites, and other corporate reports. The purpose is simple and straightforward: to show the real condition of the corporate.

Most of the theories relating to disclosure predict a negative relationship between the disclosure and the cost of capital (i.e., Healy and Palepu, 2001; Easley and O’Hara, 2004; Barth et al., 2013). Some of the empirical findings, such as Diamond and Verrecchia (1991), Botosan (1997), Easley and O’hara (2004), and Riedl and Serafeim (2011) suggest that increased disclosure reduces firm’s cost of capital by reducing the information asymmetry. There are two streams of studies that support the negative relationship between the disclosure level and the cost of capital. The first stream is represented by Amihud and Mendelson (1986) and Diamond and Verrecchia (1991), who stated that the firm’s securities have a higher cost of equity capital with the bid-ask spreading more on asset pricing because investors demand compensation for the added transaction cost. The adverse selection component and cost of equity can be reduced by disclosing more of the firm’s information. When the investors have more precise information, they are willing to place a large order in a particular firm’s stock than they otherwise would. This will result in a high demand for the firm’s securities which will increase the firm’s stock price with a reduced cost of capital.

The second stream of theoretical research suggests that the increased disclosure can reduce estimation of risk pertaining to the parameter of the payoff distribution for a firm (Barry and Brown, 1985; Clarkson et al., 1996). Klein and Bawa (1976) were the first researchers who used the estimation risk in their study and this was then followed by Barry and Brown (1985), Handa and Lin (1993), Clarkson et al. (1996), and others.

There are relatively limited empirical studies towards the effects of the disclosure of information on the cost equity capital. Botosan (1997) examined a direct association between the disclosure level and the cost equity capital for 122 firms in the manufacturing industry. She constructed her own disclosure index to be used as a proxy for the disclosure level. She found little evidence of association between the level of information disclosure and the cost of equity capital. However, she documented that the firms that had a low analyst’s following had a strong negative association between the level of information disclosed and the cost equity capital. Hail (2002) conducted a similar study by using 27 items of disclosure by the Swiss Banking Institute as a proxy for disclosure level of information. He found a negative association between the disclosure level and the cost of equity capital for 73 non-financial firms listed on the Swiss Exchange. Botosan and Plumlee (2002) used the Association for Investment Management and Research (AIMR) disclosure rankings which were annual report disclosures, other publication disclosures (timely in nature) and investor relations activities to find evidence in the association towards cost equity capital. They found that the greater the annual report disclosures the cost equity capital decreased, but the more timely disclosure of information increased the equity capital. They did not find any evidence of association between the investor relations activities and the cost equity capital. Relating to the role of quality information, Easley and O’Hara (2004) demonstrated that the quality of information could affect the asset’s price and the cost equity capital. The more precise the public (private) information, the cost of equity capital decreased (increased) (Botosan et al., 2004).

We believe that financial reporting and disclosure will continue to be a rich field of empirical enquiry. This research takes the context of Malaysia in investigating the link between corporate disclosure and cost of equity. We revisit and extend prior research of Embong et al (2012)
that found there is a significant negative relationship between disclosure and cost of equity capital for large firms and not significant for small firms. The managers of firms could strategize the firm's disclosure policy by taking into consideration that the benefit of disclosure in reducing the cost of equity may depend on the size of the firms. Instead of taking cost of capital into our research account, we specifically investigate the cost of equity; a contribution of this research to body of knowledge.

**METHODOLOGY**

**Sample**

Our sample is listed companies in Bursa Malaysia, and it is limited only to industrial based companies. We exclude financial, services, and utilities companies because their sales are irregularly disclosed. Moreover, companies in the finance sector are governed by the Banking and Financial Institution Act and have different regulations compared to others sectors. This study only covers the companies that ended their accounting reports as of 31st December annually. We differentiate between consumer based industry and other industrial firms. We also remove any firms that have missing data throughout the nine-year period. At the end, our final sample comprises 248 firms with the total pooled observations of 744 firm years over the period of 3 years with complete data.

The share prices and interest rates, which are used to determine the cost equity capital, are retrieved from worldscope and Bank Negara Malaysia website. The disclosure level data are collected from annual reports of the companies. While the other remaining three control variables such as leverage, size and liquidity are collected from World scope and DataStream.

**Methodology**

*The cost of equity*

Following Botosan (2000), the cost of equity ($K_e$) used in this research is measured based on the CAPM. The traditional Capital Assets Pricing Model defines expected returns as the sum of the expected risk free rate, the product of a firm’s estimated risk free rate, the product of a firm’s estimated market beta and the expected market risk premium.

$$K_{e,t} = R_f + \beta R_m - R_f$$

The risk free rate is equivalent to the one year of conventional interbank interest rate announced by Bank Negara Malaysia (Malaysian the Fed). The market returns is calculated based on lognormal of firm’s prices of today divided by firm’s prices of yesterday. The market beta is taken by regressing the Bursa Malaysia index under Jensen Alpha model.

**Baseline Model**

Prior research in estimating the cost of information showed there are four factors that could affect the firm’s cost of equity, namely, leverage, size, and liquidity. The basic function is given as below.

$$K_e = f(\text{leverage}, \text{size}, \text{liquidity})$$

In measuring the firm’s size, this research used the log of Market Capitalization. Meanwhile, other control variables was developed by following previous research in cost of capital (Botosan, 1997; Botosan, 2002), where leverage was measured by interest bearing debt divided by total equity. Meanwhile, liquidity is measured by (Quick ratio): cash and equivalents plus receivables over total current liabilities. Hence, the empirical regression model is as follow.

$$K_e = \beta_0 + \beta_1 \text{Leverage}_{t,i} + \beta_2 \text{Size}_{t,i} + \beta_3 \text{Liquidity}_{t,i} + \epsilon_{t,i}$$

**Estimation Models**

This research aims to investigate the role of corporate disclosure on cost of equity. This research follows Botosan (1997) in measuring the corporate disclosure ($DISCLOSE_{t,i}$). It consists of 35 items (see Appendix), where a score of “1” is given for every item disclosed and a “0” is given for every item that is not disclosed. The choice of the use of the dichotomous procedure is based on the fact that it is currently the most appropriate measurement tool available and widely used in the financial reporting literature. The score for each item will be added and equally unweighted with the possible score for each company to derive a final score for each company. The weighted score measure has some arbitrariness issues (Cooke, 1989; Healy and Palepu, 2011). Assigning weight to a disclosure item is deemed to be subjective and furthermore, this method has also been criticized due to the fact that it is difficult to identify users’ preference for items of disclosure. There are several reasons for using a dichotomous (unweighted) disclosure score in preference to a weighted disclosure score. The dichotomous method of scoring has been used in empirical studies such as Cooke (1989) and Collett and Hrasky (2005). Interestingly, some studies found almost identical results, when weighted or unweighted methods were used to capture the disclosed information that appeared in Annual Reports (Choi, 1974; Chow and Wong-Boren, 1987). In the end, the disclosure score index is calculated as follows:

$$DISCLOSE_{t,i} = \frac{\sum \text{SCORE}}{35}$$

Therefore, we introduce the corporate disclosure on our baseline model. The final model is as follow.

$$K_e = \beta_0 + \beta_1 \text{Leverage}_{t,i} + \beta_2 \text{Size}_{t,i} + \beta_3 \text{Liquidity}_{t,i} + \beta_4 \text{Disclosure}_{t,i} + \epsilon_{t,i}$$
RESULTS AND DISCUSSIONS

Descriptive Results

Table 1 shows the summary of statistic for our sample of 248 firm across the three-year period. The mean values were calculated for each variable to facilitate comparison among firm’s average. These mean values are provided including its median and standard deviation values. We provide also the statistical test for difference in the mean value for each variable. The mean value of the cost of equity for companies was about 6%, and its median was 4.23%. This implies that the values were most likely distributed normally. Similar conclusions were found on disclosure, leverage, size, and liquidity. The means were 23.8, 20.8, 2.267, and 1.883, respectively. Meanwhile, the median values were also almost similar to its means, where there were 21.905, 19.565, 2.175, and 1.01 for disclosure, leverage, size, and liquidity, respectively. This normal distribution is supported by the value of its standard deviation where it shows the values of those variables were not much deviated from its means.

Meanwhile, our t-test shows the variables are significant different from to another. For instance, there is significant difference between cost of equity and all the variables except the leverage. Even though, there is no significant difference between leverage and cost, we still can ignore the results because cost and leverage are actually relationship between dependent variable and its regressors. There is no multicollinearity issue in this conclusion. The rest, as depicted by Table 1, it implies no multicollinearity found in this research.

Size and Disclosure

Table 2 shows the descriptive of firm’s size and its disclosure. It shows interesting findings where the percentage of disclosure and companies’ size has an inconsistent result. When the firm is a small size firm, it has tendency not to disclose all the information. This can be seen at our Table 2, where most of small firms disclosed only up to 20 items. Meanwhile, the bigger size of small firms (Rm101mil – Rm 200mil), disclose around 11 to 30 items. Only 3 firms from that cluster disclose more than 30 items. Medium size firm (Rm 201mil – Rm 300mil) normally distributed from less than 10 items, 11-20 items, 21-30 items, and 31-35 items. There is no tendency of these medium size firms to disclose or not to disclose the firm’s information. Lastly, big size firms have tendency to disclose all the information. Perhaps, these big size firms, which relatively have more leverage compared to others, have obligation to report to many stakeholders (many banks, financial institution) in many countries for the financing reporting purpose.

Estimations

Table 3 shows the results of our estimation. Referring to the R2 and adjusted R2, the regression model indicated that cost of equity was well explained by the regressors such as disclosure, leverage, size, and liquidity. The R2 and adjusted R2 were 0.223 and 0.178, respectively. This means that only 22.3 percent of the variation of the cost of equity capital in the analyzed companies was explained by the variation of disclosure level information, leverage, firm’s size, and liquidity. The F-Test concludes that the model is robust enough.

Table 3 shows also that all of control variables contribute negatively. However, it is only liquidity that has significant influence on cost of equity, but not leverage and size. The value of liquidity coefficient is also relatively big, -0.981. This means high liquidity may reduce the cost of equity of a firm. This is in line with prior research such as Diamond and Verrecchia (1991).

Corporate disclosure contributes negatively and statistically significant on the cost of equity. The coefficient value is -0.102, and its standard error is 0.209. This implies that the more a firm discloses information regarding firm’s activities, the lower the cost of equity of firms. This findings support our hypothesis that there is a negative and significant relationship between corporate disclosure and cost of equity. This result is consistent with prior research such as Botosan (1997), Botosan (2002), or in Malaysia context, it is in line with Embong et al (2012).

Robustness Test

We further investigate the role of corporate disclosure on firm’s cost of equity by separating fast moving consumer goods firms (hereafter FMCG) with non FMCG-based industrial firms. FMCG companies have a tendency to have higher cost of equity compared to non-FMCG. Therefore, we have two subsets of samples.

Table 4 shows that the R2 squares of the models, FMCG and non-FMCG, are relatively good. The R2 for FMCG model is 0.267, meanwhile non-FMCG is 0.121. The F-values of both models are also significant at 1% level. Similar with above results, the control variables do not affect cost of equity significantly except the liquidity. Liquidity contributes negatively and statistically significant to cost of equity with the coefficient value of -0.857. Corporate disclosure shows a negative and significant influence on cost of equity. This implies that for FMCG firms, higher level of disclosure may lead to lower level of cost of equity.

The same conclusion is found in non-FMCG results. The control variables are negatively related to cost of equity with the values of -0.07, -4.363, and -1.095 for leverage, size, and liquidity, respectively. Yet, it is only liquidity has significant effect on cost of equity at significant level of 10%. Corporate disclosure level of non-FMCG firms has a negative and significant influence to cost of equity. The value is -0.041 meaning that for each item increases 1, it will reduce the cost of equity to 0.041. This is tally with our previous findings. In short, even though we separate the firms of FMCG and non-
FMCG, the conclusion remains the same. The level of disclosure may lead to lower level of cost of equity.

Discussion

The results of regression show that the disclosure level of information in the annual report of consumer products industry had a significant negative relationship with cost equity capital. After we differentiate between FMCG-based industry and non FMCG-based industry, the conclusion remains the same: the corporate disclosure may lower firms cost of equity. There are two arguments to explain our finding, which are (1) information cost, and (2) agency cost perspective.

The information cost arises from information differences and conflicting incentives between principal (entrepreneur) and savers. Empirical papers, such as Akerlof (1970), Botosan (1997), and Healy and Palepu (2001) explained that when disclosure is imperfect, investors bear risks in forecasting the future payoffs from their investment. If this risk is non-diversifiable, investors will demand an incremental return for bearing the information risk. As a result, firms with high levels of disclosure, and hence low information risk, are likely to have a lower cost of capital than firms with low disclosure levels and high information risk. This explains our findings that there is negative and significant relationship between corporate disclosure and cost of equity.

In agency cost perspective, there is a consequence arises because savers that invest in a business typically do not intend to play an active role in its management that responsibility is delegated to the agent (entrepreneur). The cost arises because if savers acquire an equity stake in a firm, the agent can use those funds to make investment or operating decision that are harmful to the interest of savers. Alternatively, if savers acquire a debt stake in a firm, the entrepreneur can expropriate the value of the investment by issuing additional more senior claims, by paying out the cash received from savers as a dividend, or by taking on high risk capital projects (see Smithand Warner, 1979). To avoid this issue, savers need to reduce this agency cost by having all information from the agent. Savers will increase the cost of capital as the trade-off of feeling insecure of no full-information available. This also explains our findings that a firm with low disclosure would be punished with higher cost of equity. Note this research is consistent with Botosan (1997), Riedl and Serafaim (2011), and Embong et al (2012).

CONCLUSION

The main objective of this study is to examine the relationship between the extent of disclosure level of information in the annual reports and cost equity capital for companies listed under the consumer products industry and industrial products industry. Four hypotheses were developed in this study. First, this study examined whether there was a negative relationship between the disclosure level and the cost equity capital. Second, this study investigated whether there was a positive association between the level of leverage and the cost equity capital. Third, this study investigated if there was a negative association between the firm’s size and the cost equity capital. The last hypothesis was to examine whether there was a negative relationship between the liquidity and the cost equity capital.

The findings of this research show that the disclosure level of information in the annual reports had a negative relationship with the cost equity capital for companies listed under consumer products industry, which supported Hypothesis 1. The results showed that the disclosure level variable had a significant effect at 5% level with the cost equity capital of companies. This means that companies could enjoy lower cost equity by providing more disclosure in the annual reports. This is consistent with prior studies such as Botosan (1997), Riedl and Serafaim (2011), and Embong et al (2012). Meanwhile, most of the control variables had no significant relationship with the cost equity capital except for liquidity which had a significant negative relationship with the cost equity capital.

The results of this study confirmed that the disclosure level in the annual reports is significantly negatively related to the cost equity capital. The negative relationship is consistent with the theories and ideas that the information disclosures reduced the information asymmetry to the outside investors and managers. The firms that disclosed more also reduced uncertainty of the outside investors regarding the true parameters of the payoff distribution for the firms. This would help managements in formulating management strategies regarding the companies’ future development, planning and external equity financing.

The authors acknowledge the financial support from Universiti Malaysia Sarawak.

REFERENCES


FIGURES AND TABLES

<table>
<thead>
<tr>
<th>Mean</th>
<th>Media</th>
<th>Std. Dev</th>
<th>COST</th>
<th>DISCRO</th>
<th>LEVERAGE</th>
<th>SIZE</th>
<th>LIQUIDITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0088</td>
<td>4.235</td>
<td>8.82402</td>
<td>-3.51**</td>
<td>-0.13</td>
<td>0.059</td>
<td>0.116</td>
<td>0.18</td>
</tr>
<tr>
<td>23.8095</td>
<td>21.9048</td>
<td>10.45285</td>
<td>-0.03</td>
<td>0.016</td>
<td>0.131</td>
<td>0.209</td>
<td>0.133</td>
</tr>
<tr>
<td>19.68341</td>
<td>19.565</td>
<td>11.31392</td>
<td>(0.358)</td>
<td>(0.001)</td>
<td>(0.15)</td>
<td>(0.135)</td>
<td>(0.458)</td>
</tr>
<tr>
<td>3.2627</td>
<td>2.175</td>
<td>0.57908</td>
<td>-3.99**</td>
<td>-0.13</td>
<td>0.116</td>
<td>0.18</td>
<td>0.133</td>
</tr>
<tr>
<td>1.8825</td>
<td>1.01</td>
<td>2.58542</td>
<td>(0.042)</td>
<td>(0.190)</td>
<td>(0.237)</td>
<td>(0.458)</td>
<td>(0.458)</td>
</tr>
</tbody>
</table>
Table 2
Size and its Corporate Disclosure
Size is calculated based on market capitalization. We divided into 5 level of sizes, and 4 types of disclosure items. Note that there are 35 item of disclosure in total for our content analysis of corporate disclosure.

<table>
<thead>
<tr>
<th>Size Range</th>
<th>0-10 Items</th>
<th>11-20 Items</th>
<th>21-30 Items</th>
<th>31-35 Items</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM 1 mil - RM 100 mil</td>
<td>13</td>
<td>55</td>
<td>11</td>
<td>9</td>
<td>88</td>
</tr>
<tr>
<td>RM 100 mil - RM 200 mil</td>
<td>4</td>
<td>18</td>
<td>21</td>
<td>3</td>
<td>46</td>
</tr>
<tr>
<td>RM 200 mil - RM 300 mil</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>RM 300 mil - RM 400 mil</td>
<td>0</td>
<td>6</td>
<td>12</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>&gt; RM 400 mil</td>
<td>0</td>
<td>13</td>
<td>29</td>
<td>27</td>
<td>69</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>21</td>
<td>99</td>
<td>82</td>
<td>46</td>
<td>248</td>
</tr>
</tbody>
</table>

Table 3
Model Estimations
The regression is performed using panel regression. The figure stated are the coefficient values, except numbers in parentheses which are standard error. The dependent variable is the cost of equity value of firms. The control variables are leverage, size, and liquidity. The main independent variable is corporate disclosure. The model is as follow:

\[ K = \beta_0 + \beta_1 \text{Leverage} + \beta_2 \text{Size} + \beta_3 \text{Liquidity} + \beta_4 \text{Disclosure} + \epsilon \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCORE</td>
<td>-0.102**</td>
<td>[0.209]</td>
</tr>
<tr>
<td>LEVERAGE</td>
<td>-0.04</td>
<td>[0.011]</td>
</tr>
<tr>
<td>SIZE</td>
<td>-1.537</td>
<td>[1.794]</td>
</tr>
<tr>
<td>LIQUIDITY</td>
<td>-0.981**</td>
<td>[0.322]</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>18.271***</td>
<td>[7.662]</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.225</td>
<td></td>
</tr>
<tr>
<td><strong>adjusted R^2</strong></td>
<td>0.178</td>
<td></td>
</tr>
<tr>
<td>F-Value</td>
<td>3.12***</td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Robustness Test: FMCG and non-FMCG
The regression is performed using panel regression. The figure stated are the coefficient values, except numbers in parentheses which are standard error. The dependent variable is the cost of equity value of firms. The control variables are leverage, size, and liquidity. The main independent variable is corporate disclosure. The model is as follow:

\[ K = \beta_0 + \beta_1 \text{Leverage} + \beta_2 \text{Size} + \beta_3 \text{Liquidity} + \beta_4 \text{Disclosure} + \epsilon \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCORE</td>
<td>-0.292**</td>
<td>[0.159]</td>
</tr>
<tr>
<td>LEVERAGE</td>
<td>0.01</td>
<td>[0.015]</td>
</tr>
<tr>
<td>SIZE</td>
<td>-2.205</td>
<td>[1.874]</td>
</tr>
<tr>
<td>LIQUIDITY</td>
<td>-0.857**</td>
<td>[0.414]</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>17.805***</td>
<td>[4.061]</td>
</tr>
<tr>
<td>FMCG</td>
<td><strong>R^2</strong></td>
<td>0.267</td>
</tr>
<tr>
<td><strong>adjusted R^2</strong></td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>F-Value</td>
<td>3.12***</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCORE</td>
<td>-0.041**</td>
<td>[0.171]</td>
</tr>
<tr>
<td>LEVERAGE</td>
<td>-0.07</td>
<td>[0.051]</td>
</tr>
<tr>
<td>SIZE</td>
<td>-4.363</td>
<td>[3.143]</td>
</tr>
<tr>
<td>LIQUIDITY</td>
<td>-1.095*</td>
<td>[0.615]</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>20.373***</td>
<td>[7.262]</td>
</tr>
<tr>
<td>Non-FMCG</td>
<td><strong>R^2</strong></td>
<td>0.121</td>
</tr>
<tr>
<td><strong>adjusted R^2</strong></td>
<td>0.102</td>
<td></td>
</tr>
<tr>
<td>F-Value</td>
<td>2.062***</td>
<td></td>
</tr>
</tbody>
</table>
REVISITING STOCK MARKET INTEGRATION PRE-MORTGAGE CRISIS: INSIGHT FROM BRIC COUNTRIES

Kai-Hung Wong, Chin-Hong Puah, Rayenda Khresna Brahmana
Faculty of Economics and Business
Universiti Malaysia Sarawak
Kota Samarahan, Sarawak 94300, Malaysia

ABSTRACT
This study examines long-run relationships and short-run dynamic causal linkages among BRIC countries, with the particular attention to the 2008 subprime mortgage crisis. Extending related empirical studies, comparative analyses of pre-crisis, and post-crisis periods were conducted to comprehensively evaluate how stock market integration was affected by financial crises. In general, after employing cointegration test and VAR test, the results reveal the increase of stock market integration in BRICs after the subprime crisis. The evidence also found that China stock market is the most influential among the BRICs, in which the China stock market has the ability to Granger cause the other three BRICs member countries. An important implication of our findings is that the degree of integration among countries tends to change over time, especially around periods marked by financial crises.

INTRODUCTION
One of the core questions in international finance is about market integration, which is important to both market participants and policymakers. In integrated markets, capital flows freely to where it will generate the highest return. Integrated financial markets have easier access to foreign capital, but are also more vulnerable to financial crises occurring in other areas of the world. Moreover, any increase in the degree of global financial market integration decreases the opportunity for diversification. It is thus essential to achieve a better understanding of the factors driving financial market integration.

Recent years have seen considerable attention devoted to the analysis of linkages among stock markets in different countries, especially the impact of crisis to market integration. For instance, Click and Plummer (2010) documented the high integration of market after stock market crash after financial crisis 1997. This issue is an important concern for investors because greater integration among stock markets implies that reducing the opportunities for international diversification. The loss of international portfolio benefit because of crisis may attract the attention of global investors. However, interest in this topic has also been enhanced by various conclusions among integration reference and also the fact about integrated market. Based on our knowledge, it is hardly found a research that investigating emerging market integration due to subprime mortgage; a gap which this research aims to tackle.

Prior research provides evidence that fully integrated markets, only global risks are priced (e.g. Solnik, 1974a; Sercu, 1980; Adler and Dumas, 1983) while in segmented markets only local risks are priced. Several empirical papers documented that national equity markets are becoming more integrated within the world market (e.g. De Jong and De Roon, 2005; Carrieri et al., 2007; Pukthuanthong-Le and Roll, 2008). This suggests that international equity returns are increasingly driven by global rather than by local factors. Others, such Click and Plummer (2010), and Claessens et al (2010), provide the dossiers of crisis impact on market integration.

In BRIC countries context, the issue of market integration shows various conclusion. For instance, Chittedi (2009) found there was a unidirectional causality from India to Russia and Brazil, and found the existence of cointegration relationship between BRICs and developed countries. This is supported by Sheu and Liao (2011). However, An and Brown (2010) had different opinions where none of the BRIC countries had a level of cointegration with developed countries. This is in line with Syamala and Wadhwa (2012) who found the integration between US and BRICs markets were not strong. Our Figure 1 shows also there is co-movements among BRIC’s stock markets, where the co-movements became more stronger after the subprime mortgage in 2008.

According to Bartlett (2008), BRICs had experienced a moderate to strong growth due to the greater domestic consumption and purchasing power, expanding the trade among the BRICS countries and other emerging market, and continued inflow of foreign direct investments. On the contrary, the western economies experienced slower economic growth which had resulted many of the investors were headed to BRIC countries after the subprime mortgage crisis. As such, international portfolio investment in BRICs after the subprime mortgage crisis...
has become an increasingly important for international investors. With that, a study of stock market integration in BRICs after the crisis is needed in order to provide clearer direction for investor whether the subprime mortgage crisis has caused the emerging market of BRICs to be more cointegrated. Does it sensible for investor to diversify among the BRIC countries at the same time? In other words, do BRICs stock markets interdependent after the subprime crisis? Therefore, this research aims to revisit the long run integration relationship among the stock markets of the BRICs. Further, this research also aims to analyze the short run causality among the stock markets of the BRICs and the possibility of the portfolio diversification among the BRICs.

The significance of the study of this topic is to provide a better view for international investors on the existence of long run cointegration relationship among the stock markets in BRICs especially after the subprime crisis, where it is investigated through an examination of the interrelationship and structure of linkages among the stock markets in BRICs. On the other hand, the direction of causality also been tested to see the influences on the stock markets integration among the BRICs in the short run in order to convey the significant policy implications to the relevant parties in making the portfolio diversification strategy decision.

Objective

The objective of this paper is to investigate the market integration among BRIC countries before and after the subprime mortgage crisis.

LITERATURE REVIEW

Due to its demographic and economic development, the BRIC countries are rank among the world largest and most influential economies. The emerging market economy plays an increasingly important role in global economic development as well as monetary and financial systems. After the financial crises in Asia and Russia in 1998, Turkey and Brazil in 1999, and Argentina in 2001, the financial potential and economic development of the emerging markets especially the BRIC countries were focused sharply by the investor (Jensen & Larsen, 2004).

The growth of stock market in BRICs has brought up an interest to the policymakers by expanding the financing options availability. As fast-growing emerging market, BRIC countries are one of the options to have portfolio gain for global investors. Therefore, literature regarding the integration of BRIC countries is interesting and important in international finance study (see the review of Chittedi, 2010).

The fluctuation of BRIC’s economy may result a cost of portfolio benefit. Crisis for instance, may reduce the portfolio benefit because nowadays one country is closely related to other countries activity. Gwinner et al (2008) conclude that the subprime crisis has huge impact to BRIC countries. This is supported by other empirical papers such Kregel (2009), Ghosh and Chandrasekhar (2009), and Claessens et al (2010). Kregel (2009), for instance, documented the global crisis might affect the economic development of BRIC countries and loosen the persona of their international diversification benefits. Ghosh and Chandrasekhar (2009) found that there is a cost of coupling of global crisis to Indian economy. In a wider sample, Claessens et al (2010) showed the role of financial crisis to BRIC economy, and each BRIC countries have different impact of financial crisis, especially the subprime mortgage crisis. One of the examples, Brazilian stock market, declined 50% in 6 months during the subprime mortgage crisis tragedy, and needed 2 years of recovery. This is similar with Indian stock markets, where it was very sensitive to the financial crisis. During the subprime mortgage, Indian stock markets declined rapidly up to more than 50%, yet expanded back doubling the index after adapting to the crisis condition. However, it was hit back again due to the European sovereign debt crisis, and contracted again to weaken 30% in 2 months. Different with the previous 2, Russian and China stock markets were more prudent to crisis. Even though the market dropped for 6 months, however, due to its reserves, these two stock markets recovered rapidly. These dossiers portray that each BRIC countries have different persistence of handling and disseminating subprime mortgage crisis. Hence, it is an intriguing question to know how a crisis would affect BRIC’s stock market and its integration to other stock markets during the subprime mortgage.

Recent literatures (i.e Bhar and Nikolova, 2009; Chittedi, 2009; Kenourgios et al, 2011) concluded the integration among BRICs countries become stronger and stronger due to their financial development and economic liberalization. However, due to the subprime mortgage crisis in 2008, the integration had getting weaken because each BRIC country has different and unique relationship to develop countries (see Gwinner et al, 2008; Kregel, 2009; Ghosh and Chandrasekhar, 2009; Claessens et al, 2010).

Chittedi (2009) investigated the long run equilibrium relationship among the BRICs and three major developed countries (US, UK and Japan). The daily stock market indexes from Nasdaq, FTSE-100, Nikkei-225, BOVESPA, Moscowtimes, Sensex and Shanghai Stock Exchange ranged from January 1998 to August 2008 were used. He discovered that there was a unidirectional causality from India to Russia and Brazil, and found the existence of cointegration relationship between BRICs and developed countries (see also Sheu and Liao, 2011). Sheu and Liao (2011) further documented that in the short run, there was a bilateral causality between US with Russia and China but unilateral causality from US to India and from Brazil to US.
However, An and Brown (2010) had different opinion. Based on the sample periods from October 1995 to October 2009, they noticed that apart from China, none of the BRIC countries had a level of cointegration with the US economy. The result was conformed to the study by Syamala and Wadhwa (2012). According to Syamala and Wadhwa (2012) that the degree of correlation was varied between the US and BRICs markets, and the correlation among the markets returns were not strong. The cointegration between US and India was weak but China and Russia stock markets were interdependent on US stock market. Meanwhile, Brazil stock market was not influenced by US as well as other BRIC nation. Based on their findings, the global investors have an opportunity to gain from the portfolio diversification in BRIC countries which have a lower cointegration with US stock market.

Fahami (2011) incorporated the subprime crisis to examine its impact towards BRIC and developed countries stock market linkages. The period of analysis was divided into pre-crisis (Jan 10, 2005 – Jul 22, 2007), during crisis (Jul 29, 2007 – Jan 10, 2010) and post-crisis (Jan 11, 2010 – Jul 21, 2011). Weekly closing stock indexes data obtained from seven prominent stock indexes of BOVESPA, RTS, S&P CNX 500, SSE Composite, S&P 500, FTSE 100 and Nikkei 500. The findings revealed an increased of causality relationship among the BRICs and developed countries during the subprime crisis. She further found that China was the most influential stock market before the crisis and US influenced most of the major equity markets during the crisis except for China and India.

Meanwhile, Gupta (2011) also studied the dynamic relationship of the BRICs stock markets in the financial turmoil context. The daily closing indexes from January 2008 to November 2011 were adopted based on Brazil (IBOVESPA), Russian (RTSI), India (S&P CNX Nifty) and China (SCI). The empirical results revealed that there was unidirectional causality from Russia, India, and China to Brazil, and India to Russia. However, China economy had bidirectional causality with Russia and India economy. He concluded that the Chinese economy was largely interdependent on Indian and Russian economy. Therefore, it can be hypothesized that there is stock market integration among BRIC countries, where global investors may not achieve portfolio diversification benefit by adding BRIC countries in the same basket.

**METHODOLOGY**

**Data**

In analyzing the structure of linkages and causal relationship among the stock markets in BRICs, four prominent national stock markets are selected, namely Brazil BOVESPA Index, Russia Moscow Times Index, India Bombay SE 100 Index, and China Shanghai Composite Index. All the series are transformed into natural logarithm form. The full sample period covers the period of 20113 to 2012. Data on daily closing stock indexes are gathered from the Datastream covering the periods of pre-crisis (2003 to 2008) and post crisis (2008 – 2012). These relatively shorter periods were selected because we intend to analyze the relationship between the BRICs stock markets in the midst of subprime crisis.

**Methodology**

Overall, the estimation model follows prior research where the integration is examined under long-run and short-run relationship. Firstly, this study has employed the Augmented Dickey-Fuller (ADF) unit root test to examine the order of integration for the stock markets data in the BRICs in order to check whether the examined data are stationary or not. To test the integration, this research employs Johansen and Juselius (1990) test, which developed the multivariate approach for testing long run cointegration among a set of non-stationary variables. This estimation model is important to evaluate the existence of long run stock market linkages among the BRICs. In order to examine the short-run relationships, this research run Granger causality test. It is developed to test the direction and significance of causality among the BRICs stock markets. According to Granger (1969), variable X is said to Granger cause the variable Y if past value of X could better predict Y after the controlling for past value of Y, or equivalently if the coefficients on the lagged value of X are statistically significant. The Granger causality test is determined into three types: unidirectional for both stock markets; bidirectional across the markets; or none of the relationship exists.

**RESULTS AND DISCUSSIONS**

**Unit Root Test Results**

Finance research usually uses time series data, which are assumed to be stationary. Using non-stationary data in regression analysis leads to spurious regression results. We applied the Augmented Dickey Fuller (ADF) of the unit root test to determine whether the stock market data of BRICs are stationary. Table 3 shows the ADF test results.

This test allows this research to gauge the robustness of the integration properties of market integration. The null hypothesis is that $Y_t$ has a unit root (non-stationary), that is $H_0: \alpha_2=0$, versus the alternative hypothesis that $Y_t$ is stationary or $H_1: \alpha_2<0$. Table 3 shows the results of the unit root test. Comparing the calculated ADF statistic to the MacKinnon (1991) critical value, we conclude that we failed to reject the null hypothesis as most of the variables in the level contained unit root or not stationary for both pre and post crisis of market integration properties. It indicates that all variables are integrated and have long run co-movement indication. Therefore, we continued the procedure for the co-integration test.
Cointegration Test Results

Since the time series has been confirmed as stationary at first difference in unit root test, now we can proceed to examine the long run cointegration relationship among the stock markets in BRICs by using the Johansen-Juselius cointegration test. The significant of Johansen and Juselius cointegration vector was determined through the maximum likelihood based on the trace and maximum eigenvalue statistics. To test for this cointegration or long run relationship, we employed the maximum likelihood approach of Johansen (1988) based on a vector autoregressive model:

\[ X_t = \mu + \sum_{i=1}^{p} \Pi_i X_{t-i} + \sum_{k=1}^{K} X_{t-k} + \epsilon_t \]

Where \( X_t \) is the vector variables integrated on the same order; \( \mu \) is the intercept terms for the vector, \( \Pi \) is the coefficient matrix, and \( \epsilon_t \) is the error terms which are assumed to be white noise. This equation equivalently states in the following cointegration regressions:

\[ Y_t = \beta_0 + \beta_1 X_t + \nu_t \]

Where the null hypothesis is that \( Y_t \) and \( X_t \) are not cointegrated. The specification on the trace test and maximum eigenvalue test are provided under Fraser and Oyefeso (2005) as below:

\[ \hat{\lambda}_{trace}(r) = -T \sum_{i=r+1}^{p} \ln(1 - \hat{\lambda}_i) \]

where \( \hat{\lambda}_{trace} \) is the estimated trace value, \( T \) is the number of observation and \( p \) is the number of variables.

The results of the cointegration test are documented in Table 4. It clearly indicates indicated that there is no cointegration between the BRICs in pre-crisis period as we failed to reject the null hypothesis of non-cointegration. This means there is no co-movement or market integration among BRIC countries in the long run before the subprime mortgage implying global investor might achieve portfolio diversification by adding BRIC countries in their baskets. This is in line with prior research such as Chittedi (2010), and Xu and Hamori (2012), which concluded there is no market integration among BRIC countries.

However, two cointegration equations are found in the post-crisis period for both trace and maximum eigenvalue test, hence, we can reject the null hypothesis \( r \leq 1 \) at five percent level of significance. Thus, we concluded that the cointegration among the BRICs stock markets has increased after the occurrence of US subprime crisis event in 2008. This indicates that some of the BRIC countries are tied with one another in the long run, therefore, the portfolio diversification opportunities have been diminishing through time. This is consistent with previous research such Click and Plummer (2005), Hyghebaert and Wang (2010), and Kenourgios et al. (2011), which concluded that stock market becomes more integrated after a financial crisis event.

Granger Causality Test Results

We proceed with the Granger Causality test in order to investigate the short run dynamic co-movement among BRIC’s market integration. In general, the Granger Causality test is one that can determine whether one variable is useful in forecasting another variable by revealing its causality. Despite there is a long run relationship among the BRICs stock markets, the causality directional test should be conducted to confirm the short run linkages and its causal direction. Table 5 shows the results from Granger causality test based on VECM among the stock markets in BRICs before and after the subprime crisis. The empirical results revealed that the BRICs stock market linkages has established in post-crisis period with the presence of unidirectional and bidirectional causality. The results support the findings by An and Brown (2010), Fahami (2011), Sheu and Liao (2011) who also incorporated the effect of subprime crisis to the level of integration among the BRICs stock markets.

Through the Granger causality test, we noted that the obtained results consist of four unidirectional causalities and two bidirectional relationship among the BRICs. The results revealed that there was a unidirectional causality from India to Brazil, and from China to all other BRIC countries. Meanwhile, the bidirectional causality between Brazil and Russia, and India and Russia were existed. Among the BRICs, China was the most independent stock market but Brazil stock market had the least influence. See Figure 2 for more detail results of granger causality.

Discussion

Since 2000, China and other BRICs countries economic performance had proved themselves to have a better position in the world perspective. From 2000 to 2008, the BRICs had accounted for 30% of the global increased output. During this period of time, BRICs had shared the rate of 16% to 22% in relative to the world economic output, in which China had contributed more than half to this growth1. Low labour cost and large domestic market were the main competitive advantages that brought China’s economy to the world’s second largest in 2010. Under Hu Jintao’s administration, many of the China’s foreign policies had changed and pursued more aggressively. He included more diverse alliances countries like Canada, Australia, Iran and Venezuela in order to increase its power. With an unprecedented shortage of natural resources, China invested heavily in the countries which rich in resources such as Australia,

1 Detailed discussions refer to Global Sherpa.
2 Hu Jintao is the China’s President since 2003 until present.
Africa, Brazil, Pakistan and South Asia to secure access to mineral right. With the multipolar world practice, China had developed and tied with all the major power and regional blocs to foster their relationship. With this, China has a same official position to the US on the terrorism issue. Besides that, China also increased its economic interactions with Asian countries and rebuilt its image as regional leader. In the BRIC nation, China exerted relatively greater influence after the subprime crisis due to its continuously spectacular economic development with the support of rich natural and labour resources in the countries. The massive influx of FDI by foreign investor who seek China as a platform, by utilizing its cheaper resources, to manufacture and sale in China as well as foreign markets (Chow, 2010).

In the meantime, India serves to accelerate the economic growth since after the economic liberalization in 1990s. The trade globalization had expanded Indian two-way trade (merchandise exports plus imports) which deepened the financial integration of India to the world and put India appeared to be an important trading partner (Misra, 2009). Due to the strong population and domestic consumption growth, both the China and India economy are becoming increasingly important in relative to the world.

Over the past few years, the large number of investments had flowed into Brazil due to the country strong economic background as well as economic policies and programs in controlling the taxes and public investment. The Brazil investments were so attractive to the global investors because the oversea investor can enjoy the same rights as domestic investors have with availability of the high return, buoyant and growing market. In 2000s, Brazil became the largest and most powerful countries in South America. With the large population, vast territory, and abundant resources in agricultural and mining, Brazil economy had a chance to stand as strong position in global economy. As at 2011, Brazil still is the world leader in exporting of iron ore and foodstuffs. Since April 2009, China started overtook US and became the Brazil’s largest trading partner\(^3\). In 2011, the records had shown that, Brazil-China goods trade was amounted to US$77 billion which was 28% more than Brazil-US of US$60 billion\(^4\). Therefore, this is an evidence of increased of stock markets integration between China and Brazil and unidirectional causality from China to Brazil after the subprime crisis.

After the dissolution of Soviet Union, a development of new economic model in Russia had achieved rapidly. With its large territory and rich natural resources, Russia had become one of most developing and attractive economy in the world. Russia was one of the largest natural gas producer as well as exporter in the world. In 2011, 35% of the world’s gas was belongs to Russia\(^5\). In 2010, China overtook Germany to become Russia’s largest trading partner. With the bilateral trade, China imported crude oil and gas from Russia as an input to produce equipments back to Russia. This bilateral trade agreement had fostered the integration between China and Russia. Therefore, this evidence has supported our finding on an increasing integration between Russia and China after the subprime crisis. Besides that, the unique geographical position, technological advancement, attractive tax system, extensive government support with stable social and political system had made Russia become an attractive country after the subprime crisis.

CONCLUSION

This research examines the stock market integration among the BRICs emerging market namely, Brazil, Russia, India and China. The long run cointegration and short run causality among the BRICs stock markets had been investigated in order to provide the empirical evidences of portfolio diversification opportunities to the investor in BRICs particularly after the US subprime mortgage crisis. Various econometric testing procedures were adopted to analyze the monthly data covering the periods from 2003 to 2012.

The findings revealed an increasing of long run integration of the BRICs stock markets over the post-crisis period through the adoption of Johansen-Juselius cointegration test. The Granger causality results also indicated an increasing of linkages among the BRICs stock markets after the subprime crisis with significant changes in causality directions. From that, India is Granger causes the Brazil whereas China leads all the other BRIC countries but none of them influence China. Meanwhile, there is bidirectional causality existed between Brazil and Russia, and India and Russia. In summary, China has found to be the most influential stock market while Russia and Brazil stock market performance were largely interdependent on other BRICs markets.

In connection to our findings, the important implication that the degree of stock market integration and causalities direction among the countries tend to change over time especially after the financial crisis (see, for example, Lim 2007, Yang et al. 2003, Bekaert and Harvey 1995). This important implication is crucial for monetary and financial authorities in response to the financial crises in order to implement appropriate policies for financial stability. This implication is also useful for the regional investors in managing the risk of investment whereby to gain the greatest benefits through the portfolio diversification strategies.

---

\(^3\) Details description refers The Telegraph.

\(^4\) Further information refers Rosaliene Bacchus.

\(^5\) Major discussion refers Ministry of Economic Development of Russian Federation.
As a result of stock markets in the BRIC region are integrated in the long run, the potential in obtaining the abnormal profit through portfolio diversification in BRICs is limited because the abnormal profit will be arbitraged away in the long run. Therefore, the international investor may invest in BRICs stock markets up to some extent in order to diversify the risk and earn the abnormal returns in the long run. Nevertheless, the investor still can benefit from the short run linkage that exists between some of the BRICs stock markets.

The authors acknowledge the financial support from Universiti Malaysia Sarawak.

REFERENCES


FIGURES AND TABLES

![Fig. 1 The movement of BRIC's stock market indices](image-url)
Table 1 BRIC Countries - Economy and Income, 2011

<table>
<thead>
<tr>
<th>BRIC</th>
<th>GDP (USD billion)</th>
<th>% of world</th>
<th>GDP, PPP (USD billion)</th>
<th>% of world</th>
<th>Market capitalisation (USD billion)</th>
<th>% of world</th>
<th>Population (million)</th>
<th>% of world</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>2,088</td>
<td>3.3%</td>
<td>2,185</td>
<td>2.9%</td>
<td>1,546</td>
<td>2.8%</td>
<td>195</td>
<td>2.9%</td>
</tr>
<tr>
<td>Russia</td>
<td>1,480</td>
<td>2.3%</td>
<td>2,812</td>
<td>3.7%</td>
<td>1,005</td>
<td>1.8%</td>
<td>142</td>
<td>2.1%</td>
</tr>
<tr>
<td>India</td>
<td>1,727</td>
<td>2.7%</td>
<td>4,195</td>
<td>5.5%</td>
<td>1,616</td>
<td>2.9%</td>
<td>1,171</td>
<td>17.1%</td>
</tr>
<tr>
<td>China</td>
<td>5,927</td>
<td>9.4%</td>
<td>10,170</td>
<td>13.3%</td>
<td>4,763</td>
<td>8.5%</td>
<td>1,338</td>
<td>19.6%</td>
</tr>
<tr>
<td>Total</td>
<td>11,222</td>
<td>17.8%</td>
<td>19,362</td>
<td>25.3%</td>
<td>8,930</td>
<td>15.9%</td>
<td>2,846</td>
<td>41.6%</td>
</tr>
</tbody>
</table>

Source: WDI database, World Bank.

Table 2 Summary of the Review on BRIC Countries

<table>
<thead>
<tr>
<th>Authors (Year)</th>
<th>Countries &amp; Data</th>
<th>Methodology</th>
<th>Important Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chittedi (2009)</td>
<td>BRIC Daily data (Jan 1998 - Aug 2009)</td>
<td>Unit root test (ADF, PP, KPSS)</td>
<td>There is a unidirectional causality from India to Russia and Brazil. Cointegration relationship found between BRICs and developed countries.</td>
</tr>
<tr>
<td>Fahami (2011)</td>
<td>BRIC Weekly data (Jan 2005 – Jul 2011)</td>
<td>Unit root test (ADF, PP)</td>
<td>There is increase causality among the countries during the subprime crisis. China and US is the most influential country before and during the subprime crisis respectively.</td>
</tr>
<tr>
<td>Gupta (2011)</td>
<td>BRIC Daily data (Jan 2008 – Nov 2011)</td>
<td>Unit root test (ADF)</td>
<td>There is unidirectional causality from India, Russia and China to Brazil, and India to Russia. China economy has bidirectional causality with Russia and India economy</td>
</tr>
</tbody>
</table>

Table 3 Augmented Dickey-Fuller Test Results

<table>
<thead>
<tr>
<th>Countries</th>
<th>Test equation</th>
<th>Pre-Crisis</th>
<th>Post-Crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ADF (AIC)</td>
<td>Level</td>
<td>1st difference</td>
</tr>
<tr>
<td>Brazil</td>
<td>Trend &amp; Intercept</td>
<td>-3.080 (0)</td>
<td>-7.459 (0)a</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>-0.790 (0)</td>
<td>-7.494 (0)a</td>
</tr>
<tr>
<td>Russia</td>
<td>Trend &amp; Intercept</td>
<td>-1.853 (0)</td>
<td>-7.910 (0)a</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>-0.881 (0)</td>
<td>-7.984 (0)a</td>
</tr>
<tr>
<td>India</td>
<td>Trend &amp; Intercept</td>
<td>-4.392 (2)a</td>
<td>-4.304 (5)a</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>-1.824 (0)</td>
<td>-4.320 (5)a</td>
</tr>
<tr>
<td>China</td>
<td>Trend &amp; Intercept</td>
<td>-3.949 (10)b</td>
<td>-2.011 (3)</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>-3.096 (10)b</td>
<td>-2.198 (3)</td>
</tr>
</tbody>
</table>

Notes: (c), (b), (a) denote as significant level at 10%, 5% and 1% levels respectively. The lag length selections show in parentheses are determined through the Akaike Information Criterion (AIC). The ADF test examines the null hypothesis of unit root (non-stationary) against the alternative hypothesis of no unit root (stationary).
Table 4 Johansen-Juselius Cointegration Test Results

<table>
<thead>
<tr>
<th>Lag intervals: 1 2</th>
<th>Pre-crisis</th>
<th>Post-crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$</td>
<td>$H_1$</td>
</tr>
<tr>
<td></td>
<td>Statistic</td>
<td>Probability</td>
</tr>
<tr>
<td>$r = 0$</td>
<td>0</td>
<td>$r = 1$</td>
</tr>
<tr>
<td>$r \leq 1$</td>
<td>0</td>
<td>$r = 2$</td>
</tr>
<tr>
<td>$r \leq 2$</td>
<td>0</td>
<td>$r = 3$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$r = 1$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$r = 2$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$r = 3$</td>
</tr>
</tbody>
</table>

Notes: (**), (****) denote as significant level at 5% and 1% levels respectively.

Table 5 Granger Causality Test Results Based on VECM

<table>
<thead>
<tr>
<th>Lag Interval: 1 2</th>
<th>Pre-Crisis</th>
<th>Post-Crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Countries</td>
<td>Brazil</td>
<td>Russia</td>
</tr>
<tr>
<td></td>
<td>$\chi^2$ - statistic (p-value)</td>
<td>$\chi^2$ - statistic (p-value)</td>
</tr>
<tr>
<td>Brazil</td>
<td>-</td>
<td>2.613</td>
</tr>
<tr>
<td>Russia</td>
<td>0.714</td>
<td>-</td>
</tr>
<tr>
<td>India</td>
<td>2.059</td>
<td>0.523</td>
</tr>
<tr>
<td>China</td>
<td>1.338</td>
<td>1.130</td>
</tr>
</tbody>
</table>

Notes: (**), (****) denote as significant level at 5% and 1% levels respectively. Lag length is chosen based on AIC

Fig. 2 Direction of Causality

Notes: B denotes as Brazil, R denotes as Russia, I denotes as India and C denotes as China. The symbol of $\rightarrow$ represents the unidirectional Causality, while $\leftrightarrow$ represents bidirectional causality.
ABSTRACT

This study intends to empirically explore the customer’s perception of the importance of a range of on-line retailer’s web services and after sales service. This study also explores the customers’ perceptions of the retailers’ performance in delivering these services. An online questionnaire survey has been conducted to gather the data from respondents. The data was analysed using Importance-Performance Analysis (IPA). The findings suggest areas of e-service quality where retailers could improve, based on the customers’ perceptions of the retailers’ performance against the importance of some e-service quality features and/ or services on offer. Consequently, this study highlights that retailers should take active steps to understand their customers' requirements, before developing an online customer services strategy. From a practical perspective, retailers could also apply the questionnaire developed for this study to canvas the opinions of customers, to help identify areas in which their performance needs to be improved.

INTRODUCTION

Parasuraman et al., (2005), Cristobal et al., (2007) Loiacono et al., (2007), Yang et al., (2003) are examples of a growing body of work that has focused upon customers’ perceptions of the web-sites service quality that they experience, when either visiting, performing or transacting with a website. The customers’ perceptions of e-service quality are largely based on studies of the traditional service quality measurement. For example, SERVQUAL (Parasuraman et al., 1988) has been widely used in previous offline service quality studies, has been adopted and adapted in e-service quality studies by other researchers (e.g. van Riel et al., 2003; Kaynama and Black, 2000). To correspond with the changing environment, the founding authors of SERVQUAL further developed and enhanced the service quality dimensions to tailor the needs of online offerings by introducing i.e. E-S-QUAL and Rec-S-Q (Parasuraman et al., 2005). This work is recognised as the most comprehensive work on e-service quality (Bauer et al., 2006).

In recent years, significant numbers of new e-service quality measurement models have been developed and tested. For example: PeSQ (Cristobal et al., 2007); e-TRansQual (Bauer et al., 2006); PIROQUAL (Francis and White, 2002); e-TailQ (Wolfinbarger and Gilly, 2003); E-S-QUAL and Rec-S-Q (Parasuraman et al., 2005). In contrast, some studies merely register lists of e-service quality dimensions (e.g. Yang et al., 2003; Yang and Fang, 2004; Long and McMellon, 2004, Cox and Dale, 2002), without exploring the affects and implications of such dimensions.

Nevertheless, when studies have been directly aimed at the virtual experience they have tended to focus on the technical evaluations of web sites interface features such as: the web site design (e.g. Dabholkar, 1996; Liu and Arnett, 2000; Kaynama and Black, 2000; Cox and Dale, 2002; Siu and Cheung, 2001); and, the accuracy and relevance of information (e.g. Liu and Arnett, 2000; Yoo and Donthu, 2001; Loiacono et al., 2007; Cao et al., 2005), rather than the softer service elements. Some studies suggest that consumers have become more concerned with being able to reliably order and receive ordered goods (Barnes and Vigden, 2001). Thus, the technical evaluation focus has been shifted to the assessment of on-line transaction processing (Parasuraman et al., 20005; Wolfinbarger and Gilly, 2003), which potentially renders softer issues as unimportant e.g., ease of use, responsiveness to complaints, availability of information.

The previous body of literature can be criticised in a number of key respects. For example, there has been a tendency to focus solely on consumers’ perceptions of the service providers’ ability to satisfy a variety of service dimensions (e.g. Wolfinbarger and Gilly, 2003; Parasuraman et al., 2005), but without explicitly addressing how important each dimension is to their overall satisfaction to the same set of features / services.

This is an important gap to be filled, because customers are not likely to be impressed by service providers who deliver an ultra fast service, if speed is of no importance to them. Consequently, it is necessary to better
understand consumers’ perceptions of both the perceived importance and the performance of a range of service dimensions, as this will enable service providers to better align their service delivery with their customer’ expectations (O’Neil et al., 2001). Therefore, the purpose of this study is twofold: i) to gain greater understanding of customers’ experiences and perceptions of service quality, when using on-line shopping web-sites; ii) to investigate the customers’ perception of the importance and performance of a variety of elements of e-service quality.

**RESEARCH STRATEGY**

The internet was used to distribute the questionnaire to the respondents because it is perceived to be a relatively cheap and efficient means of conducting a customer survey (Granillo and Wheaton, 2004; Ekman et al., 2006). To ensure that the participants had access to internet facilities, which is commonly identified as a potential drawback of this strategy (Solomon, 2001; Granillo and Wheaton, 2004), the study targeted students, who were known to have access to the Internet. The students should reasonably represent the underlying population of online customers based on the fact that both the university students and the UK population show an increasing trend in online shopping (Jamie, 2010; Wearden, 2011, Rigby, 2012).All questions were adapted from prior e-service quality studies (e.g. Parasuraman et al., 2005; Wolfinbarger and Gilly, 2003; Tih and Ennis, 2006; Bauer et al., 2006; Cao et al., 2005; Yang et al., 2003; Ribbink et al., 2004). The questionnaire has also gone through a series of pre-testing and pilot testing stages.

Respondents were asked to rate their perceptions using 7 point Likert scale, based on their experience of Internet shopping at a specified online store. These questions were explicitly compared and contrasted service performance with importance. The questionnaire was distributed to a sample of 800 randomly selected students, studying at an established university in the UK. Respective students received an invitation email, which has a link to the online questionnaire. The students should reasonably represent the underlying population of online customers based on the fact that both the university students and the UK population show an increasing trend in online shopping (Jamie, 2010; Wearden, 2011, Rigby, 2012).All questions were adapted from prior e-service quality studies (e.g. Parasuraman et al., 2005; Wolfinbarger and Gilly, 2003; Tih and Ennis, 2006; Bauer et al., 2006; Cao et al., 2005; Yang et al., 2003; Ribbink et al., 2004). The questionnaire has also gone through a series of pre-testing and pilot testing stages.

Respondents were asked to rate their perceptions using 7 point Likert scale, based on their experience of Internet shopping at a specified online store. These questions were explicitly compared and contrasted service performance with importance. The questionnaire was distributed to a sample of 800 randomly selected students, studying at an established university in the UK. Respective students received an invitation email, which has a link to the online questionnaire. The students should reasonably represent the underlying population of online customers based on the fact that both the university students and the UK population show an increasing trend in online shopping (Jamie, 2010; Wearden, 2011, Rigby, 2012).All questions were adapted from prior e-service quality studies (e.g. Parasuraman et al., 2005; Wolfinbarger and Gilly, 2003; Tih and Ennis, 2006; Bauer et al., 2006; Cao et al., 2005; Yang et al., 2003; Ribbink et al., 2004). The questionnaire has also gone through a series of pre-testing and pilot testing stages.

Respondents were asked to rate their perceptions using 7 point Likert scale, based on their experience of Internet shopping at a specified online store. These questions were explicitly compared and contrasted service performance with importance. The questionnaire was distributed to a sample of 800 randomly selected students, studying at an established university in the UK. Respective students received an invitation email, which has a link to the online questionnaire. The students should reasonably represent the underlying population of online customers based on the fact that both the university students and the UK population show an increasing trend in online shopping (Jamie, 2010; Wearden, 2011, Rigby, 2012).All questions were adapted from prior e-service quality studies (e.g. Parasuraman et al., 2005; Wolfinbarger and Gilly, 2003; Tih and Ennis, 2006; Bauer et al., 2006; Cao et al., 2005; Yang et al., 2003; Ribbink et al., 2004). The questionnaire has also gone through a series of pre-testing and pilot testing stages.

The steps involved in the IPA analysis for the web service are as follows:

**Steps 1 and 2** – determining the mean values for both Importance and Performance and conducting Paired sample t-Test for each items. The mean values were found to be significantly different at the level of 1 percent ($r=0.255; p < 0.001$). The e-service quality features recording an overall mean performance score of $m= 5.74$, against overall importance score of $m= 6.00$ (Table 2).

**Step 3** – conducting factor analysis. Four items were dropped from this analysis due to their low cross loadings (Hair et al., 2010). Eight factors emerged from the factor analysis: *FULFILLMENT, ORDER RETURN, IMPROVEMENT, EASE OF USE, INFORMATION*, *WEBSITE QUALITY, EASY OF USE, INFORMATION*, *EASE OF USE, INFORMATION*. 

**Importance-Performance Analysis (IPA)**

The IPA analysis is employed to gain a deeper understanding of the customers’ perceptions of the website service quality. The ‘importance’ dimension is used to canvas a person’s general assessment of the significance of a particular attribute, whilst the ‘performance’ dimension is used to assess how well that person’s requirements, with respect to that same attribute, are satisfied (Chu and Choi, 2000). For example, in the context of e-service quality, respondents could be asked to rate both the performance of a specified web-site in satisfying their need for ‘after sales service’, as well as the ‘importance’ of ‘after sales service’ to their experience of on-line shopping at the same site.

The IPA technique emerged from the earlier work of Martilla and James (1977). Although this technique has been widely used in the service quality domain (e.g. Ennew et al., 1993), as yet, it has rarely been applied in the context of electronic commerce (O’Neill et al., 2001). The findings of the IPA technique will be analysed in a four quadrant of Importance-Performance Matrix (O’Neill et al., 2001; Martilla and James, 1977; Chu and Choi, 2000), as shown in Figure 1. The step by step procedures in conducting the IPA analysis, adapted from O’Neill and Palmer (2004) and O’Neill et al., (2001).

**FINDINGS AND DISCUSSION**

The respondents’ demographic background is presented in Table 1. The statistics reveals that most of the respondents’ have a mixture of various education backgrounds. Most importantly, more than half of the respondent’s shops more than 12 times a year and almost 40 percent have experienced more than 6 years in online shopping. These findings indicate that the respondents are experienced enough with the online shopping encounter.

**Web Services**

The steps involved in the IPA analysis for the web service are as follows:

**Steps 1 and 2** – determining the mean values for both Importance and Performance and conducting Paired sample t-Test for each items. The mean values were found to be significantly different at the level of 1 percent ($r=0.255; p < 0.001$). The e-service quality features recording an overall mean performance score of $m= 5.74$, against overall importance score of $m= 6.00$ (Table 2).

**Step 3** – conducting factor analysis. Four items were dropped from this analysis due to their low cross loadings (Hair et al., 2010). Eight factors emerged from the factor analysis: *FULFILLMENT, ORDER RETURN, IMPROVEMENT, EASE OF USE, INFORMATION*, *WEBSITE QUALITY, EASY OF USE, INFORMATION*. 

**Importance-Performance Analysis (IPA)**

The IPA analysis is employed to gain a deeper understanding of the customers’ perceptions of the
WEB-SITE RESPONSIVENESS, SECURITY and PRIVACY.

Step 4 – Paired sample t-Test for each factor on Importance and Performance. The results are presented in Table 3. In analyzing the mean value of performance versus importance, three negative values have been identified for the FULFILLMENT, ORDER RETURN, and IMPROVEMENT components. These significant negative difference could imply that the retailers’ e-service quality performance is at a level much lower than the customers’ perceptions of their importance. The two positive values for the EASE OF USE and INFORMATION components indicate good news for the retailers, as their performance perceived by customers exceed the importance. However, as neither of these results is significantly different, thus the impact is negligible. The neutral values for the WEB-SITE RESPONSIVENESS, SECURITY and PRIVACY components signify that there are no differences in customer perception of the components’ importance and performance - the retailers address these issues very effectively.

Step 5 – Charting each factors onto the IPA matrix presented using cross-hair matrix. The matrix facilitates in identifying more clearly, between the stronger and weaker factors (O’Neill et al., 2001). Figure 2 illustrates the eight components of e-service quality that fall equally into two opposite quadrants.

Four of the constructs in Quadrant B (i.e. FULFILLMENT, EASE OF USE, SECURITY, and WEB RESPONSIVENESS) were rated above average, both in terms of their importance and performance. These findings suggest that all the four constructs are performing at an appropriate ‘keep up the good work’ level. However, ORDER RETURN, IMPROVEMENT, PRIVACY and INFORMATION are in Quadrant C. These factors were rated as below average in both performance and importance. On the face of it, these four areas are of low priority.

After Sales Service

The steps involved in the IPA analysis for the after sales service are as follows:

Steps 1 and 2 – The Paired sample t-Test for (Step 1 and 2) the after sales service (as in Table 4) indicate that the difference between after sales service performance and importance are statistically significant (t = -6.812; p < 0.001). The overall performance mean for the after sales service is m=5.10, whilst the importance mean is m=5.82.

Step 3 – Factor analysis. All items were ultimately used, as none needed to be dropped from the analysis. Three extracted; CUSTOMER SERVICE, COMPENSATION and CONTACT.

In assessing the reliability measure for each of the.

Step 4 – The paired sample t-Test (Step 4) for the performance versus importance mean analysis is presented in Table 5, and demonstrates negative values for all three components of after sales service. The negative values indicate that, from the respondents’ perspective, the retailers’ after sales service performance is at a level much lower than its perceived importance. The results of the paired sample t-Tests indicate that all three components exhibit significant differences between their performance and importance, in the area of after sales service, where all differences are significant at 1% level.

Step 5 – Figure 3 illustrates clearly demonstrates that the three components fall in two opposite quadrants, as discussed below:

On one hand, CONTACT and CUSTOMER SERVICE components signify that the retailers should ‘keep up the good work’ (Quadrant B) as they are broadly aligning their performance with their customers’ expectations. On the other hand, the positioning on the COMPENSATION (Quadrants C) suggests that the respondents’ regard it as being of relatively low importance and performance, and it can therefore be rated as a ‘low priority’ (Martilla and James, 1977). However, the large and significant gap, between performance and importance, suggests that retailers should really keep an eye on this facet of their after the eight components of e-service quality that fall equally into two opposite quadrants.

DISCUSSION AND CONCLUSION

This study is among the first to employ the IPA, specifically focusing upon the customers views of the service quality that they receive for their online shopping. The study highlights a number of important insights into the effectiveness of retailers’ current e-service quality practices.

Firstly, all the four factors positioned in Quadrant B of web services in the IPA matrix, is in the ‘ideal’ desired position in the matrix. It indicates that all the four factors reflects an optimal performance, would be perceived that these elements to be performing well above average in relation to the e-services elements deemed to be important by customers. The four factors that fall into this ‘comfort’ area should be maintain by the online retailers, in order to ensure that they have the capabilities and abilities to deliver the best values of online shopping experience to their customers.

Secondly, Quadrant C of web services reflects the fact that certain aspects of the e-service quality are not performing to their full potential. However, given that respondents have recorded a relatively low importance and performance ranking for all the four factors, underperformance in this area may not be so serious and can be safely ignored. However, from the operational
perspectives, it can be argued that all elements in this quadrant should be devise around their ability to meet and exceed customer expectations. In particular, the underperformance of ORDER RETURN and IMPROVEMENT should be taken seriously as prior research has demonstrated that the same order return issue can impact upon customer service quality perception and satisfaction (Jiang and Rosenbloom, 2005; Cristobal et al., 2007). Furthermore, the PRIVACY factor as it is one of the important risky elements and major obstacles, that might lead to customer resistance for the customers to perform their online shopping (Miyazaki and Fernandez, 2001; Lee and Lin, 2005; Wolfinbarger and Gilly, 2003).

Thirdly, for the after sales service, the IPA metrics shows that CUSTOMER SERVICE and CONTACT are in the ‘ideal position’. However, the fact that there are significant negative gaps between performance and importance, for both these constructs (see Table 3), suggests that there is absolutely no room for complacency. suggests that the respondents’ regard it as being of relatively low importance and performance, and it can therefore be rated as a ‘low priority’ (Martilla and James, 1977). However, the large and significant gap, between performance and importance for COMPENSATION for the after sales service, suggests that retailers should really keep an eye on this facet of their after the eight components of e-service quality that fall equally into two opposite quadrants.

Whilst highlighting some important areas, this study is limited, as the empirical evidences and findings of this study are appropriate only to the online UK retailers study. Any generalisations to a wider population, for example, to other industries or to other countries should be done with caution due to structural, cultural, social, political or economic differences, between populations.

ACKNOWLEDGEMENT

The author acknowledges the financial support from Universiti Malaysia Sarawak.

REFERENCES


Jamie, (2010). Student’s Online Spending Habits. The Students Room. Available at: http://insight.thestudentroom.co.uk/1n51ght/students-online-spending-habits/ Access Date: 16th April, 2012


Liu, C. & Arnett, K.P. (2000). Exploring the Factors Associated With Web Site Success In The


**FIGURES AND TABLES**

Fig. 1 The Importance-Performance Matrix
Table 1: Respondent Profile

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>97</td>
</tr>
<tr>
<td>Female</td>
<td>114</td>
</tr>
<tr>
<td>Total</td>
<td>201</td>
</tr>
<tr>
<td>Education Background</td>
<td></td>
</tr>
<tr>
<td>Diploma</td>
<td>69</td>
</tr>
<tr>
<td>1st degree</td>
<td>49</td>
</tr>
<tr>
<td>Masters degree</td>
<td>8</td>
</tr>
<tr>
<td>Others</td>
<td>75</td>
</tr>
<tr>
<td>Total</td>
<td>201</td>
</tr>
<tr>
<td>Online Shopping Experience</td>
<td></td>
</tr>
<tr>
<td>Less than 1 year</td>
<td>10</td>
</tr>
<tr>
<td>Between 1-3 years</td>
<td>52</td>
</tr>
<tr>
<td>Between 4-6 years</td>
<td>63</td>
</tr>
<tr>
<td>More than 6 years</td>
<td>74</td>
</tr>
<tr>
<td>Total</td>
<td>201</td>
</tr>
<tr>
<td>Respondents Online Shopping Frequency</td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>6</td>
</tr>
<tr>
<td>Between 2-5 times</td>
<td>26</td>
</tr>
<tr>
<td>Between 5-8 times</td>
<td>31</td>
</tr>
<tr>
<td>Between 9-12 times</td>
<td>27</td>
</tr>
<tr>
<td>More than 12 times</td>
<td>111</td>
</tr>
<tr>
<td>Total</td>
<td>201</td>
</tr>
</tbody>
</table>

Fig. 2  Web Service Importance-Performance Matrix

Note: Overall Importance Mean = 6.0;
Overall Performance Mean = 5.7

Fig. 3  After Sales Service Importance-Performance Matrix

Note: Overall Importance Mean = 5.7;
Overall Performance Mean = 5.0
### Table 2 Web Service I-P Mean And Paired Sample t--Test

<table>
<thead>
<tr>
<th>Importance</th>
<th>Mean</th>
<th>Performance</th>
<th>Mean</th>
<th>P - I</th>
<th>t-value</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/WQ 1</td>
<td>6.31</td>
<td>6.01</td>
<td>-0.295</td>
<td>-3.620</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 2</td>
<td>5.28</td>
<td>5.42</td>
<td>0.141</td>
<td>1.232</td>
<td>.220</td>
<td></td>
</tr>
<tr>
<td>C/WQ 3</td>
<td>5.89</td>
<td>6.01</td>
<td>0.115</td>
<td>1.437</td>
<td>.153</td>
<td></td>
</tr>
<tr>
<td>C/WQ 4</td>
<td>6.74</td>
<td>6.15</td>
<td>-0.590</td>
<td>-9.754</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 5</td>
<td>6.78</td>
<td>6.28</td>
<td>-0.551</td>
<td>-7.472</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 6</td>
<td>6.72</td>
<td>6.19</td>
<td>-0.524</td>
<td>-9.669</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 7</td>
<td>6.00</td>
<td>5.43</td>
<td>-0.567</td>
<td>-6.333</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 8</td>
<td>6.66</td>
<td>5.45</td>
<td>-1.215</td>
<td>-7.037</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 9</td>
<td>6.42</td>
<td>5.77</td>
<td>-0.654</td>
<td>-9.493</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 10</td>
<td>5.96</td>
<td>6.20</td>
<td>0.237</td>
<td>2.951</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>C/WQ 11</td>
<td>5.91</td>
<td>5.90</td>
<td>0.010</td>
<td>1.152</td>
<td>.251</td>
<td></td>
</tr>
<tr>
<td>C/WQ 12</td>
<td>5.94</td>
<td>6.07</td>
<td>0.135</td>
<td>1.565</td>
<td>.135</td>
<td></td>
</tr>
<tr>
<td>C/WQ 13</td>
<td>6.12</td>
<td>5.51</td>
<td>0.611</td>
<td>5.232</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 14</td>
<td>5.99</td>
<td>6.09</td>
<td>0.090</td>
<td>1.274</td>
<td>.024</td>
<td></td>
</tr>
<tr>
<td>C/WQ 15</td>
<td>6.35</td>
<td>6.01</td>
<td>-0.343</td>
<td>-7.275</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 16</td>
<td>6.15</td>
<td>6.34</td>
<td>0.186</td>
<td>2.482</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td>C/WQ 17</td>
<td>6.41</td>
<td>6.46</td>
<td>0.051</td>
<td>0.791</td>
<td>.430</td>
<td></td>
</tr>
<tr>
<td>C/WQ 18</td>
<td>5.96</td>
<td>5.32</td>
<td>-0.677</td>
<td>-5.940</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 19</td>
<td>5.92</td>
<td>5.67</td>
<td>-0.242</td>
<td>-2.762</td>
<td>.006</td>
<td></td>
</tr>
<tr>
<td>C/WQ 20</td>
<td>6.12</td>
<td>5.96</td>
<td>-0.154</td>
<td>-1.612</td>
<td>.103</td>
<td></td>
</tr>
<tr>
<td>C/WQ 21</td>
<td>6.45</td>
<td>6.22</td>
<td>-0.224</td>
<td>-3.119</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>C/WQ 22</td>
<td>6.09</td>
<td>5.90</td>
<td>-0.190</td>
<td>-2.077</td>
<td>.099</td>
<td></td>
</tr>
<tr>
<td>C/WQ 23</td>
<td>5.81</td>
<td>5.50</td>
<td>-0.324</td>
<td>-2.159</td>
<td>.032</td>
<td></td>
</tr>
<tr>
<td>C/WQ 24</td>
<td>5.81</td>
<td>5.68</td>
<td>-0.127</td>
<td>-1.758</td>
<td>.080</td>
<td></td>
</tr>
<tr>
<td>C/WQ 25</td>
<td>5.78</td>
<td>5.34</td>
<td>-0.440</td>
<td>-4.751</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 26</td>
<td>5.52</td>
<td>4.49</td>
<td>-1.125</td>
<td>-0.250</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 27</td>
<td>5.79</td>
<td>5.13</td>
<td>-0.654</td>
<td>-4.741</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C/WQ 28</td>
<td>4.93</td>
<td>4.55</td>
<td>-0.372</td>
<td>-3.161</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>C/WQ 29</td>
<td>5.21</td>
<td>5.00</td>
<td>0.210</td>
<td>1.162</td>
<td>.241</td>
<td></td>
</tr>
</tbody>
</table>

Note: p<0.05

### Table 3 Paired Sample t-Test – The Level of Adoption

<table>
<thead>
<tr>
<th>Importance</th>
<th>Performance</th>
<th>Mean</th>
<th>Mean</th>
<th>P - I</th>
<th>t-value</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSVC1</td>
<td>6.36</td>
<td>5.41</td>
<td>-0.95</td>
<td>-5.589</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC2</td>
<td>5.71</td>
<td>5.53</td>
<td>-0.17</td>
<td>-2.015</td>
<td>.312</td>
<td></td>
</tr>
<tr>
<td>DSVC3</td>
<td>6.00</td>
<td>5.49</td>
<td>-0.51</td>
<td>-4.154</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC4</td>
<td>6.01</td>
<td>5.47</td>
<td>-0.54</td>
<td>-4.745</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC5</td>
<td>6.40</td>
<td>5.37</td>
<td>-1.03</td>
<td>-6.190</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC6</td>
<td>6.10</td>
<td>5.34</td>
<td>-0.76</td>
<td>-5.035</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC7</td>
<td>6.04</td>
<td>5.33</td>
<td>-0.71</td>
<td>-4.775</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC8</td>
<td>5.90</td>
<td>5.34</td>
<td>-0.56</td>
<td>-4.246</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC9</td>
<td>5.59</td>
<td>5.42</td>
<td>-0.17</td>
<td>-1.925</td>
<td>.350</td>
<td></td>
</tr>
<tr>
<td>DSVC10</td>
<td>5.79</td>
<td>5.53</td>
<td>-0.26</td>
<td>-1.561</td>
<td>.123</td>
<td></td>
</tr>
<tr>
<td>DSVC11</td>
<td>5.07</td>
<td>4.12</td>
<td>-0.95</td>
<td>-5.455</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC12</td>
<td>5.44</td>
<td>4.18</td>
<td>-1.26</td>
<td>-6.800</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC13</td>
<td>5.75</td>
<td>4.77</td>
<td>-0.98</td>
<td>-5.707</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>DSVC14</td>
<td>5.53</td>
<td>4.15</td>
<td>-1.38</td>
<td>-7.648</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

Note: p<0.05

### Table 4 After Sales Service I-P Mean And Paired Sample t-Test

#### Table 5 Paired Sample t-Test - After Sales Service

<table>
<thead>
<tr>
<th>Importance</th>
<th>Performance</th>
<th>Mean</th>
<th>Mean</th>
<th>P - I</th>
<th>t-value</th>
<th>Sig. (2-tailed)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUSTOMER SERVICE</td>
<td>6.080</td>
<td>5.411</td>
<td>-0.627</td>
<td>-5.539</td>
<td>0.000</td>
<td>Significant</td>
<td></td>
</tr>
<tr>
<td>CONTACT</td>
<td>5.774</td>
<td>5.151</td>
<td>-0.623</td>
<td>-4.579</td>
<td>0.000</td>
<td>Significant</td>
<td></td>
</tr>
<tr>
<td>COMPENSATION</td>
<td>5.404</td>
<td>4.440</td>
<td>-0.935</td>
<td>-5.756</td>
<td>0.000</td>
<td>Significant</td>
<td></td>
</tr>
</tbody>
</table>

Note: p<0.05
ABSTRACT

This study aims to investigate the relationship between carbon dioxide emissions (CO\textsubscript{2}), international tourism receipt (TR) and international tourist arrival (TA) in Malaysia and South Korea by using annual time series data from 1980 to 2009. The tests employed for the paper are Augmented Dickey-Fuller (ADF) and Philips-Perron (PP) unit root test, Johansen and Juselius Cointegration Test and Vector Error Correction Model (VECM). Malaysia exhibits a long-run relationship among the variables and a short-run causality running from tourism receipt to CO\textsubscript{2} emission. Meanwhile, South Korea exhibits a short-run causality from tourist arrival and tourism receipt to CO\textsubscript{2} emission. Therefore, the governments in both countries should persistently monitor their current environmental policy so that the policy will consistently be paralleled with the current and future needs of the tourism sector.

INTRODUCTION

The tourism sector is one of the most lucrative sectors in the economy in these past few decades. According to Othman et al. (2012), tourism is defined as a temporary movement of people to destination outside of the workplace and the usual place of residence, the activities that they do while staying at the destination and the facilities provided to meet their needs. In the global world, many developed and developing countries are shifting towards economic development based on the tourism industry. Consequently, most developing countries recognize tourism as the engine of growth in their effort to economically develop their countries (Awang and Abdul Aziz, 2011). Therefore, it is essential for the government to identify the potential and strength of the sector in order to bring more profit to tourism-related industries and as a result generate more income and job opportunities for the country as a whole.

According to World Development Indicator (2012), international tourism receipt can be defined as the expenditures by international inbound visitors, such as the payments to national carriers for international transport and it also consist of any other prepayment made for goods or services received in the destination country. Furthermore, this receipt also may include from passenger transport items and same-day visitors. Environment is important to the tourism sector. The environmental qualities become one of the most important aspect in enhancing the growth of tourism sector. Environmental quality can be defined as a characteristics that relate to the natural environment as well as the built environment including air and water or pollution, noise and the potential effects which such characteristics may have on physical and mental health come from human activities. Over the last decade, the relationship between economic development and environmental quality has been extensively explored.

TOURISM EFFECTS ON THE CARBON DIOXIDE EMISSIONS

There are long-standing debate regarding the relationship between economic growth and environmental quality. Generally, the relationship between these factors is presented using Environmental Kuznets Curve theory. The Environmental Kuznets Curve (EKC) hypothesis explained that the holds points of economic development and environmental degradation can be expressed by an inverted-U curve (Ni et al., 2010). To investigate the Environmental Kuznets Curve hypothesis typically the researchers utilized environmental quality-as
measured by CO₂ and economic growth—measured by GDP. Hung and Shaw (2006) found that there is a significant relationship where the economic growth highly contributes to the CO₂ emission compared to other environmental quality. Furthermore, the significant relationship between these factors has been studied by Sanglimsuwan (2011) and Boopen and Vinesh (2011) in Mauritius. Meanwhile, a study by Tiwari (2011) in India from the period 1971 until 2005 found that there are bi-causality between CO₂ emissions and GDP and between CO₂ and energy consumption.

As there are several studies that focused on the relationship between GDP and environmental quality, this study aims to analyze the tourism sector as the contributor of GDP with the variables from tourism sector. Thus, this research begins with a series of studies by focusing in the general topics on the relationship between environmental quality and tourism, which can lead to a better understanding of the importance of the two variables in current world. Then, this studies look into the specific environmental quality variables which is CO₂ emission and tourism specifically, international tourist arrivals and international tourism receipt. The number of tourist arrivals and tourism receipt variable is considered as the variables that affect the CO₂ emission.

Tourism can create beneficial effects on environment by assisting conservation through serving as a tool to finance the maintenance of natural environment (Huang, 2012). In most popular tourist destinations, the tourists gather together extensively in a short time, to cause a series of environmental problems, such as heavy traffic, atmospheric pollution, noise pollution, trash stench, and the pollution of the bathing beach water, which are the most typical environmental problems (Lee et al., 2009).

As the international tourist arrival is rising, the increasing demand on the tourism sector leads to the increase of related goods and services production such as the land clearance for built up more hotels, recreational parks and tourism accommodation. The activities including afforestation, reforestation, and deforestation as a potential means to reduce greenhouse gases (GHGs) in atmosphere and, thus, to meet the emission on reduction targets allocated to the developed countries. The tourism sector consumes land, ecosystems and threatened species that has had unpleasant impacts through land clearance, construction and overconsumption on coral reefs, coastal wetlands (especially mangroves), rainforests and mountainous areas (McGregor et al., 2009). All this activities needs to be carried out to encourage the international tourists to visits.

Transport industry became a key factor of tourism development that facilitated the movement of tourists from their sources to the destination. The transportation contributes a large portion of carbon dioxide emission and its emission is growing faster (Tol, 2007). According to Jatuporn and Chien (2011), the
linkages of tourism development to CO₂ emissions are mostly defined as transportation and economic activities through domestic energy consumption. The tourism development variables used by researcher are international tourism receipts, international tourism expenditures, tourism deficit, natural resources depletion and net forest depletion to investigate relationship with CO₂ emission. The forecasting on the relationship between CO₂ emission and international tourist arrival by Mayor and Tol (2008) is revealed, the increasing incomes will lead to the average tourist being able to take more trips a year by the end of the century hence carbon dioxide emissions follow the growth pattern until 2060 when emissions per passenger-kilometer start to fall due to improvements in fuel efficiency. This forecasting used world annual data from year 1950 until 2005 by applying the Hamburg Tourism Model. Therefore, according to Hamilton et al. (2005), the relationship between tourism and CO₂ can depend by the tourist arrivals and what they should prefer.

Moreover, the previous researchers found that the transportation has positive relationship between energy consumption and CO₂ of tourism industry, the main effect of emission growth influenced by five factors including energy consumption and CO₂ in tourism, energy intensity, expenditure size and the industry size is developed by Liu et al. (2011) using Decomposition analysis Chengdu data distributed in which over 50,000 tourists during 1999 until 2004. A study on the total energy use and CO₂ emissions relationship with international passenger air travel to (and from) New Zealand by using visitor arrival data collected by Statistics New Zealand, it estimated a total energy use of 27.8 PJ found that positive relationship between the international tourist arrival and the international air travel (CO₂) (Becken, 2002).

The demand on good transportation services is developed with the increasing number of road, water and air transportation availability. Transport via air travel contributes more to CO₂ emission level because the air travel plays a major role to the tourism sector (Becken, 2002). Growing demand on air travel causes the combustion of fossil fuels to release a range of greenhouse gases including methane (CH₄), nitrous oxide (NO₂) and CO₂ emission. Hence, CO₂ emission is the most important one as the atmospheric concentration increases due to human activities (Ubaidillah et al. 2012). The main impact of the increasing CO₂ emission by tourism demand is from air transportation had been identified by Becken (2002); Tol (2007); Mayor and Tol (2007) and Jatuporn and Chien (2011). Tourism also played a most important role in the growth of air travel and being a third biggest industry behind petroleum and automobile (Becken, 2002). Thus, the international tourist arrivals are being the main contributor in increasing CO₂ emission because the demand for air travel is increasing. According to Becken (2002), the environmental impacts of air travel come from energy use and greenhouse gas and travelling by air requires huge amounts of energy and releases CO₂ emission into the atmosphere. In the case of small-island developing states, by promoting air travel to islands induces growing tourists flows that need to be carefully managed with regard to the three dimensions of sustainability: economic, social, and environmental (Becken, 2002). Due to increasing in tourism receipt, tourist arrival and CO₂ emission, this study need to identify whether these variables have causality or not.

Therefore, this study aims to identify the long-run and short-run relationship between CO₂ emission as the environmental quality proxy and tourism demand factors. Malaysia and South Korea have been selected as a case study to present the developing and developed country and to compare the significant of those factors to their CO₂ emission. Those countries are selected because Malaysia is one of the ten countries that achieve highest international tourist arrival in 2010 (United Nation World Tourism Organization (UNWTO), 2012). Meanwhile, South Korea as a developed country had the highest CO₂ emission among the ten countries in the 2008 world data (EIA, 2011).
TOURISM GROWTH AND RELATED ENVIRONMENTAL QUALITY REGULATIONS IN MALAYSIA AND SOUTH KOREA.

Malaysia is one of the fastest growing in terms of tourism industries among developing country and is ranked ninth place in the world tourism rankings in 2010 as shown in Table 1. Tourism sector in Malaysia becomes the second largest sector that contributes to the Gross Domestic Product (Othman et al., 2012). Malaysia's tourism also affects other economic sectors; it stimulates the economic progress of developing economies and its importance in gaining widespread recognition (Mazumder et al., 2009). Malaysia is one of the countries enhances the tourism development through ecotourism and has formulated some plans for this sector. The plans include the East Coast Economic Region (ECER) which covers the states of Kelantan, Terengganu, Pahang and the district of Mersing, Johor. The ECER Master Plan is developed to guide the development of this region over the next 12 years covering the 9th, 10th and 11th Malaysia Plans. It aims to transform into major international and local tourism destination (East Coast Economic Region (ECER), 2007).

Tourism in Malaysia depends heavily on environmental resources, including clean beaches, warm climate, and cleans air. However, the rising number in international tourists demand will cause imbalance in environmental quality particularly in the CO_2 emission. The tourists are the main reason in increasing the CO_2 emission because they get involved in the services activities including transportation, shopping, food, entertainment and others (Bhuiyan et al., 2012).

As a significant segment to South Korea national economy, the tourism final demand and value-added revenue had contributed to 3.5 percent of the Gross Domestic Product of South Korea in 1988 and the exaggerated rapid growth of tourism sector in the next two decades had boost up this country GDP by 6.6 percent in 2008 (Seo et al., 2009). South Korea tourism sector does not only influence the economic growth but it also affects other aspects which include environmental quality. The influence on the environment is one of the most significant impacts, specifically, which can be observed easily and shortly during the peak tourist season. In South Korea, the tourist destinations might burden a majority of tourists on the peak tourist season on account of the small area and high population density, typically, in the summer east seacoast and winter ski field (Lee et al., 2009).

The Basic Law on Sustainable Development is established in 2007 wherein South Korea can actively involve in international efforts for sustainable development such as countermeasures against global warming (Environmental Quality South Korea, 2012). The law stated that in reducing climate change which aimed to build a low-carbon society by shifting the country's economic growth strategy from growth-oriented development to low carbon and green growth, is prepared and made public. The Presidential Committee on Green Growth, in which most of the main cabinet members participate, is established in 2009, and the Basic Law on Low-Carbon and Green Growth is enacted at the end of December 2009 to support the Committee's activities (Environmental Quality South Korea, 2012).

The tourists are responsible for 4.4 per cent of global CO_2 emission and emissions are increasing at an average rate of 3.2 percent per year up to 2035 (Bhuiyan et al., 2012). In Malaysia and South Korea the CO_2 emissions have along with the increase of international tourist arrival. In 1979, Malaysia's CO_2 emission based on Figure 1 is 2.02 metric tons per capita and it is reached 7.57 metric tons per capita in 2008 while international tourist arrival increases from 2.04 million visitors in 1979 to 22.05 million visitors in 2008 as shown in Figure 2.

In the case of South Korea as shown in Figure 3, it shows that in 1979 the CO_2 emission is 3.53 metric tons per capita increase to 10.40 metric tons per capita in 2008 while international tourist arrival also rising from 1.13 million visitors to 6.89 million visitors as shown in Figure 4. Furthermore, the income of international tourist arrivals will lead to the international tourism receipt rising at the same time. In Korea, tourism receipts also rise by 23 per cent to hit a record of 12 billion US dollars.
in 2010 meanwhile Malaysia achieved a total of 24.6 million tourists arrived and spent RM56.5 billion in 2010. This purpose of this paper is to identify the short-run and the long-run relationship between Tourism sector and CO₂ emission for Malaysia and South Korea. The following sections will discuss on the methodology and the empirical techniques and findings and discussions. The final section will discuss on the conclusion.

METHODOLOGY AND EMPIRICAL TECHNIQUES

Data and Methodology

This study uses time series data collected from World Bank Data, Korea Tourism Organization and Putrajaya Department of Tourism Malaysia. This study gathered data of the number of international tourist arrival (TA), international tourism receipt (TR) and CO₂ emission (metric tons per capita) in Malaysia and South Korea from 1980 to 2009. The time-series data period are chosen by taking into consideration the availabilities of tourism and emission data for both countries. CO₂ is Carbon Dioxide Emission (metric tons per capita), TA is International Tourist Arrivals, TR is International Tourism Receipt (Malaysia in MYR million and South Korea in US1000). All data are transformed into natural logarithm. The studies from Tol (2007); Huang (2012); Lee et al., 2009; and Mayor and Tol (2008) used international tourist arrival and environmental quality as a variable in the non-linear model whereas Jatuporn and Chien (2011), Tol (2007); Zaman et al. (2011) and Nademi and Najibi (2011) used CO₂ and tourism demand as a variables.

This study investigates two models which are South Korea as a proxy of a developed country whereas Malaysia as a developing country. Typically researchers used the tourism demand variables including tourism expenditure, tourism receipt, tourism deficit and others as dependent variables. This study performs the Augmented Dickey-Fuller (ADF) and Philips-Perron (PP) unit root test, Johansen and Juselius Cointegration Test and Vector Error Correction Model (VECM). The Cointegration test and Granger Causality test application in determine the relationship between environmental quality and tourism had been used in study by Lee et al. (2009) by using South Korea annual data from the year 1999 to 2008 and Pulina and Biagi (2006) by using Sardinia (Italy) annual data over the period 1955 to 2004 (Lee et al., 2009).

Augmented Dickey-Fuller (1979) and Philips-Perron Unit Root Test

Augmented Dickey-Fuller (ADF) and Philips-Perron (PP) test will be applied to test the stationary. The simple Dickey-Fuller unit root test described above is valid only if the series is an AR(1) process. If the series is correlated at higher order lags, the assumption of white noise disturbances t is violated. Phillips and Perron (1988) propose an alternative (non-parametric) method of controlling for serial correlation when testing for a unit root. The PP method estimates the non-augmented DF test equation, and modifies the t-ratio of the coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic.

Johansen and Juselius Cointegration Test

The transformation of the data can be employed to induce stationary if a time series is found to be non-stationary (Lee et al., 2009). However, cointegration should be tested firstly because the method might bring about losing the information of the data. If non-stationary time series has cointegration, there is no spurious regression because of a long-run stationary relationship or equilibrium relations among the variables.

Vector Error Correction Model

The direction of the Granger causality can only be detected through the vector error correction model (VECM) derived from the long-run cointegrating vectors. Furthermore, to indicating the direction of causality amongst variables, the VECM also allows to distinguish between short-run and long-run Granger causality. The innovative methodology enables one to test theories that generated Granger causality implications using the environment of ECM and VECM that includes the error term.
component. The Granger causality test is very sensitive to the number of lag used in the analysis (Gujarati, 1995). According to this approach, a variable x is caused by a variable y if x can be predicted better from the past values of both x and y than from past values of x alone and vice versa (Granger, 1969). The error correction terms (ECTs) can be consistently obtained from the corresponding lagged residuals of the single equation cointegration regression. The relevant error correction term (ECTs) must be included in the VAR to avoid misspecification and omission of the important constraints. The dynamic interaction between all the macroeconomic variables would provide coherent policy stance for an economy.

FINDINGS AND DISCUSSIONS

Unit Root test

According to table 2 and table 3, it is found that for ADF and PP tests in Malaysia and Korea at level, all the variables fail to reject the unit root null hypothesis. However, at first difference for both ADF and PP tests, all test statistics indicate significance at 5 percent level of significance. The variables LCO$_2$, LTA and LTR are stationary indicating that the variables are in fact integrated at order one, I(1). Therefore, Johansen and Juselius Cointegration test are conducted since the variables are stationary.

Cointegration Analysis

According to Lee et al. (2009), if a time series is found to be non-stationary, the transformation of the data can be employed to induce stationarity. If non-stationary time series has cointegration, there is no spurious regression because of a long-run stationary relationship or equilibrium relations among the variables.

From the Johansen and Juselius Cointegration test results shown in table 4, it is indicated that the null hypothesis of the zero cointegration (r=0) is rejected by both trace and maximum eigenvalue test at 5% significance level. The results indicate that for the Malaysian case, the variables do not drift apart and share at least one common stochastic trend in the long-run. Thus, the vector error correction model (VECM) test is applied.

However, the result in table 5 shows that there is no cointegration vectors exist since the null hypothesis is not rejected at 5% significance level for both maximum eigenvalue and trace statistic. Therefore, this implies that South Korea does not support the long-run equilibrium relationship between tourist arrival, tourism receipt and CO$_2$ emission. Thus, we proceed to the short-run Granger Causality test.

Vector Error Correction Model (VECM) Result

As a result of the presence of a single cointegrating vector in the co-integration test, the detection of cointegration implies that proper VAR systems consists of error correction term (ECT) is required to study the dynamic relationship between the variables. The direction of the Granger causality can only be detected through the vector error correction model (VECM) derived from the long-run cointegrating vectors. Furthermore, to indicate the direction of causality amongst variables, the VECM also allows distinguishing between short-run and long-run Granger Causality. Table 6 illustrates the results of Granger causality test in a VECM framework and Error Correction Term (ECT).

Table 6 shows the causal relationship between LCO$_2$, LTA and LTR based on VECM framework. In the short-run, it is found that only LTR granger causes LCO$_2$. In other word, there is unidirectional Granger causality relationship running from LTR to LCO$_2$ as shown in figure 5. Based on the result, the error correction term (ECT) for $\Delta$LCO$_2$ carries the correct sign that is negative, less than one and statistically significant at 5% level, giving evidence that $\Delta$LCO$_2$ is adjusted towards the long-run equilibrium level. ECT is used to measure the speed of the temporal adjustment to long-run equilibrium. The estimated coefficient of ECT for $\Delta$LCO$_2$ is about 0.512 of the short-run deviation will be adjusted towards the equilibrium annually or it indicates that about 51.2 percent of adjustment occurs in a year. In
short, the speed of adjustment of CO$_2$ is approximately two years to adjust back to the equilibrium.

**Short-run Granger Causality Test Results**

Based on the causality test result in table 7, it is found that LTA and LTR granger cause LCO$_2$ short-run. This shows that the number of tourist as well as tourist receipt does affect the CO$_2$ emission in South Korea but only in the short-run (Figure 6).

**CONCLUSION**

According to the analysis, the results show that there is no existence of long-run relationship among the variables in South Korea. This is because it is difficult to verify the impact of tourism demands to CO$_2$. For the case of South Korea, the demand for tourism may not be the main contributor to the growth of CO$_2$ in this country. As South Korea is the world's tenth largest emitter of CO$_2$ in 2008, according to CDIAC (2013), the largest contributor to CO$_2$ emission in South Korea are fossil-fuel consumption and cement production which provide 139 million metric tons of carbon. Those contributors come from industrial activities particularly in manufacturing sector not tourism sector.

In contrary, Malaysia is found to have long-run relationship among the variables. However, the speed of adjustment of CO$_2$ only takes two years to adjust back to the equilibrium. Based on the statistical analysis, there is a long-run relationship between TR, TA and CO$_2$ emission. In addition, there is a short-run causal relationship running from tourism receipt to CO$_2$ emission. This is due to the fact that Malaysia has become the third largest source of income from foreign exchange in tourism (UNWTO, 2012) in which the major activities engaged in are sightseeing and visiting interesting places plus other activities such as shopping, visiting theme parks and beaches (Ibrahim, Zahari, Sulaiman, Othman and Jusoff, 2009). These activities contribute towards enhancing the international tourism receipt. Meanwhile, the environmental degradation as a result of the tourism related industries or sectors such as transportation; service and communication sector has triggered greater GHG emission specifically the CO$_2$ emission through the supply or process of providing the goods and services to the tourists. According to Liu, Feng and Yang (2011), the CO$_2$ emission in tourism, energy intensity, expenditure size and industry size effect are generally found to be main drivers of emission growth.

South Korea have restricted the environmental policy for instance by implementing the Natural Environment Conservation Act 1991 that contribute to the environmental sustainability which bring forth the importance of protection of the environment specifically in the tourism sector. They used eco-friendly technology especially to the potential firm that contributes to the CO$_2$. In the case of Malaysia, greater reinforcement pertain to the environmental aspect of tourism sector is crucial to circumvent the sector from becoming one of the major contributor of CO$_2$ emission in the country. Specific rules or regulation should be implemented in order to provide a better standard for the sector in the long-run. In addition, the tourism sector should move forward towards green tourism as the sector is much more emphasized on the conservation of the flora and fauna, preserving the heritage and wildlife thus reducing as much pollution as possible. As a country that is resource and environmentally well-endowed, this aspect should be put forward as a competitive advantage and be utilized wisely in a long-term.

**ACKNOWLEDGEMENTS**

Financial support from Universiti Malaysia Sarawak (UNIMAS) is gratefully acknowledged. All remaining flaws are the responsibilities of the authors.

**References**


Carbon Dioxide Information Analysis Center, 2013, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.


Environmental Quality South Korea, 2012, *Environmental Policy and Environmental Fiscal Reform of South Korea: Comparative Study to those of Japan*.


Table 1 The World Tourism Rankings in 2010

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>International tourist arrivals (in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>France</td>
<td>76.80</td>
</tr>
<tr>
<td>2</td>
<td>United States</td>
<td>59.75</td>
</tr>
<tr>
<td>3</td>
<td>China</td>
<td>55.67</td>
</tr>
<tr>
<td>4</td>
<td>Spain</td>
<td>52.68</td>
</tr>
<tr>
<td>5</td>
<td>Italy</td>
<td>43.63</td>
</tr>
<tr>
<td>6</td>
<td>United Kingdom</td>
<td>28.13</td>
</tr>
<tr>
<td>7</td>
<td>Turkey</td>
<td>27.00</td>
</tr>
<tr>
<td>8</td>
<td>Germany</td>
<td>26.88</td>
</tr>
<tr>
<td>9</td>
<td>Malaysia</td>
<td>24.58</td>
</tr>
<tr>
<td>10</td>
<td>Mexico</td>
<td>22.40</td>
</tr>
</tbody>
</table>


Fig. 1 Malaysia’s CO₂ emission in metric tons per capita in 1979 to 2008
(Source: World Bank Data, 2012)

Fig. 2 Malaysia’s International Tourist Arrival in millions in 1979 to 2008
(Source: World Bank Data, 2012)

Figure 3: Korea’s CO₂ emission in metric tons per capita in 1979 to 2008
(Source: World Bank Data, 2012)

Figure 4: Korea’s International Tourist Arrival in millions data in 1979 to 2008
(Source: World Bank Data, 2012)

Figure 5: Short-run Causality Direction in Malaysia
Fig. 6 Short-run Causality Direction in South Korea

Table 2 Augmented Dickey-Fuller (ADF) and Philips-Perron (PP) unit root test for Malaysia

<table>
<thead>
<tr>
<th>Level</th>
<th>ADF</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td></td>
<td>Linear trend</td>
<td>Linear trend</td>
</tr>
<tr>
<td>LCO</td>
<td>-1.055[0]</td>
<td>-1.708[0]</td>
</tr>
<tr>
<td>LTA</td>
<td>-0.535[0]</td>
<td>-2.813[0]</td>
</tr>
</tbody>
</table>

First Differences

| LCO            | -5.918          | -5.853[1]** |
| LTA            | -5.337[0]*      | -0.252       |
| LTR            | -5.182          | -0.675       |

Note: LCO = log carbon dioxide emission LTA = log tourist arrival, LTR = log tourism receipt. Asterisks (**) denote significant at 5 percent level. Figure in parentheses are the lag lengths.

Table 3 Augmented Dickey-Fuller (ADF) and Philips-Perron (PP) unit root test for South Korea

<table>
<thead>
<tr>
<th>Level</th>
<th>ADF</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td></td>
<td>Linear trend</td>
<td>Linear trend</td>
</tr>
<tr>
<td>LCO</td>
<td>-1.597[0]</td>
<td>-0.733[0]</td>
</tr>
<tr>
<td>LTA</td>
<td>-1.691[0]</td>
<td>-1.495[0]</td>
</tr>
<tr>
<td>LTR</td>
<td>-2.440[0]</td>
<td>-1.637[0]</td>
</tr>
</tbody>
</table>

First Differences

| LCO            | -5.003[0]*     | 5.394[0]*    | 5.003[1]*      |
| LTA            | 4.710[0]*      | 4.837[0]*    | 4.710[0]*      |
| LTR            | 3.744[0]*      | 4.009[0]*    | 3.733[1]*      |

Note: LCO = log carbon dioxide emission LTA = log tourist arrival, LTR = log tourism receipt. Asterisks (**) denote significant at 5 percent level. Figure in parentheses are the lag lengths.
lagged values of the independent variables, the significant of the error correction term(s) $\Delta$ is the first differences operator. Asterisks (*) indicate statistically significant at 5 percent level

Table 7 South Korea Short-

<table>
<thead>
<tr>
<th>Null Hypothesis:</th>
<th>F-Statistic</th>
<th>p-value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTA does not Granger Cause LCO$_2$</td>
<td>4.883*</td>
<td>0.017</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>LCO$_2$ does not Granger Cause LTA</td>
<td>0.077</td>
<td>0.926</td>
<td>Do not reject $H_0$</td>
</tr>
<tr>
<td>LTR does not Granger Cause LCO$_2$</td>
<td>2.838*</td>
<td>0.079</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>LCO$_2$ does not Granger Cause LTR</td>
<td>1.039</td>
<td>0.370</td>
<td>Do not reject $H_0$</td>
</tr>
<tr>
<td>LTR does not Granger Cause LTA</td>
<td>2.526</td>
<td>0.102</td>
<td>Do not reject $H_0$</td>
</tr>
<tr>
<td>LTA does not Granger Cause LTR</td>
<td>0.256</td>
<td>0.776</td>
<td>Do not reject $H_0$</td>
</tr>
</tbody>
</table>
AN ECM ANALYSIS OF THAI TOURISM DEMAND IN MALAYSIA

Fung-Thai Thien, Chin-Hong Puah and Mohd Khairul Hisyam Hassan
Faculty of Economics and Business, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.

ABSTRACT

Tourism industry is one of the important service industries that play a crucial role in the development of Malaysian economy. Thailand has been ranked among the top five tourist generating countries for Malaysia since last decade. This paper intends to determine factors affecting Thai tourism demand in Malaysia from the macroeconomic perspective by utilizing quarterly data from 2000Q1 to 2011Q4. The empirical results indicate that there is a long-run equilibrium relationship between Thai tourism demand and the specified explanatory variables, which are Thai real income, real travel cost and exchange rate. Specifically, higher Thai real income will reduce Thai tourist to Malaysia, while higher real travel cost and stronger currency attract more Thai visitor to Malaysia.

INTRODUCTION

Malaysian government has given attention to its tourism industry following the poor performances of Malaysia’s primary commodities and manufacturing industry, which were resulted by the oil price drop and economic recession in the mid-1980s. The Ministry of Culture, Arts and Tourism was established by Malaysian government in 1987, and was upgraded to Ministry of Tourism in 2004.

In order to promote Malaysia to tourist around the globe, a series of promotion campaign was implemented since 1990s. The famous Visit Malaysia Year was carried out in 1990, 1994, 2000 and 2007. At the same time, Malaysia also organized annual events such as the World Amateur Inter-team Golf Championship, the Rainforest World Music Festival and the Borneo International Jazz Festival which attract mass arrival of international tourist.

According to World Travel and Tourism Council (WTTC) (2012), Malaysia’s tourism industry contributed 14.8 percent (RM 125.4 billion) to its GDP in 2011. WTTC (2012) forecasted this contribution to increase further to 15.1 percent of GDP (RM 130.8 billion) in 2012. Furthermore, Malaysian tourism industry contributed 12.9 percent (1.56 million jobs) of total employment in 2011 and is projected to contribute to 13.2 percent (1.6 million jobs) of total employment in 2012.

In addition, the role of tourism industry is highlighted in the Economic Transformation Programme (ETP), which is introduced to lead Malaysian economy towards achieving the status of high income nation by 2020. A successful ETP will divert Malaysia into service-based economy in which service sector will contribute to 65 percent of GDP. Under ETP, 12 National Key Economic Areas (NKEAs) are identified as the key engines of growth in which tourism industry is included in the list.

Malaysia has a strong position for its tourism industry in which Malaysia was ranked in top 10 for both the most visited country and tourist arrival in the world, as well as ranked top 15 in the global receipts (PEMANDU, 2010). Therefore, it is important to study the factors affecting tourism demand in Malaysia as this particular industry will play an important role in the future development of Malaysian economy.

Thailand was ranked the second in tourist generating countries for Malaysia from 2000 until 2006. Nevertheless, Thailand’s ranking has dropped to be the third tourist generating country for Malaysia started in 2007. Despite this, the average annual growth for Thai visitor arriving into Malaysia is 11.8 percent for the period 2000 to 2011. Table 1 shows the number of Thai tourist arrival into Malaysia, their market share and annual growth from 2000 to 2011.

As depicted in Table 1, Thai tourist arrival has shown an upward trend from 2000 to 2005, with the highest tourist arrival in 2005 (accounted for 11.6 percent of total tourist arrival in Malaysia). However, since 2006, Thai tourist arrival has shown a downward trend until 2011 with the highest negative annual growth of 14.1 percent in 2007. The decreasing Thai tourist arrival in recent years has raised the needs to study the factors affecting Thai tourist arrival into Malaysia.

Objective

The objective of this paper is to empirically examine the Thai tourism demand in Malaysia from a macroeconomic perspective.

LITERATURE REVIEW

In the study of tourism demand, tourist arrival has been utilized by a number of studies such as Dritsakis...
change in income. Given that demand increases more than proportionate to a demand for tourism in Malaysia:

**MODEL, DATA AND METHODOLOGY**

Other researchers such as Ouerfelli (2008), Botti et al. (2007) and Brida et al. (2008) also obtained results supporting elastic coefficient for income in the cases of France, Mexico and Tunisia, respectively. The elastic income coefficient indicates that tourism is a luxury good given that demand increases more than proportionate to a change in income.

On the other hand, Garin-Munoz (2006) pointed out that transportation cost or travel cost is also an important explanatory variable to be included in tourism demand study. This variable measures the travel expenditure of tourist to travel from their origin country to the tourism destination and this expenditure makes up a considerable part of total expenditure of their trip. Dritsakis (2004) and Nelson et al. (2011) are among the researchers that have included travel cost in examining tourism demand. Both authors utilized the airfare to measure the travel cost in their study.

Nevertheless, Mohd Salleh et al. (2007) revealed that difficulty exists in determining the exact flight that tourist abroad and hence the amount of airfare paid. As such, Mohd Salleh et al. (2007) suggested to proxy travel cost using the crude oil price. This proxy is also used in the study of Garin-Munoz and Montero-Martin (2007), Garin-Munoz (2006), Mohd Salleh et al. (2008) and Habibi and Abdul Rahim (2009), for example.

In addition, exchange rate is evidenced to be a significant variable to explain tourism demand by the studies of Dritsakis (2004), Ouerfelli (2008), Mohd Salleh et al. (2007) and Mohd Salleh et al. (2008), to name a few. The appreciation of origin country’s currency against the destination’s currency will encourage tourist to travel as their spending power have increase and are able to purchase more goods and services in tourism destination.

According to Purchasing Power Parity, the depreciation or appreciation of the exchange rate also reflects the difference in the rate of prices (or the difference in inflation rates) in origin country and destination country. Therefore, exchange rate variable carries the information of both price levels in tourist origin country and tourism destination. The price levels carry information of cost of living in the destination as compared to the country of origin.

The following function is used in examining Thai demand for tourism in Malaysia:

\[ TA = F(Y, TC, EX) \]  

(1)

where \( TA \) represents Thai tourist arrival to Malaysia, \( Y \) proxies the real income of Thai, \( TC \) is the real travel cost and \( EX \) is the exchange rate between Thailand and Malaysia.

A positive relationship is expected to exist between Thai real income and tourism demand in Malaysia. This indicates that an increase in Thai real income will encourage more Thai tourist to visit Malaysia and vice versa, a decline in Thai real income subsequently will reduce the Thai tourist visiting Malaysia. Thai real income is computed by deflating Thai GDP with its consumer price index (CPI).

Meanwhile, real travel cost is expected to have a negative relationship with Thai tourist inflow to Malaysia. Higher real travel cost would increase the total cost of the trip and reduce the visiting Thai tourist to Malaysia. The real travel cost proxy used in this study is real crude oil price. Real crude oil price is calculated by deflating crude oil price with the CPI of the US.

The existence of a positive relationship between exchange rate and tourism demand is expected, which in particular, the appreciation of Thai Baht will encourage more Thai visitor travelling to Malaysia and vice versa, Thai Baht depreciation will decrease the number of Thai tourist visiting Malaysia. The exchange rate variable used in this study is the ratio of currency between Malaysia and Thailand.

The tourism demand model to be examined in this study is expressed as follow:

\[ LTA = \beta_0 + \beta_1 LY + \beta_2 LTC + \beta_3 LEX + \varepsilon_t \]  

(2)

Data for Thai tourist arrival is collected from Tourism Malaysia\(^1\), while the data used to compute Thai real income and exchange rate are collected from International Financial Statistics (IFS) published under International Monetary Fund (IMF) and for travel cost, crude oil price data is obtained from Global Economic Monitor published by World Bank. All variables are transformed into natural logarithm form before conducting any estimation. The period covered in this study is from 2000Q1 to 2011Q4, utilizing quarterly observation.

The empirical testing procedures start off by examining the time series properties of the variables utilizing the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationary test (Kwiatkowski et al., 1992). The next step after identifying the time series properties of the variables is to test for long-run relationship among the variables by implementing the Johansen and Juselius multivariate cointegration test (Johansen and Juselius, 1990).

Once the long-run cointegration relationship is established, Error-Correction Model (ECM) is carried out to determine the interaction among the variables. ECM is

\(^1\) Data of tourist arrival from Thailand is downloaded and compiled from http://corporate.tourism.gov.my/research.asp?page=facts_figures.
also able to investigate the speed of adjustment of the dependant variable back to equilibrium when there are changes in explanatory variables. Short-run Granger causality test result is obtained from ECM before running a series of diagnostic tests to ensure that the model is robust.

EMPIRICAL RESULTS

Stationary Test Results

This study utilized KPSS stationary test to determine the time series properties of variables. Optimal lag length is chosen based on Barlett Kernel estimation with Newey-West bandwidth. Table 2 shows the results of KPSS stationary test. The null hypothesis of stationary is rejected at the level, but is not able to be rejected after first differencing. Hence, all variables are integrated of order one or I(1) as illustrated in Table 2.

Multivariate Cointegration Test Results

Once the variables are identified to be integrated of order one, Johansen and Juselius multivariate cointegration test is implemented to determine the number of cointegrating vector(s) in the model. The null hypothesis of zero cointegration (r = 0) is rejected by both maximum eigenvalue and trace statistics at 5 percent level of significance as shown in Table 3. However, we are not able to reject the null hypothesis of at most one cointegrating vector. Thus, the existence of a single cointegrating vector in the model lead us to presume that there is a long-run relationship among the variables in the model.

Normalized Cointegrating Vector Estimates

The evidence of long-run equilibrium relationship found in the cointegration test allows us to normalize the obtained cointegrating vector with respect to the dependant variable (LTA) in order to obtain the long-run estimates of the independent variables. The following equation shows the estimated results acquired from vector error correction estimates:

\[ \text{LTA} = 43.707 - 6.135L + 1.564LTC + 8.087LEX \]

The t-statistic is reported in the parentheses and indicates that all the estimated coefficients are statistically significant at 1 percent level. The estimated sign of real income is negative, which contradicts the theory. Higher Thai real income in this case will lead to decrease in Thai tourist arrival to Malaysia. The possible explanation behind this phenomenon is that Thai tourist with higher income seek for alternate tourism destination in other countries rather than their neighboring country. The increase in income will encourage Thai tourist to travel to other tourism destinations which previously may be perceived as expensive and unaffordable for them.

Real travel cost is identified to have a positive relationship with Thai tourist arrival into Malaysia. In other words, increase in real travel cost will increase visiting Thai tourist to Malaysia. This positive relationship can be explained by considering the possibility that higher real travel cost will subsequently increases airfare which discourages Thai visitor to travel via airline services. Therefore, as an alternative, they use their own transport to travel to Malaysia, their neighboring country.

Furthermore, Thai visitor is also more probable to travel to Malaysia using their own transport as they can take the advantage to buy petrol in Malaysia which is cheaper due to the subsidies provided by Malaysian government. A trip to Malaysia using their own transportation would be pleasant for Thai tourist as they can enjoy their travelling trip in Malaysia and return home purchasing cheaper petrol at the same time.

Last of all, exchange rate has the plausible sign and in line with past literature. The positive relationship between exchange rate and tourism demand implies that appreciation of Thai Baht will increase Thai tourist travelling to Malaysia. Thai visitor will perceive that they are wealthier when their currency appreciates as more unit of Malaysian Ringgit can be exchange by a unit of Thai Baht. This situation will lead Thai to perceive that travelling and spending in Malaysia is cheaper than before.

Estimation of Error-Correction Model

In Table 4, the summarized estimation of ECM for Thai tourism demand in Malaysia is presented. The model passes all the diagnostic tests. The estimated residuals are normally distributed with constant variances, free from serial correlation and misspecification. In addition, the CUSUM and CUSUM of squares stability tests indicate that the model is stable within the period of study.

The estimated error-correction term (ECT) is statistically significant and has the correct negative sign for its coefficient, which implies that the variables in the system are cointegrated in long-run (Kueh et al, 2009). As suggested by the ECT value, approximately 25.1 percent of short-run deviations of tourism demand will be adjusted in quarter basis towards the long-run equilibrium state.\(^2\)

Table 5 demonstrates the short-run causality test results obtained from the ECM. The null hypothesis of no causal relationship is rejected for real income and real travel cost at 5 and 10 percent significant levels, respectively. This implies that tourism demand is influenced by both variables in the short-run. This suggests that Thai visitors are influenced by the wealth effect, where they perceive they are wealthier when their real income increases and travel more often to Malaysia in

\(^2\) A dummy variable (Dum2001Q4) is included in the ECM to capture the impact of terrorist attack in September 2011.
the short-run. Moreover, Thai travelers are also responsive to the changes in travel cost, where they have the alternative to travel to Malaysia using their own transport when travel cost increases in the short-run.

**SUMMARY AND CONCLUSION**

This study examines the factors affecting Thai tourist arrival into Malaysia from macroeconomic point of view. The independent variables utilized in this study are real income, real travel cost and exchange rate. First of all, time series properties of the variables are identified to be $I(1)$ by the KPSS stationary test. Next, Johansen and Juselius multivariate cointegration test confirmed the existence of long-run equilibrium relationship among the variables under study.

The normalized equation shows that two of the variables have reversal sign and one with plausible sign. The reversal sign of real income is possible considering Thai visitor may travel to other countries in the long-run when their real income increases. They feel that they afford to travel to other countries rather than only their neighboring country.

On the other hand, the positive sign of real travel cost is viewed with the possibility that increase in real crude oil price (proxy for real travel cost) will increase airfare and subsequently, Thai visitor may choose to travel to Malaysia using their own transport. Another probable explanation is the airfares travelling to Malaysia are relatively cheaper if compared with airfares to other countries, although both airfares may have increase.

Exchange rate is the only variable with plausible sign in line with past literature. In the long-run, appreciation of Thai Baht will increase Thai tourist to Malaysia as one unit of Thai Baht can be exchange for more units of Malaysian Ringgit. In other words, goods and services in Malaysia are relatively cheaper and this will attracts more Thai tourist into Malaysia.

ECM is performed to determine the interaction between the variables of study. Results obtained from the ECM once again confirm the existence of long-run equilibrium relationship among the variables in the model. Short-run deviations will be adjusted back towards the long-run equilibrium. In addition, Granger causality test is carried out and the results indicate that real income and real travel cost can be used to predict the future movement of Thai tourist.

Policy implications that can be suggested in this study are to improve the attractions of Malaysia’s tourism, and increase the promotion of Malaysia’s tourism in Thailand. In the long-run, less Thai tourist will travel to Malaysia because as their real income increase, they start to travel to other countries. Therefore, tourism authorities in Malaysia should increase the attractiveness of their tourism destination to ensure that the Thai visitor revisit Malaysia because of these attractions. These attractions may exist in the form of uniqueness of its cultures, products and the environment in the destinations.

Also, Thai visitor who are satisfied of their trip to Malaysia or, are attracted by the uniqueness of Malaysia’s tourism will share their experiences with their friends and families. Therefore, increasing promotion of Malaysia’s tourism is parallel with the spread of information about Malaysia’s tourism to attract potential Thai visitor. The spread of information together with the timely promotion of Malaysia’s tourism is able to grasp Thai tourist as these information will ease the concern regarding safety of the destination.

Additionally, by knowing that Thai tourist are likely to travel to Malaysia when the travel cost increase, Malaysia should prepare to welcome Thai tourist when an increase in the real crude oil price (the proxy for real travel cost) is observed. Once again, a timely promotion campaigns in Thailand during the period of increasing travel cost is expected to attract more Thai tourist arrival.

In short, tourism sector in Malaysia plays an important role in developing its economy, not only in the past but in the future as well. As WTTC has pointed out that tourism industry has a wider impact on an economy via its direct and also indirect or induced contributions, it is essential to study the factors affecting tourist arrival to Malaysia. This study focused on determining the factors affecting tourist arrival to Malaysia for the case of Thailand, who ranked the third in the tourist generating countries for Malaysia. With the knowledge of factors affecting tourist arrival to Malaysia, Malaysia tourism authorities can design appropriate policies to attract more international visitor.

**ACKNOWLEDGEMENT**

The authors acknowledge the financial support of Universiti Malaysia Sarawak via Dana Principal Investigator: 03(DPI03)/793/2011(03).

**REFERENCES**


Tourist Arrival

<table>
<thead>
<tr>
<th>Year</th>
<th>Tourist Arrival (Person)</th>
<th>Market Share (Percentage)</th>
<th>Annual Growth (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>940,215</td>
<td>9.2</td>
<td>88.6</td>
</tr>
<tr>
<td>2001</td>
<td>1,018,797</td>
<td>8.0</td>
<td>8.4</td>
</tr>
<tr>
<td>2002</td>
<td>1,166,937</td>
<td>8.8</td>
<td>14.5</td>
</tr>
<tr>
<td>2003</td>
<td>1,152,296</td>
<td>10.9</td>
<td>-1.3</td>
</tr>
<tr>
<td>2004</td>
<td>1,518,452</td>
<td>9.7</td>
<td>31.8</td>
</tr>
<tr>
<td>2005</td>
<td>1,900,839</td>
<td>11.6</td>
<td>25.2</td>
</tr>
<tr>
<td>2006</td>
<td>1,891,921</td>
<td>10.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>2007</td>
<td>1,625,698</td>
<td>7.8</td>
<td>-14.1</td>
</tr>
<tr>
<td>2008</td>
<td>1,493,789</td>
<td>6.8</td>
<td>-8.1</td>
</tr>
<tr>
<td>2009</td>
<td>1,449,262</td>
<td>6.1</td>
<td>-3.0</td>
</tr>
<tr>
<td>2010</td>
<td>1,458,678</td>
<td>5.9</td>
<td>0.6</td>
</tr>
<tr>
<td>2011</td>
<td>1,442,048</td>
<td>5.8</td>
<td>-1.1</td>
</tr>
</tbody>
</table>


Table 2: KPSS Stationary Test Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level</th>
<th>First Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTA</td>
<td>0.234 (4)***</td>
<td>0.200 (3)</td>
</tr>
<tr>
<td>LY</td>
<td>0.234 (3)***</td>
<td>0.168 (2)</td>
</tr>
<tr>
<td>LEX</td>
<td>0.154 (1)**</td>
<td>0.058 (3)</td>
</tr>
<tr>
<td>LEX</td>
<td>0.169 (0)**</td>
<td>0.169 (1)</td>
</tr>
</tbody>
</table>

Notes: LTA = natural log of tourist arrival, LY = natural log of real income, LTC = natural log of travel cost, and LEX = natural log of exchange rate. Asterisks (**) and (***)) indicate significant at 5 percent and 1 percent levels, respectively.

Table 3: Johansen and Juselius Cointegration Test Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>H0</th>
<th>H1</th>
<th>λmax</th>
<th>λtrace</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTA, LY, LTC, LEX</td>
<td>r = 0</td>
<td>r = 1</td>
<td>29.414**</td>
<td>57.604**</td>
</tr>
<tr>
<td></td>
<td>r ≤ 1</td>
<td>r = 2</td>
<td>19.406</td>
<td>28.189</td>
</tr>
<tr>
<td></td>
<td>r ≤ 2</td>
<td>r = 3</td>
<td>5.326</td>
<td>8.783</td>
</tr>
<tr>
<td></td>
<td>r ≤ 3</td>
<td>r = 4</td>
<td>3.456</td>
<td>3.456</td>
</tr>
</tbody>
</table>

Notes: r is the number of cointegrating vectors. Asterisk (**)) indicates significant at 5 percent level. The optimal lag length for VAR model is three, which is selected based on Schwert (1987) criteria.
### Table 4: Estimation of Error-Correction Model

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.009</td>
<td>0.283</td>
<td>0.778</td>
</tr>
<tr>
<td>ΔLY_{t-1}</td>
<td>2.076</td>
<td>2.272</td>
<td>0.028</td>
</tr>
<tr>
<td>ΔLRCOP_{t-2}</td>
<td>-0.306</td>
<td>-1.738</td>
<td>0.089</td>
</tr>
<tr>
<td>Dum2001Q4</td>
<td>-0.643</td>
<td>-3.028</td>
<td>0.004</td>
</tr>
<tr>
<td>ECT(-1)</td>
<td>-0.251</td>
<td>-2.908</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Diagnostic Tests:

- $R^2$: 0.350
- $R^2$: 0.286
- JB: 3.357 [0.186]
- AR[2]: 2.995 [0.061]

<table>
<thead>
<tr>
<th></th>
<th>ARCH[1]</th>
<th>RESET</th>
<th>CUSUM</th>
<th>CUSUM$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.212</td>
<td>0.487</td>
<td>Stable</td>
<td>Stable</td>
</tr>
</tbody>
</table>

Notes: JB is the Jarque-Bera statistic for testing normality, AR[2] and ARCH[1] are the Langrange Multiplier tests for 2nd order serial correlation and ARCH effects, respectively. RESET refers to the Ramsey RESET specification test while CUSUM and CUSUM$^2$ refer to CUSUM and CUSUM of squares stability tests.

### Table 5: Short-run Granger Causality Test Results

<table>
<thead>
<tr>
<th>H$_0$</th>
<th>p-value</th>
<th>Causal Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LY does not Granger cause LTA</td>
<td>0.028</td>
<td>Reject H$_0$</td>
</tr>
<tr>
<td>LTC does not Granger cause LTA</td>
<td>0.089</td>
<td>Reject H$_0$</td>
</tr>
</tbody>
</table>
ABSTRACT

This study analyses the impact of the International Financial Reporting Standard (IFRS) compliance practice by non-financial corporations listed on equity capital market of Bursa Malaysia. Research methodology embodies content analysis on annual reports to determine the level of IFRS compliance practice by the companies listed on Bursa Malaysia. Multiple regression analysis examines the association between the levels of IFRS compliance practice of the firms and their value relevance. The findings of this study highlights that the International Financial Reporting Standard compliance practice has an impact on the firms’ value relevance of earnings and solvency in their equity market in Malaysia. This finding contributes international financial reporting standard as a reliable quality for being the value relevance in equity capital market.

Keywords: International Financial Reporting Standard, content analysis, and value relevance

1. INTRODUCTION

International Accounting Standards Committee (IASC) issued the International Accounting Standards (IAS). The IASC ran from 1973 through to 2001; it was succeeded by the IASB. Since 2001, IASB has been issuing standards referred as IFRS. To avoid unnecessary confusion, it refers to international accounting standards broadly as IFRS hereafter. The issued standards of Malaysian Accounting Standard Board (MASB), initially known as the Financial Reporting Standards (FRSs) which are named as Malaysian Financial Reporting Standards (MFRSs). Both FRS and MFRS are equivalent to IFRSs. The listed companies on Bursa Malaysia are mandated to use IFRS/FRS in 2005 and MFRS standard by 2012.

The current accounting standards issued by MASB for the private companies are highly converged with IFRS. However, the implementation of International Financial Reporting Standard compliance practices by the firms need to convey the quality of accounting information to be beneficial in the equity capital market. This study focuses on non-financial firms, which are implementing the Malaysian Financial Reporting Standards (compliance with IFRS). The constituent of the companies listed on main market of Bursa Malaysia includes different industrial sectors. The main concern on choosing of non-financial corporations bases on the consistency of observed data and validity of measuring the level of IFRS compliance among them.

The main objective of the study is to analyses the firms’ accounting quality with the level of IFRS compliance practice in the main market of Bursa Malaysia by using content analysis based on the mandatory disclosure index of selected IFRSs commonly reported by the firms from different sectors. It also aims to examine the value relevance of earnings and solvency of the firms to predict the level of higher IFRS compliance.

2. LITERATURE REVIEW

Financial reporting quality with IFRS compliance has been a very common issue around the world since early years of 2000. Especially, Union exchange-listed companies are required to use The International Financial reporting Standards (IFRSs) starting from 2005 fiscal year. Based on the official estimation for 2005 (Hughes, 2007), in the initial year of EU adoption IFRS, some 8000 financial statements were prepared in accordance with IFRS. Many studies examined the economic consequences of mandatory IFRS reporting around the world; and it highlighted that the effects on reducing cost of equity capital and stock market liquidity around the time of IFRS started implementation. Thus the capital market benefit occurred only in the countries which have IFRS compliance practice (Aisbitt, 2006; Daske et al., 2008). There is a series of studies conducted on IFRS compliance in the year 2000s; and a great deal of studies in the earlier period find that there were many companies which were non-compliance with the International Accounting Standard/ International Financial Reporting Standard (IAS/IFRS) in different countries at that time (Cairns, 1999; Street & Gray, 1999; Tower et al., 1999). Later studies highlighted on cross-listed firms to identify the level of IFRS compliance practice with mandatory
disclosure requirements of some selected IFRSs. It indicates that the greater level of compliance with IFRS around 2005 (Al-Shiab, 2008; Carlin et al., 2009). Their research instruments to measure the level of disclosure compliance with IFRS applied self-constructive indices based on the selected IFRS used in annual reports of firms. The level of compliance in emerging markets, mean value of compliance score for the entire period (2005-2007) was 63 percent. Hassan et al. (2009) study on the level of compliance based on financial instruments standards of IFRS compliance in Malaysia; their findings indicate that the improvement of the compliances were from 26 percent in 1999 to 58.9 percent in 2003. Malaysian firms are mandated to use IFRS or MFRS after their studies and the IFRS compliance scores should be more or less similar to the other emerging markets.

Dake and Günther (2006) examined the financial reporting quality with the financial statement of Austrian, German and Swiss firms adopted internationally recognised standards either IFRS or U.S. Generally Accepted Accounting Principle (U. S. GAAP). They provided evidence that disclosure quality increased significantly by using IFRS in these European countries. Ball et al. (2003) and Francis et al. (2003) found that firms with desirable earnings which increase financial information quality. Leuz et al. (2003) presented a capital market model with rational expectations, in which better information improves the coordination between firms and investors regarding the investment decisions. The financial information quality reduces misalignment risk between the firm and investors.

Several researchers conducted empirical analysis on voluntary and mandatory disclosures in financial reports relating to the equity market. The prior studies frequently used proxies for earning and accrual quality to measure the accounting quality (Dechow and Dicheve, 2002; McNichols, 2002). The accrual accounting quality measure was based on how well accruals shown in cash flows at that time. The measure of accrual accounting quality has been used in a variety of ways in their regression models, often as a measure of information quality or financial reporting quality (McNichols, 2002). Prior studies measured the accounting quality; their findings highlights that a performance of a firm can be matched with unexplained accrual measures (Kothari et al., 2001). Those measures of accounting quality were based on the financial value results and it might be indirectly measured the accounting quality. Barth et al. (2008) measured the accounting quality using earnings management, timely loss recognition, and value relevance matrices. Following the prior researches, they predicted that firms with higher quality earnings exhibit less earnings management, more timely loss recognition, and higher value relevance of earnings and equity book value. There are many studies in measuring the convergence of national financial reporting standards to IFRS, observing on the accounting standard issuers, and not on the firm level of the standard implementation. For example, Xiaohui QU (2008) measured the accounting quality with China Accounting Standards (CAS) compared with IFRS. The researchers used fuzzy cluster model to test 33 CAS to examine if CAS complies with IFRS.

Despite the increased attention to enhance higher quality of financial reporting, the term “financial reporting quality” is vague and difficult to define. Although IASB has developed a set of high quality global accounting standards which requires transparent and comparable information in financial reports, it does not present a specific definition of high quality accounting. In addition, Penman (2007)’s notion of accounting quality is centered on the usefulness of information for investors. Under this perspective, accounting quality upholds not only the interest of investors but also increases the public interest. Although there has not been a clear definition of accounting quality at the firm level, several studies measure accounting quality with value relevance (Daske & Günther, 2006; Hung & Subramanian, 2007; Penman, 2007; Barth et al., 2008; Armstrong et al., 2009). Higher quality of International Financial Reporting is beneficial in that it provides superior firm-specific information that can serve towards reducing information asymmetry and limiting agency conflicts through improved external monitoring (Bushman & Smith, 2001). Accomplishing the higher quality reporting in Malaysia, Malaysia Accounting Standard Board (MASB) has been converging Malaysia financial reporting standards to IFRS. Barth et al. (2008, p.491) interprets “accounting amounts that are more value relevant as being of higher quality”. Barth et al. (2008) also determine that if the firms that apply IAS/IFRS have higher accounting quality than firms that do not; and that accounting quality improves after the firms have adopted IAS/IFRS.

Financial statements are regarded as being of high quality if they have three attributes: transparency, comprehensive disclosures and comparability (Pownall and Schipper, 1999). Transparent financial statements means that “standards that reveal the events, transactions, judgments and estimates underlying the statements, and their implications” (Pownall and Schipper, 1999, p. 262). Comprehensive disclosures relate to providing of all mandatory or voluntary information in financial reports; consequently reasonable assurance is delivered to prevent misleading decisions by investors. Financial statements are comparable if they possess similar transactions and events are accounted and disclosed consistently. Comparability is one of the four qualitative characteristics underlying IFRS and it is important that an investor can analyse the annual reports of the different companies in
the same sector and be confident that the results are comparable, in order to make investment decisions.

Following Pownall and Schipper (1999) and Barth et al. (2008), this current study defines the accounting quality of firms, which have already adopted the International Financial Reporting Standards, is assessed by comparing with IFRS terminology disclosures (mandatory disclosures) included in model financial reports published by the International Financial Reporting Standard Board. Thus, this study defines the good practice of IFRS compliance practice based on the disclosure contents included in annual reports. IFRS compliance practice is defined here as the firms’ compliance with the IFRS mandatory disclosure requirements of the model of the International Financial Reporting Standards (IFRSs) issued by IASB and disclosing them together with the other compulsory requirements in their annual financial reports.

In other words, the higher the level of IFRS compliance in the annual reports, the better the quality of accounting. The measurement method of IFRS compliance practice is defined as its attributed property of accounting quality is consistent with the measures of accounting quality used by Bova & Pereira. (2012) who in turn adopted the measures applied in Kenya, Financial Reporting Award (the FiRe Awards). The measures of accounting quality applied in this current study are slightly different from their study, in that local factors and contemporary issues of IFRS have been taken into consideration in the setting of measurement attributes by using the fourteen items of IFRS mandatory disclosure and other voluntary disclosures such as corporate social responsibility and investor relationship information.

Accounting quality with value relevance studies were widely used to capture the combined attributes of reliability and accuracy of accounting information (Barth, 1994; Schipper & Vincent, 2003; Bartow et al., 2005). Daske et al. (2008) examined the heterogeneity in economic consequences of IFRS adoptions around the world, by distinguishing that firms have considerable option to adopt IFRS, mandatory or voluntary adoption,. They also found that the economic consequences of IFRS adoptions depend on the extent to which firms make material changes to their accounting policies. Kinsey et al. (2006) and Jermakowiez et al. (2007) extended the work of Dask et al (2008) by defining heterogeneity as common law versus code law countries abide by to adopt IFRS. Malaysia is a common law country.

The value relevant issue is well-known in the accounting literature (Pope, 1992). Value relevance means that there is a statistical association between financial information and stock prices or returns (Agostino et al. 2008). The value relevance is described as “value relevance is the ability of financial statement information to capture or summaries information that effects share values” (Hellström, 2006, p.325). The literature reveals that the accounting quality measures in prior research are mostly in all financial value relevance. Some country specific studies examined the value relevance of certain accounting items and the earnings focused on the regression of returns on earnings based on U.S GAAP, IAS and German GAAP (Clinch, 1998; Bartow, 2005).

Many studies on value relevance are widely conducted by prior researchers to capture the combined attribute of reliability and accuracy of accounting information (Schipper, 2003). The value relevance of different accounting principles has been explored in the accounting literature by using either an event study or an association analysis (Holthausen and Watts, 2001). Jermakowiez, et al. (2007) studied the value relevance of the DAX-30 German companies to determine whether voluntary adoption of IFRS caused in more value relevance after adoption of IFRS as compared to before. They analysed the association between book value and the market value of equities using a long window (value relevance of 12 months) and short window (information content within 6 days of earning announcements) period. Many previous studies on this association investigate if financial reporting data explains market capitalisation and changes over long window period of time (Barth, 1994; Choi et al. 1997). Ohlson model (Ohlson, 1995; Feltham and Ohlson 1995) stipulates that share prices can be written as a linear function of earnings and equity book value, given a dividend valuation model; and it has been employed in a number of studies. (Chandra and Balachandran 1992, Harris et al. 1994, Easton and Sommers 2003). Many European accounting researches expected to have an impact on the relationship between accounting data and stock prices (Devalle et al., 2009). Hung and Subramanyam (2007) indicated that the implication of IFRS adoption for improving accounting quality and its value relevance should be assessed.

3. METHODOLOGY

This study extends prior research by focusing on the implementation of IFRS at the firm level rather than accounting standard regulator or issuer because the accounting standards which have been issued by MASB originated from IAS and IFRS, and the scope of terminologies and measurement criteria are compatible with IFRS1. The assessment method of accounting quality

1 Sources: Accounting standards of different countries are available on http://www.iasplus.com, http://www.masb.org.my
with high IFRS compliance in this study is novelty and operational content analysis, to assess directly the accounting quality with mandatory disclosure requirements of IFRS, appearing in the firms’ annual reports.

We refer to the idea of financial value relevance ratios to be the proxies as the independent variables which can predict the accounting quality with higher IFRS compliance in our accounting quality model. Among the financial value relevance, current ratio is included instead of fund flow ratio as applied in the prior research, as a proxy for liquidity position of a firm. Current ratio reflects the test of accrual concept more than fund flow ratio does. The financial value relevance ratios are leverage ratio and earnings ratios consistent with prior research of Hribar et al. (2009), used as control variables in our accounting model.

Our study introduces content analysis on accounting quality with the level of IFRS compliance based on the disclosure requirement of fourteen IFRSs shown in Table 1. It focuses on the required disclosure statements regarding 14 IFRSs and voluntary statements such as corporate social responsibility statement in the firms’ financial reports, to assess the level of accounting quality with IFRS compliance by using predetermined measurement index (Appendix 1). It also examines the equity stock market reflection on the international financial reporting as an issue of accounting quality with higher IFRS compliance. Stock price reflection of the firms is analysed as informational efficiency in equity stock markets of the firms in emerging market like Bursa Malaysia.

The following are hypothesized to achieve the research objectives of study has mentioned in the introduction:

H₁: International Financial Reporting Standard compliance practice by the firms listed on main market of Bursa Malaysia are significantly higher.

H₂: Accounting quality with IFRS compliance practice by the firms is significantly associated with value relevance indicators.

These hypotheses are to be tested by using the research modules (equation 1 and 2) shown in the following section, analytical techniques and empirical analysis.

3.1 Sources of Data and Data Screening

Our sample consists of the non-financial companies listed on the main market of Bursa Malaysia. There are 637 companies from different industrial sectors, listed in the main market of Bursa Malaysia in 2010. Since the study mainly focuses on non-financial corporations and their quoted equity stocks; systematic sampling procedure is employed to get the required sample size. Finally, the sample consists of 158 firms from the four sectors of main market: Commerce, Industry, Construction, and Trade and Service sectors. This study only focuses on the non-financial corporations from the four major sectors of Bursa Malaysia because the total number of constituent companies in these sectors can provide the required information of annual reports and their stock prices are also available in the internet.

The annual reports of 158 non-financial corporations, for the fiscal years from 2005 to 2010, were downloaded from their websites, to assess the level of IFRS compliance practice. The measures of the other proxy variables are based on the information from the fiscal year 2010 annual report of an individual firm so that the measures taken on the accounting quality with higher IFRS compliance are consistently fair by having the same situation to all firms after mandating the accounting standards in Bursa Malaysia in 2005.

Some companies from each sectors have no available website; some have websites however annual reports are not available within the first six months in 2010 (2010 annual reports should be published during the first 6 months, 1st January to 30th June, 2010), those companies are excluded. Final selection determines all available equity stock of non-financial corporations from the four industrial sectors and come up with the systematic random sample size of 158 non-financial corporations and their equity shares.

3.2 Analytical Techniques and Empirical Analysis

The analytical techniques adopted by this study aims to find out the determinants of IFRS compliance practice, in the online financial reports released by the non-financial corporations in their websites linked to main market of Bursa Malaysia. The methodology consists of quantitative method and triangulation approach. Content analysis and multiple regression analysis are carried out based on the cross sectional data collected from annual reports 2010 of the firms. Content analysis assesses the level of IFRS compliance practice based on the disclosure statements related to 14 IFRSs. The content analysis uses the measures on the disclosure requirements of fourteen IFRS standards which include property plant and equipment, business combination, tangible assets, impairment of assets; stock and long term contracts; deferred tax; leases; retirement benefits; marketable securities, research and development expenditure, revenue recognition, contingent liabilities and contingent assets, segmental reporting, and related party disclosure (shown in Appendix 1). The reason of choosing these standards is concerned with the consistent assessment on the level of IFRS compliance in annual reports of all non-financial companies from the different industrial sectors comprised in the observation. These measures are the properties of accounting quality with higher IFRS compliance.
Validity and reliability of measuring the higher level of IFRS compliance, content analysis program is constructed by setting queries on key disclosure requirements of each and every fourteen IFRS standard in the QSR NVivo 9 software (Qualitative Data Analysis). The annual reports of the firms in word files are saved in the input unit of the software; and a standard format of annual report issued by the International Accounting Standard Board is stored in the memory storage area of the software. The content analysis is run based on the queries of mandatory and voluntary disclosure statements (Appendix 1) of a particular IFRS standard.

We ran the queries in the software to examine the contents of disclosure requirements to find out an individual firm’s compliance percentages of each IFRS standard for a firm. Then, the total compliance percentages of the fourteen IFRS compliance scores are computed as a percentage of the highest score possible for the firm. The following equation describes the mathematic formula to calculate the weighted average of 14 IFRS compliance scores of a firm; the method is also consistent with the studies of Street & Gray (2002) and Al-Shiab (2008).

\[
\text{AccQ}_{it} = \frac{\sum_{j=1}^{14} X_j}{S_i} \quad (Eq.1)
\]

Where \( j \) means a score of single IFRS, from one to fourteen items of IFRS; \( X_j \) is an attribute of IFRS measures. The total number of applicable IFRS standards for a firm \( (S_i) \) divides the sum of these compliance scores for each company. The \( \text{AccQ}_{it} \) is the mean value of the total IFRS compliance of firm \( i \) at time \( t \). The mean value of the total IFRS compliance is calculated as \( 0 < \text{AccQ}_{it} < 1 \). This calculation method is giving unequal weighting to the disclosure items in 14 different IFRS standards. It is logically considered that real situation of different firms might imply unequally for each IFRS in their annual reports.

### 3.3 IFRS Compliance Practice in Malaysia

Using a comprehensive disclosure index of selected IFRSs for which previous research stated as significantly considers that a firm’s annual report disclosure at least 65 percent of IFRS compliance in annual reports as higher IFRS compliance, and any number lower as lower level of IFRS compliance. We use the accounting quality with IFRS compliance dataset of 158 companies with 14 IFRS standards and reaches 2212 of the total observations. To assess the level of IFRS compliance in Bursa Malaysia, this study postulates the hypothesis that accounting quality with non-financial corporations in main market of Bursa Malaysia may highly comply with the International Financial Reporting System (IFRS).

The result of the content analysis provides a data set of the level of IFRS compliance of the firms. The data set of the accounting quality with the level of IFRS compliance the firms is used to run the analysis of mean value that shown t-statistic (14.89) is significant at 1% level. Thus, this provides the significant evidence to conclude that the accounting quality with the higher IFRS compliance positively differs from mean value of IFRS compliance level (65%). This suggests that the postulated hypothesis has been accepted and it concludes that the non-financial corporations listed on the main market of Bursa Malaysia highly have a higher level of accounting quality with IFRS compliance. It may confirm that they have implemented the higher IFRS compliance in their financial reporting.

### 3.4 Financial Value Relevance Indicators

Additionally, this study extends to examine the association between accounting qualities with IFRS compliance and financial value relevance; the proxies for financial value relevance included in the following model concern with institutional investors’ interests in the market. Financial value relevance is reflected by the following: leverage ratio (LeveRatio) and net asset backup per share (NetAssetPS) as proxies for firm financial position; return on equity (ROE) a proxy for the profitability; earnings per share (EPS), price earnings ratio (P/E), share growth rate (DGR) as proxies for investors’ ratios; and firm’s current ratio (FCR) as a proxy for the liquidity position. Computations of these variables are based on the data available from 2010 annual reports of the firms.

The following hypothesis is postulated for this purpose: A firm’s financial value relevance may predict the IFRS compliance practice. A regression model as shown below is to arrive at this objective.

\[
\text{AccQ}_{it} = 0.737 + (-0.008)\text{LeveRatio}_{it} + 0.002\text{NetAssetPS}_{it} + 0.001\text{ROE}_{it} + 0.001P/E_{it} + 0.027DGR_{it} + 0.001CR_{it} + \epsilon_{it} \quad (Eq.2)
\]

Where, \( \text{AccQ}_{it} \) = the accounting quality with IFRS compliance practice of \( i^{th} \) firm at time “t” annual report;

\( \text{LeveRatio}_{it} \) = Leverage Ratio of \( i^{th} \) firm at time \( t \);

\( \text{NetAssetPS}_{it} \) = Net asset backup per share of \( i^{th} \) firm at time \( t \);

\( \text{ROE}_{it} \) = Return on Equity of \( i^{th} \) firm at time \( t \);

\( \text{EPS}_{it} \) = Earnings per share of \( i^{th} \) firm at time \( t \);

\( P/E_{it} \) = Price earnings ratio of \( i^{th} \) firm at time \( t \);

\( DGR_{it} \) = Dividend growth rate of \( i^{th} \) firm at time \( t \);

\( CR_{it} \) = Current ratio of \( i^{th} \) firm at time \( t \); and

\( \epsilon_{it} \) = the intercept.
\[ \beta_1, \beta_2, \beta_3, \ldots, \beta_7 = \text{the coefficients of partial correlation on factors (1-7); and } \varepsilon_{1u} = \text{residual errors.} \]

4. FINDINGS AND DISCUSSIONS

Spearman and Pearson correlations among the variables resulted in the regression model are shown in Table 1. There is no multicolinearity relationship and the model concerned with mitigating the limitation of multiple regression analysis. Table 2 presents the results of the analysis using the multiple regression models; the ANOVA results show that F statistics 4.697 is significance at 1% level of probability and the R square is 42%. The variables, returns on equity, earning per share and current ratio are not statistically significant. However, leverage ratio, price earnings ratio, share growth rate and current ratio are statistically significant. Accounting quality regresses on Leverage ratio, Earnings per share, Price Earnings ratio and dividend growth rate. (see Table 2).

\[ \text{AccQ}_{it} = 0.737 + (-0.008)\text{LeveRatio}_{it} + 0.002\text{NetAS}_{it} + 0.001\text{ROE}_{it} + 0.001\text{P/E}_{it} + 0.027\text{DGR}_{it} + 0.001\text{PCR}_{it} + \varepsilon_{2u} \quad (Eq.3) \]

The results highlight that accounting quality with higher IFRS compliance is negatively associated with the leverage ratio of a firm and positively associated with price/earnings ratio and share growth rate. It suggests that higher quality with IFRS compliance underlines with prudence concept, the more compliance with IFRS can covey meeting prudent concept. A limited number of prior research conducted in EU markets also mentioned that IFRS compliance and leverage ratio is negatively associated such as Kothari et al. (2005). However, they did not mention about prudence concept. The finding of our study could conclude that a significant negative association between leverage ratio and the higher IFRS compliance indicates the prudence concept underling of IFRS system. Moreover, the accounting quality with higher IFRS compliance is significantly and positively associated with profit indicators such as earnings per share (EPS), price earnings ratio (P/E ratio) and dividend growth ratio (DGR) at 5% significant level. Therefore, from these findings, it might infer that accounting quality with higher IFRS compliance delivers the benefits of reliable information for the investors and it might be evading financial crisis in the Bursa Malaysia.

5. CONTRIBUTION AND CONCLUSION

This study contributes an evidence of the existence of higher accounting quality with IFRS compliance practice and it has value relevance in the equity capital market. The empirical results provide evidence for the existence of accounting quality with value relevance so that investors can rely on the financial information reported by the firms in Bursa Malaysia. The finding can contribute to the accounting standard setter to overview of their accounting standards published and companies are mandated to follow the standards. It can increase foreign direct investment by the International investors, and to achieve more foreign investment flows into Malaysia.

REFERENCES

German stock market perspective. *Journal of Accounting, Auditing and Finance* 20, 95-119.


TABLES

Table 1: Accounting Quality with Financial Value Relevance Correlation Matrix

<table>
<thead>
<tr>
<th>Variables</th>
<th>Expected Sign</th>
<th>Coefficient</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-</td>
<td>0.753</td>
<td>64.883</td>
<td>0.000</td>
</tr>
<tr>
<td>Leverage</td>
<td>-</td>
<td>0.098***</td>
<td>2.879</td>
<td>0.096</td>
</tr>
<tr>
<td>Market P/S</td>
<td>+</td>
<td>0.001</td>
<td>0.359</td>
<td>0.577</td>
</tr>
<tr>
<td>ROE</td>
<td>+</td>
<td>0.001</td>
<td>0.677</td>
<td>0.539</td>
</tr>
<tr>
<td>EPS</td>
<td>+</td>
<td>0.001***</td>
<td>2.476</td>
<td>0.014</td>
</tr>
<tr>
<td>P/E</td>
<td>+</td>
<td>0.001**</td>
<td>2.704</td>
<td>0.090</td>
</tr>
<tr>
<td>ROE</td>
<td>+</td>
<td>0.001</td>
<td>2.883</td>
<td>0.043</td>
</tr>
<tr>
<td>E</td>
<td>+</td>
<td>0.003</td>
<td>0.415</td>
<td>0.675</td>
</tr>
<tr>
<td>R</td>
<td>+</td>
<td>0.424</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROA</td>
<td>-</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard errors of estimate</td>
<td>0.071</td>
<td>1.826</td>
<td>4.697</td>
<td>0.000</td>
</tr>
</tbody>
</table>

This table reports the regression results of the accounting quality and financial value relevance. This study finds significant association between IFRS compliance practice (Actuality) (dependent variable) and financial value relevance (leverage ratio and share growth rate). *** Significant at 1% level, ** significant at the 5% level, and * significant at 10% level.

Acknowledgement: We gratefully acknowledge the financial support from the Faculty of Economics and Business and Fundamental Research Grant Scheme (Ref: FRGS/2/2013/SS05/UNIMAS/02/1).
ABSTRACT

This study examines the presence of Day of the Week effect in returns on four indices: Kuala Lumpur Composite Index, Mid 70, Top 100 and EMAS of Bursa Malaysia. It essentially investigates evidence for the premise that the rational investors and perfect markets exists. The daily returns over a period of 18 years from 1996 to 2014 are tested using both parametric and non-parametric statistics. The study finds strong evidence for day of the week effect, that the mean daily returns of every trading day of the week are not equal. There is no evidence for random walk hypothesis. Friday returns are positive and significantly higher than that of other week days. Monday returns are negative, and these results display no difference between test methods, parametric or non-parametric.

Key words: Day of the week, Anomaly, Efficiency, Stock, Malaysia.
JEL Code: G11, G12 and G14

1. INTRODUCTION

Since Eugene F. Fama coined the term “efficient market” in 1970 there has been numerous attempts to sustain the Efficient Market Hypothesis (EMH) which emphasizes that the asset prices fully reflect ‘available’ information. However, these attempts find no evidence in favour of EMH (Islam, Watanapalachaikul and Clark, 2007). The EMH holds that, as a result of competition, equilibrium prices in financial markets that incorporate all relevant information. Hence simple strategies cannot beat stock markets because the prices follow a random walk. If the stock prices follow a random walk, the market is said to be efficient and the prices are not predictable based on the past prices as the price would be dependent upon the new information. The fact that the return volatility is not independent of the time (i.e. time variant; the presence of calendar anomalies) evidences against random walk theory. If the information flows efficiently, investors share similar information creating no systematic way of exploiting trading opportunities, but an investor would outperform using unique strategies if anomalies exist. Therefore empirical studies on seasonality are at least motivated by (1) appetite for evidence for EMH and (2) information for judgments of managers and investors.

Financial theory issues such as volatility, predictability, speculation and anomalies are also related to the efficiency issue (Islam, Watanapalachaikul and Clark, 2007) and are all interdependent. Therefore, EMH tests have taken the form of testing the issues of predictability, anomaly, seasonality, volatility and the existence of bubbles. Seasonal, momentum and disposition effects in markets are some of the contrary evidence that has been the outcome of such empirical efforts, received by EMH. Some studies suggest that the limited support for EMH likely to be a result of inappropriate test methods, excess volatility mean reversions, non-linearity, anomalies and bubbles in the stock markets. For instance, Chien, Lee, and Wang (2002) state that the limitations of the existing efficiency test methods contribute to conflicting outcomes. The seasonality in stock markets could be attributed to inappropriate statistical methods. This makes the evidence on seasonality an ‘invalid challenge’ for EMH. In contrast, Number of studies including French (1980), Lankonishik & Levi (1982), Harris (1986), Wang (1997), Nippani & Greenhut (2011) confirm the distribution of stock returns according to the day of the week. Evidence on day-of-the-week effect also suggests that behavioural factors would be priced and makes invalid the ‘Homo Economicus’ assumption that the humans make perfectly rational decisions. Research on seasonal effects suggests an existence of crowded behaviour in markets even though the markets are efficiently adjusted to new information. Hence price adjustments do not show rational economic decisions and it is not the weaknesses of test methods but irrationality of the market.

Day of the week effect is the most widely documented among the seasonal effects, it has been comprehensively investigated in different markets (Rahman, 2009). Basdas (2011) observes that studies generally adopt regression analysis with dummy variables,
ANOVA, non-parametric tests or Generalized Auto Regressive modelling for testing seasonality effect. However, Chien, Lee, and Wang (2002) suggest non-parametric tests to avoid upward biasness in test statistics derived from dummy variable models, particularly when the return volatility of a period observed is higher than the other periods. The present study uses both parametric and non-parametric statistics in examining the seasonality. If the day of the week effect is true in an efficient market, the anticipations would make it difficult to make money out of seasonality. Nippani and Greenhut (2011) find a reversal of week end effect after a long period of week end effect, suggesting that the market follows seasonality in making money. According to Muhamad and Rahman (2010), Malaysian Stock exchange has shown day of the week effect in the past. Extending their work, the present paper examines the day of the week effect in Kuala Lumpur Stock Exchange with a view that if the day of the week patterns are anticipated in markets, anticipations could make the effect diminishing subsequently and disappearing later. More specifically, the objective of the present study is to test the Day of the Week effect in Malaysian Stock Exchange and examine whether results are conflicting under different test methods. The paper proceeds as: related literature in section 02, method of research in section 03, results in section 04 and conclusions in section 05.

2. LITERATURE REVIEW

Many studies confirm negative Monday returns and propose investor information asymmetry to be the potential reason. French (1980) shows a significant negative Monday returns in equity market on S&P 500; Jaffe and Westerfield (1985) show evidence from equity markets in US, UK, Canada, Australia, and Japan. Subsequently, Bishara (1989) confirm week end effect in Canada. Ruchika and Datta (2012) find evidence for day of the week effect in BRIC countries. Agathee (2008) find that the Friday returns are higher relative to other trading days, and mean returns over the week are not different from zero. These findings have no difference in emerging markets (Aggarwal and Rivoli, 1989). For instance, Gunathilaka (2013) reports negative Monday returns in Colombo, Muhammad & Rahman (2010) find Monday effect in Malaysia, Wong and Ho (1986) document day-of-the-week effect on the Singapore stock market. Wong and Ho (1992) further confirm the negative Monday returns in Singapore, Malaysia, Hong Kong and Thailand.

However these findings are not universal; cross country evidence make the researchers troublesome. For instance, negative Monday returns in Romania (Patev, 2003), in contrast positive Monday returns in Egypt (Aly, Mehdian and Perry, 2004), no week day effect in Mauritius (Rojid S, Seetanah B, 2009), limited day of the week effect in Istanbul (Basdas, 2011) and positive Monday returns while negative Tuesday returns in Indian capital market (Mahendra and Damini, 2006). Fan (2002) find that Friday returns are lower than other trading days. Nippani and Greenhut (2011) conclude that the week end effect, a long observed anomaly in Canadian market, shows a reversal from 1998. Reversals also can be a result of market developments. Ruchika and Datta (2012) observed that the stock markets became efficient after the introduction of future trading. Thus the integration of debt markets, derivatives and equity markets would bring the efficiency, as the market become more sensitive for full information. However, the equity market anomalies may not pertinent in debt markets, for instance, Jordan and Jordan (1991) discover no meaningful difference in mean daily returns for fixed income securities. It is also interesting to note that government securities, the so-called risk free securities, do display day of the week end effect as observed by Singleton and Wingender (1994) who confirmed Weekend effect for treasury securities, by examining closing prices of treasury securities of various maturities.

Some studies propose that the institutional investors are less likely to trade on Mondays narrowing width of the market and thus returns go negative (Abraham and Ikenberry, 1994). Rationally, the institutional investors’ role may be less active in markets where positive Monday returns are observed, for instance, Mahendra and Damini (2006) find positive Monday returns in India; Aly et al (2004) in Egypt. In markets where institutional investors play a greater role more pronounced day of the week patterns in returns and volumes are present (Richard and Laura, 1995). According to Chen and Singal (2003) the weekend effect is due to short sales. As short sellers cannot close their positions during week end, they tend to close their positions by the weekend to avoid the risk. Most importantly, they find that stocks with high short interest show a greater weekend effect than others. This suggests that random walk may not be applicable in the short run, as short sellers increase in a market. Further, Worthington and Higgs (2009) find random-walk behaviour in long term (monthly returns) than short term (weekly returns). On the other hand, anomalies in short term may be attributed to statistical methods used; over and under reactions become marginal or disappear in the long run. Chien, Lee, and Wang (2002) argue that the application of parametric models to test returns with different volatilities have contributed largely to seasonality evidence. Hence if the returns are time-dependent in Malaysian context, the question of its continuation even after markets are integrated may draw attention to test methods as well.
3. METHODOLOGY

This study uses the returns on indices of four portfolios as reported by Bursa Malaysia: Kuala Lumpur Composite Index (KLCI), Mid 70, Top100 and EMAS. KLCI is a market capitalization weighted index that is a portfolio of 30 largest companies on main market of Bursa Malaysia. While KLCI consist of 30 largest companies of the market, the next 70 companies in the main market are included in Mid 70 index, and the sum of these two constituents are reflected in Top100 index. All these indices do comply with FTSE liquidity and investability standard criteria, and are jointly operated by FTSE and Bursa Malaysia. The FTSE Bursa Malaysia EMAS index is the sum of all the constituents of Top100 and FTSE Bursa Malaysia Small Cap index. Hence these four indices cover a substantial part of the entire market.

The sample period includes daily returns over approximately 18 years from 1st January 1996 to 31st March 2014. The time series of these indices are obtained from Datastream database of Thompson Reuters. The behaviour of the indices over the sample period is depicted in the figure 01.

Figure 1: The Returns on the indices over the sample period.

3.1 Model Specification

The dummy variable regression model has been extensively employed in prior studies to investigate seasonal effects in stock markets (Chien, Lee and Wang, 2002). The model regresses period dummies with the returns under the assumption that the variances of stock returns are constant throughout the test period. The model is expressed as;

\[ R_t = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \beta_5 D_5 + e_t \]

Model 01

Where,

\( R_t \) is the continuous daily market return as captured by an index I and measured as;

\[ R_t = \ln \frac{I_t}{I_{t-1}} \]

D1 – D5 are Dummy variables equal to 1 if t is a tested day and 0 otherwise;

\( \beta_1 - \beta_5 \) are average return of each respective day;

\( e_t \) is the stochastic disturbance term.

The hypothesis is that, in absence of any day of the week effect in returns, the variance of average return of each respective day which is the slope of D1 to D5 should not be different from zero. (H0=\( \beta_1=\beta_2=\beta_3=\beta_4=\beta_5 \)). Model is estimated with no common intercept to represent the average return of the observed day (where the dummy is equal to 1), if the model is estimated with a common intercept, the slope of the exogenous variable/s in the model should measure the difference between the average return of the observed day and other day/s, which should not be different from zero if there is no day of the week effect. Hence, the slope would measure the difference between average return in the day observed and the rest of the days. If the average return in the day of the week is the same as that of the rest, then the estimate of slope (\( \beta_d \)) should not be different from zero. In other word, the common intercept would represent the average return of the week, commonly.

One of the conditions required for random walk is that the return series should contain unit root (Mishra, 2010; Gahlot & Datta, 2012). If the series at first difference is stationery, the series follow weak form efficiency. Hence, the study examines the Null Hypothesis of unit root (\( \delta = 0 \)) at level and first difference in the time series of returns on the indices using Augmented Dicky-Fuller test which takes the following specification. Hence the alternative hypothesis is \( \delta < 0 \).

\[ \Delta R_t = \alpha + \beta R_t + \gamma R_{t-1} + \delta_1 \Delta R_{t-1} + \ldots + \delta_p \Delta R_{t-p} + e_t \]

Model 2

Where, \( R_t \) is the return at time t,

\( \alpha \) is a constant intercept,

\( \beta \) is the coefficient on a time trend and the lag order of the autoregressive process.
Chien et al. (2002) conclude that the widely used dummy variable regression model may wind up the results and other methods like ANOVA, pair-wise comparison approach would be appropriate for seasonality testing. They show that the test statistics of dummy variable regression model is biased upward. In particular, if the return volatility is relatively higher in the test period than the rest of the periods, the regression statistics tend to reject the null hypothesis incorrectly. The present study examines the seasonality of stock returns of Malaysian capital market with parametric methods and suggested nonparametric statistics in order to determine whether it is factual anomalies or inappropriate statistical methods. Hence the study performs a hypothesis test of the equality of population medians for a one-way design using the Kruskal-Wallis test, where the test statistic (H) indicates equality; $\bar{I}1 = \bar{I}2$, where the $\bar{I}$ is the population medians.

4. FINDINGS AND DISCUSSIONS

The table 1 gives the descriptive statistics of the four indices used in the study. The highest average returns are drawn from the large 30 portfolio while both highest Friday positive returns and largest Monday negative returns are drawn from Mid70 portfolio. Negative Monday and positive Friday returns are observed for all four indices. On average, Friday positive returns are significantly higher than other days.

Table 1: Descriptive Statistics of the Daily Returns of the Sample Period (1996 – 2014)

<table>
<thead>
<tr>
<th></th>
<th>LCI (E(R))</th>
<th>ID70</th>
<th>TOP100</th>
<th>EMAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Monday</td>
<td>0.0881</td>
<td>0.1354</td>
<td>0.0897</td>
</tr>
<tr>
<td></td>
<td>Tuesday</td>
<td>0.0104</td>
<td>0.0152</td>
<td>0.0136</td>
</tr>
<tr>
<td></td>
<td>Wednesday</td>
<td>0.0548</td>
<td>0.0508</td>
<td>0.0557</td>
</tr>
<tr>
<td></td>
<td>Thursday</td>
<td>0.0062</td>
<td>0.0071</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>Friday</td>
<td>0.0793</td>
<td>0.1399</td>
<td>0.0896</td>
</tr>
<tr>
<td>Mean</td>
<td>5 Days</td>
<td>0.0190</td>
<td>0.0094</td>
<td>0.0136</td>
</tr>
<tr>
<td>Std dev</td>
<td>5 Days</td>
<td>0.3450</td>
<td>0.3800</td>
<td>0.3320</td>
</tr>
<tr>
<td>Maximum</td>
<td>5 Days</td>
<td>0.810</td>
<td>9.980</td>
<td>1.270</td>
</tr>
</tbody>
</table>

KLCI, MID 70, TOP 100 are composite index of Bursa Malaysia. 4750 Total number of observations

4.1 Random Walk

The table 2 shows the results of the Augmented Dickey-Fuller (ADF) unit root test, where the statistic $t_\alpha$ is less than the critical value at 1% level and is significant at 1% level in all the circumstances. Thus the series both at level and first difference show no unit root (i.e. the returns are stationary). This suggests that the Bursa returns do not follow a random walk, and the market follows a weak form. The ADF test equation further explains that the constant and trend are insignificant, the series at level are mean centred and autocovariances do not depend on time.

Table 2: Unit root tests

<table>
<thead>
<tr>
<th></th>
<th>KLCI</th>
<th>TOP100</th>
<th>MID70</th>
<th>EMAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.*</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Critical value: 1% level</td>
<td>-3.4315</td>
<td>-3.4315</td>
<td>-3.4315</td>
<td>-3.4315</td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>2.0008</td>
<td>2.0007</td>
<td>1.9990</td>
<td>2.0006</td>
</tr>
<tr>
<td>F-statistic</td>
<td>828.3676</td>
<td>708.1419</td>
<td>692.9364</td>
<td>692.9364</td>
</tr>
<tr>
<td>P-value</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>First Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF test statistic</td>
<td>-24.7725***</td>
</tr>
<tr>
<td>Prob.*</td>
<td>0.0000</td>
</tr>
<tr>
<td>Critical value: 1% level</td>
<td>-3.4315</td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>2.0024</td>
</tr>
<tr>
<td>F-statistic</td>
<td>890.5910</td>
</tr>
<tr>
<td>P-value</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

*MacKinnon (1996) one-sided p-values. *** significant at 1 per cent, ** significant at 5 per cent and *significant at 10 per cent.

4.2 Dummy Variable model

The panel A of the table 3 represents the results of dummy model estimated; coefficients and their corresponding p-values. Accordingly, coefficients for Monday negative returns and Friday positive returns are significant at 5 per cent level in all the four indices while Wednesday returns are positive and insignificant in all four indices. Remaining day returns are mixed and insignificant.
Table 3: Regression Results
Dependent variable: Rt, Sample (adjusted): 1January 1996 to 14 March 2014; Includes observations: 4750 after adjustments.

Panel A: OLS estimations

<table>
<thead>
<tr>
<th>Day of the week</th>
<th>KLCI Coefficient (Prob.)</th>
<th>Mid 70 Coefficient (Prob.)</th>
<th>Top100 Coefficient (Prob.)</th>
<th>EMAS Coefficient (Prob)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>-0.0010** (0.0141)</td>
<td>-0.0014** (0.0025)</td>
<td>-0.0090** (0.0380)</td>
<td>-0.0011** (0.0089)</td>
</tr>
<tr>
<td>Tuesday</td>
<td>0.0003 (0.4942)</td>
<td>-0.0002 (0.7333)</td>
<td>0.0001 (0.7535)</td>
<td>0.0000</td>
</tr>
<tr>
<td>Wednesday</td>
<td>0.0009 (0.2028)</td>
<td>0.0006 (0.2562)</td>
<td>0.0006 (0.1975)</td>
<td>0.0005</td>
</tr>
<tr>
<td>Thursday</td>
<td>8.08E-05 (0.843)</td>
<td>7.08E-05 (0.8743)</td>
<td>-1.11E-05 (0.9795)</td>
<td>0.0058</td>
</tr>
<tr>
<td>Friday</td>
<td>0.0011** (0.0058)</td>
<td>0.0014** (0.0018)</td>
<td>0.0009** (0.0383)</td>
<td>0.0011** (0.0154)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0031</td>
<td>0.0042</td>
<td>0.0020</td>
<td>0.0021</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>1.8523</td>
<td>1.7269</td>
<td>1.8533</td>
<td>1.8301</td>
</tr>
</tbody>
</table>

*** significant at 1 per cent, ** significant at 5 per cent and *significant at 10 per cent.

Thus, the estimations confirm the day of the week effect invalidating the random walk hypothesis in Bursa Malaysia. As Chien, Lee, and Wang (2002) argue that the test statistics of the dummy variable regression model are biased upward and estimators are identical only at the maximum likelihood only (i.e. the variance of the population T should be equal to variances of ‘p’ observations representing a tested day where the dummy is 1, and ‘q’ observations representing other days where the dummy is zero). Therefore pair-wise comparison should be performed to investigate the difference between any two means of factor levels.

4.3 Pair-Wise comparisons

The table 4 reflects the results of non-parametric test: Kruskal-Wallis statistic. Table 4 states the result of the test that there is a variation between Monday and other days effects; similarly Friday and other days effects are also different, these finding are also highly significant and consistence with the above findings.

Table 4: pair-wise comparisons

<table>
<thead>
<tr>
<th></th>
<th>Pair: EMAS</th>
<th>Median</th>
<th>Avg. Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>50</td>
<td>0.0000167</td>
<td>2418.0</td>
</tr>
<tr>
<td>Other days</td>
<td>800</td>
<td>0.0000000</td>
<td>2205.7</td>
</tr>
<tr>
<td>Test statistic (H) (Prob.)</td>
<td>18.21</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Friday</td>
<td>50</td>
<td>0.0005888</td>
<td>2510.7</td>
</tr>
<tr>
<td>Other week days</td>
<td>800</td>
<td>0.0000000</td>
<td>2341.7</td>
</tr>
<tr>
<td>Test statistic (H) (Prob.)</td>
<td>1.55</td>
<td>.001</td>
<td></td>
</tr>
</tbody>
</table>

5. CONTRIBUTION AND CONCLUSION

This study contributes an evidence of the existence of days of the week effect on return and volatility of indices in Bursa Malaysia. The empirical results provide evidence for the existence of the days of the week anomaly for all indices consistently and persistence. Portfolio rebalancing, information processing, and lack of information seem to be effective on higher volatility and lower returns of Mondays. Days of the week anomaly in Malaysia is especially interesting as it could support the proposition that these anomalies are general, world-wide phenomena rather than the result of a special type of institutional arrangement in Malaysia.

In conclusion, the findings of these calendar anomalies in all indices have important implications for practitioners and academics. For practitioners, it affects the asset allocation, hedging decisions and the timing of security issuances by firms. For academics, it has implications for asset pricing and performance evaluation. Moreover, it is highlight that these market indices in Bursa Malaysia are consistently volatile that can contribute to further research by answering to the question of which market index is the most reflect in the equity market.

REFERENCES


Acknowledgement: We gratefully acknowledge financial support from the Faculty of Economics and Business and Fundamental Research Grant Scheme (Ref: FRGS/2/2013/SS05/UNIMAS/02/1).
A STUDY ON INFLUENCE OF JOB STRESS FACTORS ON ORGANIZATIONAL EFFECTIVENESS

Ho-Soung Kong
Department of Business Administration
Namseoul University
Cheonan, Chungcheongnam-do, S.KOREA

ABSTRACT

In this Research, objectives are to review efficient job management plan by analyzing what impacts and relationship are provided mutually on decision-making participation and organizational trust, organizational effectiveness respectively through task characteristics, role overload, role conflict and role ambiguity, which are job stress factors, with the target of workers of Korean mid-sized firms.

It was investigated that Korean corporations need efforts to reduce job stress factors that are expressed in job stress by reducing job stress through clear job description of their conducting jobs through job analyses. Second, it was found out that Korean corporations are emphasizing importance of humanistic business management so that workers are informed of situations of organizations and are invited to participate in decision-making. Third, in order to possess and maintain high quality human resources who are bases of competence that is a key of organizations by reducing job stresses, it is required to manage organization trust of organization members. Fourth, it was resulted that organizational trust, and organization effectiveness have impacts relationships and organization trust was found out as an element which provides the most impacts on organizational effectiveness.

THEORITICAL CONSIDERATIONS

Job Stress

Job stress can be defined as requirements in work roles possessing extreme or harmful characteristics. In addition, it is harmful physical and emotional reaction which is generated when job requirements do not comply with capabilities or resources and needs of laborers.

As task characteristics decides their attitude and behavior, according to nature of task people take charge, when people find interests in tasks, then motivation is provided and they feel satisfactory about their jobs and also can increase accomplishments. So it can be said that task characteristics are primary products of organizations’ behaviors and processes and provide main impacts on organization culture.

Role ambiguity means uncertainty of job implementation which appears as information required for implementing jobs is not provided sufficiently. Role ambiguity can be generated when supervisors’ appropriate information transmission is poor, in case of confused information transmission, when behavior that organization members must take is uncertain and expected result of specific role is not clear.

According to Matterson & Ivancevich(1993), role conflict is generated when job requirements do not comply with individual standards and value of employees and job requirements, etc or when jobs require behaviors conflicting to moral and value system of individuals.

Role overload is also called as business overload and means “levels of taking charge of excessive burdens irrationally due to time pressure and quantity of works.” Role overload can be viewed from two aspects. First, it is quantitative overload and it is generated when too much works are assigned or sufficient time for
jobs is not allowed. Second, it is qualitative overload and it appears when employees feel that they do not have technical capability and authority required for job implementation.

Because majority of people think that they have lots of knowledge about their jobs, they experience significant amounts of stresses if they are denied to participate in decision-making processes related to their jobs.

It was found out that increase of participation in decision-making is deeply related to not only job satisfaction but also to employees’ health and welfare. If quality of participation is increased as possible as can and if clear expression of opinions is exchanged mutually, negative results such as stresses or its related tensions will be reduced. (Yang Choon-Hee et al., 2005)

Organizational Trust

Trust means firm belief and dependence about quality and characteristics of people and things or truth of statements and is necessary for maintaining stable and social relationship and also means nature or basic orientation of individuals who easily trust other people or things (Lee Kyung-Ho, 2010).

Fairness means people feel and recognize that procedure or standards are applied fair and consistent and are classified into procedure fairness, distribution fairness and interaction fairness.

Consideration means that conveniences and cares are provided to organization members based on spirit of human oriented business management so that employees do not have inconveniences in organization and family life and organizations support growth and development of each individual and can be classified into two dimensions.

First, people feel consideration when they feel sense of growth of and development of each individual and when the base for self-fulfillment of desires is built and second, when organizations provide members with conveniences and have no inconveniences in implementing jobs as organizations pay attention to work environment.

Openness means whether there is communication between organization members within organizations, establishment of family mood, whether organizations are innovative and creative organizations and whether organizations are responding to business management environment flexibly (Lee Kyung-Ho, 2010). And it also means that organization can have pride of organization for their members with tangible and intangible characteristics which can occupy competitive superiority.

Organizational Effectiveness

Organizational effectiveness means levels of belief to accept goal and value of organization through achievement of businesses by organization members and levels of passion to stay as organizationnal members with spontaneity exerting efforts for organizations and is expressed in royalty and devotion for organizations. (Mowday, 1979)

Job satisfaction is attitude about jobs. Importance of job satisfaction is as following. First, job satisfaction is that happy workers are efficient, innovative and care a lot. Second, people who like their jobs continue to maintain smooth human relationships inside and outside of organizations. Third, viewing from the position of organizations, if job satisfaction is increased, turnover or absent rate is reduced and also can obtain effects of increase of productivity.

Organizational commitment provides impacts on creation of accomplishments of employees more than job satisfaction and if level of organizational commitment is higher, possibility of increase of accomplishment is higher. And organizational commitment becomes a useful index for organizational effectiveness. (Cho Guk-Haeng, 2007).

Organization immersion can be classified into emotional and calculative. Emotional organization immersion means royalty and solidarity with organizations based on belief for goals and values of organizations, will to try efforts for organizations and strong will to stay continuously in organizations. Calculative organization immersion means royalty for organizations based on external compensations such as remuneration or promotion, etc.

DESIGN OF RESEARCH STUDY

Model of Research

In this research, we have analyzed by establishing models as below with regard to impacts of task characteristics, role overload, role conflict, role ambiguity and lack of decision-making participation, which are factors of job stress, on organizational trust and organizational effectiveness. Research model for implementing actual proof research is like Figure 1.

Hypothesis

Structural relationships between job stress factors and organizational trust and organizational effectiveness

<Hypothesis 1> Task characteristics will provide plus(+) impacts on organizational effectiveness.

<Hypothesis 2> Role overload will provide minus(-) impacts on organizational effectiveness.

<Hypothesis 3> Role conflict will provide minus(-) impacts on organizational effectiveness.

<Hypothesis 4> Role ambiguity will provide minus(-) impacts on organizational effectiveness.

<Hypothesis 5> Lacks of decision-making participation will provide minus(-) impacts on organizational effectiveness.
Hypothesis 6: Task characteristics will provide plus (+) impacts on organizational trust.

Hypothesis 7: Role overload will provide minus (-) impacts on organizational trust.

Hypothesis 8: Role conflict will provide minus (-) impacts on organizational trust.

Hypothesis 9: Role ambiguity will provide minus (-) impacts on organizational trust.

Hypothesis 10: Lacks of decision-making participation will provide minus (-) impacts on organizational trust.

Relationships between organization trust and organization effectiveness

Hypothesis 11: Organizational trust will provide plus (+) impacts on organizational effectiveness.

EVIDENTIARY ANALYSIS

In order to find out demographic statistics characteristics of total 160 samples used in statistics analysis of this research, we have carried out Frequency Analysis by using SPSS 18.0 statistics package program for finding out demographic statistics characteristics and for trustworthiness, we have used Cronbach’s alpha value and carried out factors analysis by using AMOS 18.0 statistics package program for confirming a single dimension.

For finding out impacts of stress factors on organization trust, knowledge sharing and organization effectiveness, by using Structural Equation Model (SEM), we have analyzed cause-and-effect relationships. For appropriateness judgment, we have evaluated research unit appropriateness centering on X2, GFI, AGFI, RMR, NFI and CFI, etc.

To judge internal consistency of variables used in this research, trustworthiness verification was carried out. Internal consistency or not can be confirmed through Cronbach’s Alpha Coefficient Value and in general, if it is over 0.6, trustworthiness is guaranteed. (Hair et al., 2006). As results of analyses, it is judged that internal consistency is secured because values over 0.7 were induced for all factors, (See Figure 2, Figure 3)

CONCLUSION

In this research, we have reviewed efficient job management plans by analyzing what impacts and relationships are provided mutually by task characteristics, role overload, role conflict and role ambiguity, which are job stress factors of laborers of mid-sized firms, on lacking decision-making participation, organizational trust, and organizational effectiveness respectively. Based on the above research, we have summarized verification results of hypothesis obtained through analysis of actual proof and suggest strategic implications.

First, it is situation that Korean corporations do not have preparation of proper clear job description and job description about conducting job through job analyses. In order to increase accomplishments of organizations, efforts to reduce emergence of job stresses factors by reducing job stresses are required.

Second, it is significantly meaningful that importance of humanistic business management informing situations of organizations to members and inviting to participate in decision-making is being emphasized.

Third, it is necessary to manage organization trust of organization employees for possessing and maintaining high quality human resources that are base of competence which is a key of organizations by reducing job stress.

Fourth, organizational trust and organizational effectiveness have impacts relationships and as an element which provides most impacts on organization effectiveness, organization trust (0.351) can be listed. Therefore, organizations require efforts to build open and initiative organization culture and communication activation plan for inducing solidarity of organization members in order to increase organization trust.

REFERENCES


Cho Guk-Haeng, 2007, Fairness and organization immersion, job satisfaction and organization effectiveness, Korea Academic Information.


Korea Occupational Safety and Health Agency, 2005, Job stress evaluation management.


Miles R., 1975, An empirical test of causal inference between role perception of conflict and ambiguity and
various personal outcomes”, *Journal of Applied


**FIGURES AND TABLES**

![Figure 1 Research Model](image1)

**Figure 1 Research Model**

![Figure 2 Structure Equation Model Analysis Result](image2)

**Figure 2 Structure Equation Model Analysis Result**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Direction</th>
<th>Coefficient</th>
<th>t-value</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>+</td>
<td>0.212</td>
<td>3.664</td>
<td>Adoption</td>
</tr>
<tr>
<td>H2</td>
<td>-</td>
<td>-0.459</td>
<td>-4.482</td>
<td>Reject</td>
</tr>
<tr>
<td>H3</td>
<td>-</td>
<td>-0.213</td>
<td>3.664</td>
<td>Adoption</td>
</tr>
<tr>
<td>H4</td>
<td>-</td>
<td>-0.126</td>
<td>1.284</td>
<td>Reject</td>
</tr>
<tr>
<td>H5</td>
<td>-</td>
<td>-0.151</td>
<td>-1.135</td>
<td>Adoption</td>
</tr>
<tr>
<td>H6</td>
<td>+</td>
<td>0.311</td>
<td>2.638</td>
<td>Adoption</td>
</tr>
<tr>
<td>H7</td>
<td>-</td>
<td>-0.085</td>
<td>-1.150</td>
<td>Reject</td>
</tr>
<tr>
<td>H8</td>
<td>-</td>
<td>-0.091</td>
<td>-0.978</td>
<td>Reject</td>
</tr>
<tr>
<td>H9</td>
<td>-</td>
<td>-0.271</td>
<td>-2.638</td>
<td>Adoption</td>
</tr>
<tr>
<td>H10</td>
<td>-</td>
<td>-0.199</td>
<td>-2.985</td>
<td>Adoption</td>
</tr>
<tr>
<td>H11</td>
<td>+</td>
<td>0.351</td>
<td>4.767</td>
<td>Adoption</td>
</tr>
</tbody>
</table>

***: p<0.01, **: p<0.05. χ²=676.535, DF=478, GFI=0.836, AGFI=0.807, NFI=0.845, IFI=0.935, TLI=0.928, CFI=0.934 RMR=0.062
STUDY OF THE EFFECTS OF JOB BURNOUT ON ORGANIZATIONAL EFFECTIVENESS AND TURNOVER INTENTION

Yongman Kwon
Department of Business Administration
Namseoul University
Chenan, Chungcheongnam-do, S.KOREA

ABSTRACT
This study attempted to investigate an efficient human resources management plan by enhancing organizational performance and reducing the factors which influence turnover intention after analyzing the effects of job burnout on organizational effectiveness and turnover intention. To test a research model, a questionnaire survey on job burnout, organizational effectiveness (organizational commitment, job satisfaction) & turnover intention and demographic characteristics was conducted against workers in Seoul and Gyeonggi area. After explaining the purpose of the study to them at work, data were obtained through a standardized questionnaire survey. For this, 232 copies of questionnaires were used for analysis from March 2 to 30, 2014. To figure out the effects of job burnout on organizational effectiveness and turnover intention, causality was analyzed using the Structural Equation Model (SEM). According to hypothesis testing, a total of five hypotheses were chosen, and the following results were obtained: Job burnout had a significant effect on organizational commitment (-0.456), job satisfaction (-0.488) and turnover intention (0.501). In terms of relations between organizational effectiveness and turnover intention, organizational commitment (-0.350) and job satisfaction (-0.199) decreased turnover intention. Considering the reality that Korea is the second highest among OECD member countries with 2,092 hours (2012, yearbook of labor statistics) in terms of average annual work hours, it is urgent to correct the high influence of job burnout on organizational effectiveness and turnover intention. It’s been known that negative attitude against the organization such as cynicism, heavy stress and loss of health have a critical effect on job satisfaction and organizational commitment (Baing, Kelloway & Frone 2005). Therefore, it is important to discover and analyze these obstacles which hinder employees from being absorbed in their work with attachment.

INTRODUCTION:
Recently, rapid changes in business environment have caused a considerable amount of stress to employees. Organizational changes such as introduction of a team organization, in particular, have asked them to carry out their duties through collaboration with people from diverse classes. In fact, this kind of burnout in the process of fulfilling duties decreases organizational performances and makes them leave their job. In fact, employees’ burnout could be a big barrier for the organization to attain the organizational goal as well. Therefore, it is important to enhance organizational performances through job burnout management. Hence, this study aims to i) investigate the effects of employees’ burnout on organizational effectiveness and turnover intention and ii) reduce the burnout and improve organizational performances by figuring out structural relations between organizational effectiveness and turnover intention.

THEORETICAL BACKGROUND AND HYPOTHESIS PRESENTED
Job burnout
Burnout is created by diverse causes, not by just a single cause, as a result of continuous and repetitive emotional tension in maintaining relationship with various people for a long time. After all, it occurs due to accumulation of stress at work such as heavy workload, excessive job demand and uncomfortable work environment. Therefore, burnout is not a temporary phenomenon but a consequence of continuous and ceaseless activities.

Burnout is influenced by social atmosphere and work environment as well as personal circumstances. It is important to prevent burnout and study the influential factors in terms of organizational effectiveness. Maslach & Jackson (1981) conceptualized the components of burnout into emotional exhaustion, low personal accomplishment and depersonalization.

‘Emotional exhaustion’ refers to psychological burnout caused by excessive interactions with people. The symptoms include weariness, lack of energy, fatigue, loss of interest and lack of confidence. ‘Low personal accomplishment’ means a diminished willingness to accomplish or attain a particular goal at work. Lastly, ‘depersonalization’ is a kind of negative response to other people.
Burnout is very contagious. Once it occurs, it spreads across the organization fast. After all, burnout is the result of employees’ emotional exhaustion, depersonalization and low personal accomplishment (Maslach & Jackson, 1985).

**Organizational effectiveness**

Organizational effectiveness is a concept viewed from multiple levels and aspects, which can be divided into economic & financial effectiveness and psychological & behavioral effectiveness. Profitability, growth and total sales are the typical examples of the economic & financial effectiveness indexes while psychological & behavioral effectiveness indexes include employees’ morale, job satisfaction, organizational commitment and turnover intention. Mowday (1979) insisted that organizational effectiveness is employees’ belief to accept organizational goal and values through work achievement, voluntariness and degree of desire to stay as a member of the organization. He said that it is expressed through loyalty and devotion to the organization. In this study, the psychological & behavioral effectiveness indexes ‘job satisfaction’ and ‘organizational commitment’ were examined.

Organizational commitment is a concept relating to mindset or psychological state which refers to employees’ attachments to their organization. In his study (1979), Mowday defined that organizational commitment is employees’ determination to stay with their organization with the willingness to attain the organizational goal. Organizational commitment is a concept which represents relationship between individuals and organization. In general, it refers to employees’ strong belief and attachment to the organizational goal and value, determination to work hard for the organization and willingness to stay as a member of the organization. Therefore, high organizational commitment enhances organizational performances and reduces turnover intention.

Ivancevich & Matteson (1983) said that job satisfaction is a series of attitude that an individual poses against his/her duties. If job satisfaction is a desirable or positive emotional state to employees as the complex of perception, emotion and behavioral tendency, job dissatisfaction could be a discomfort or negative emotional state. If job satisfaction is the result of satisfaction, it could have an effect on the improvement of productivity or organizational commitment. On the contrary, job dissatisfaction may cause undesirable results such as decline in work performances, absence, tardiness and job change.

Employees have a direct effect on corporate performances. From a long-term perspective, they also have a significant influence on corporation-customer relationship. Allen & Meyer (1984) defined turnover intention as a degree of employees’ willingness to leave their current job. In other words, they want to leave the organization voluntarily because of a certain reason. Even though turnover intention is not a complete indicator of leaving a job, it is very important to manage turnover intention because it is used as a sign of quitting a job in many businesses.

**Hypothesis presented**

In Figure 1, the model has been developed and analyzed to figure out the effects of job burnout on organizational effectiveness.

Employees are less satisfied with their job when they feel burnout at work. Daley (1979) said that employees experience work burnout when they are assigned with heavy workload or work that they cannot handle in the relationship between job burnout and turnover intention. Therefore, job burnout and organizational effectiveness are closely related to each other, and the following hypotheses were developed:

Hypothesis 1: Job burnout has a negative effect on organizational commitment.

Hypothesis 2: Job burnout has a negative effect on job satisfaction.

Hypothesis 3: Job Burnout has a positive effect on turnover intention.

As employees have stronger organizational commitment, they are less willing to leave their job. In other words, unless they have an obvious complaint against the organization or decide to leave their job because of personal reasons, they wouldn’t leave the organization. In their previous study (1990), Meyer and Allen explained that there is a negative relationship between organizational commitment and turnover intention.

Job dissatisfaction means low compensation by the organization. According to exchange theory, it means that there is a high possibility that employees may leave their job. Angle and Perry (1981) said that job satisfaction is correlated with turnover intention. Therefore, the following hypotheses were developed:

Hypothesis 4: Organizational commitment has a negative effect on turnover intention.

Hypothesis 5: Job satisfaction has a negative effect on turnover intention.

**EVIDENTIARY ANALYSIS**
Sampling and Research Method

In this study, a questionnaire survey was conducted against workers in Seoul and Gyeonggi area. After explaining the purpose of the study to them at work, data were obtained through a standardized questionnaire survey. A total of 232 copies of questionnaires were used for analysis from March 2 to 30, 2014. To figure out the effects of job burnout on organizational commitment, job satisfaction and turnover intention, causality was analyzed using the SEM.

To test the goodness-of-fit of the research unit and model, the following indexes were used: Chi-squared statistic ($X^2$), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), root mean square residual (RMR), normal-fit index (NFI) and comparative fit index (CFI).

Reliability and Validity Testing

In this study, covariance matrix-based confirmatory factor analysis was conducted using a four factor-included measurement model. According to the analysis, the model fit satisfied the reference level (GFI≥0.9, AGFI≥0.9, NFI≥0.9, IFI≥0.9, CFI≥0.9, RMSR≤0.05) proposed by Hair et al. (2006). In addition, the factor loadings of measurement items on the four factors were all statistically significant (p<0.01). To investigate convergent validity, addition, construct reliability (C.R.) and average variance extracted (AVE) were calculated. In terms of C.R., all factors exceeded the threshold. In terms of AVE, they were all at least 0.5.

Hypothesis Testing

To test the research model, the structural equation modeling was conducted using the AMOS 18.0. The analysis results ($X^2=677.028$, df=478, GFI=0.846, AGFI=0.827, NFI=0.855, IFI=0.935, TLI=0.948, CFI=0.954 and RMR=0.061) mostly satisfied the goodness-of-fit judgment standards proposed by Hair et al. (2006). The structural equation results unveiled in Figure 2.

The said analysis results can be summarized in Table 3.

CONCLUSION

The hypothesis testing results found that job burnout has a considerable effect on organizational commitment (-0.456), job satisfaction (-0.488) and turnover intention (0.501). Considering the reality that Korea is the second highest among OECD member countries with 2,092 hours (2012, yearbook of labor statistics) in terms of average annual work hours, it is urgent to correct the high influence of job burnout on organizational effectiveness and turnover intention. In relations between organizational effectiveness and turnover intention, organizational commitment (-0.350) and job satisfaction (-0.199) decrease turnover intention. It’s been known that negative attitude against the organization such as cynicism, heavy stress and loss of health have a critical effect on job satisfaction and organizational commitment (Baing, Kelloway & Frone 2005). Therefore, it is important to discover and analyze these obstacles which hinder employees from being absorbed in their work with attachment. Because collaboration with people from diverse classes causes stress at work and could eventually generate work burnout (Yang C. H. & Kwon Y. M., 2005). Therefore, there should be efforts to improve coordination and unity and promote communication among groups.

REFERENCES


Mowday, R. T., Steers, R. M. & Porters, L. W., 1979, The Measurement of Organizational Commitment,

**FIGURES AND TABLES**

**Figure 1** Research Model

**Figure 2** Structure Equation Model Analysis Result

**Table 1** Reliability Verification

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sub measurement</th>
<th>Numbers of questions</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Burnout,</td>
<td></td>
<td>3</td>
<td>0.817</td>
</tr>
<tr>
<td>organizational effectiveness</td>
<td></td>
<td>4</td>
<td>0.933</td>
</tr>
<tr>
<td>Job Satisfaction,</td>
<td></td>
<td>5</td>
<td>0.910</td>
</tr>
<tr>
<td>Turnover Intention</td>
<td></td>
<td>5</td>
<td>0.908</td>
</tr>
</tbody>
</table>

**Table 2** Validity Verification

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sub measurement</th>
<th>Numbers of questions</th>
<th>C. R.</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Burnout,</td>
<td></td>
<td>3</td>
<td>0.81</td>
<td>0.66</td>
</tr>
<tr>
<td>Job Satisfaction,</td>
<td></td>
<td>5</td>
<td>0.93</td>
<td>0.66</td>
</tr>
<tr>
<td>organizational effectiveness</td>
<td></td>
<td>4</td>
<td>0.94</td>
<td>0.78</td>
</tr>
</tbody>
</table>

\[X^2 = 693.028, \text{df} = 477, p = 0.000, \text{GFI} = 0.849, \text{AGFI} = 0.822, \text{NFI} = 0.860, \text{IFI} = 0.952, \text{TLI} = 0.946, \text{CFI} = 0.951, \text{RMSR} = 0.051\]

**Table 3** Hypothesis Verification Result

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Hypothesis Direction</th>
<th>Path Coefficient</th>
<th>t</th>
<th>p</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>0.456</td>
<td>3.975</td>
<td>**</td>
<td>Adoption</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>0.488</td>
<td>4.015</td>
<td>**</td>
<td>Adoption</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>0.501</td>
<td>4.125</td>
<td>**</td>
<td>Adoption</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>0.350</td>
<td>5.144</td>
<td>**</td>
<td>Adoption</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>0.199</td>
<td>2.767</td>
<td>**</td>
<td>Adoption</td>
</tr>
</tbody>
</table>

\[\chi^2 = 677.028, \text{df} = 478, \text{GFI} = 0.846, \text{AGFI} = 0.827, \text{NFI} = 0.855, \text{IFI} = 0.935, \text{TLI} = 0.948, \text{CFI} = 0.954, \text{RMSR} = 0.061\]
A MODEL OF KNOWLEDGE SHARING FOSTERING THE LEARNING CAPABILITY – A CASE STUDY WITHIN THE AEROSPACE INDUSTRY

Jarl Simonsen, Aurilla Aurelie Arntzen Bechina*
GKN Aerospace Norway AS,
Kongsberg, Norway
*Buskerud and Vestfold University College,
Faculty of Technology and Maritimes Science
Kongsberg, Norway

ABSTRACT
Engineering firms and more specifically aerospace industry in high cost countries are facing stringent challenges in order to stay competitive. Knowledge in this industry sector is considered as a vital asset for companies facing competition from low cost country. A Knowledge Management (KM) strategy aims at improving the ability of organizations to execute business and manufacturing operations in a more efficient and effective way. Knowledge sharing processes intra or inter companies are essential for fostering a learning organization. Our research study intends to investigate the factors enabling a Norwegian company to become a sharing and learning organization. To this end, qualitative and quantitative assessments of this company have been carried out. Based on the analysis of the collected data, a general framework inspired by the SECI model of Nonaka and Konno has been delineated. This framework indicates the processes to implement in order to foster the learning capability of the company.

INTRODUCTION
The last two decades, engineering companies are facing harsh challenges due to the fast revolution in technology landscape; customer’s needs changes, fierce competitions, globalisation, and so forth. In the aerospace industry, complex requirements of the government and high expectation of stakeholders has created for the firms the need to be more innovative and to react more quickly to the demands of their customers while coping in a sustainable and cost effective way to the environmental issues and constraints. In addition, the product development timeline is more and more reduced in order to be able to compete in global world where some countries still benefit from the lower salaries of their employees. Norway as a higher cost country often has to rely on higher competence, experience and high skills in fewer employees to do the same job. GKN Aerospace Norway AS (GAN) is one example of a company within the manufacturing industry, located in a high cost country and in a competitive market. GAN is a company specializing in the manufacture of high precision aerospace engine components. Therefore, knowledge about technology, technical processes, regulations, marketing is vital for the firm’s survival. Hence, engineering firms need to be able to acquire the appropriate knowledge and to make it available for all the employees. Learning is essential for firms’ survival. More and more companies recognise that corporate knowledge is vital for business and manufacturing processes improvement and even more especially in high cost country such as Norway. Knowledge and Knowledge Management (KM) have gained a tremendous attention. KM processes are ranging from creating, capturing, searching, using and disseminating corporate knowledge, in order to enhance organisational performance (Ndlela and du Toit 2001). Amongst the processes, the knowledge sharing process is the most important for fostering the learning capability of an organisation. Failures or success of such KM projects have been widely discussed in the literature (King and Marks 2008; Vaccaro, Parente et al. 2010; Bechina Arntzen and Sols 2012). However, the reasons of failing KM projects have been poorly addressed. Some studies pointed out that one explanation could be related to the lack of participation in Knowledge sharing process (Hendriks 1999; Bartol and Srivastava 2002; King and Marks 2008).

Thus, the Company GAN, encountering many of the challenges listed above, is recognising that to become a learning organisation, there is a need to facilitate the knowledge sharing intra and inter organisations.

Our research study intends to investigate the factors and the processes enabling the capture and transfer of knowledge in order for a manufacturing company to become a sharing and learning organization. To achieve this goal, a cross approach based on qualitative and quantitative methods was adopted. The validation of our investigation is made through a concrete case study of the...
GKN Company. The Analysis of the collected data, help us to delineate a general framework encompassing the most relevant methods or techniques allowing the knowledge sharing and the organizational learning. This framework uses the model of knowledge conversion and SECI as defined by (Nonaka and Takeuchi 1995; I. Nonaka and Konno 1998).

The next section presents the basic concept of knowledge and Knowledge Management with a special focus on knowledge sharing. The section three, describes the adopted methodology while the section 4 outline the analysis of the collected data and presents the delineated models.

A SHARING AND LEARNING ORGANIZATION

KM concepts

Most of the organisations and especially engineering firms are recognising that knowledge is one of the most competitive resources for the dynamic global business and manufacturing environment (Sharif, Zakaria et al. 2005). To illustrate this recognition, the last decade many companies have strongly focused on organising creating, transferring, searching, and sharing Knowledge under the roof of so-called Knowledge Management (Hildreth and Kimble 2002). Building Competence/skill of employees has proved to be essential to the successful transfer and diffusion of technologies and knowledge. Engineering companies are often engaged in capacity building and knowledge engineering sharing efforts in increase production levels and to perform work routines more efficiently. Studies have focused either on knowledge sharing inter-organizations (Husman & Brandt, 2001) or inter-units in a firm (Joseph G. Davis, Eswaran Subrahmanian et al. 2002). Knowledge sharing contributes to foster the learning capability of the organisations. However, it is acknowledged that knowledge sharing as a nebulous concept is challenging to establish and to promote (Boisot 1998; Dixon 2002). One the reason is due to poor definition of Knowledge. In addition, confusion is often made between knowledge, skills and experiences. Some definitions of the 3 concepts keys are given below.

There is a general consensus on agreeing that there is so far no one single definition of knowledge (Boisot 1998; Dixon 2002). One popular definition is from Davenport & Prusak (2000), they define the knowledge as “a fluid mix of framed experience, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information”. Hunter (1999) focuses more on actionable knowledge and suggests rather “Knowledge is information in the mind, in a context which allows it to be transformed into actions”.

According to Brooking (1999) knowledge is defined “as information in context with understanding to applying that knowledge”. The wide-based knowledge definitions highlight there are several forms of knowledge; tacit, explicit, implicit and systemic knowledge at the individual, group and organizational levels (Johannessen, Olaisen et al. 2001; Ichijo and Nonaka 2007)

Explicit knowledge (EK) has a tangible dimension that can be easily captured, codified and communicated. It can be shared through discussion or by writing it down and stored into some repositories, documents, notes, etc. Examples might include a telephone directory, an instruction manual, or a report of research findings. In contrast Tacit Knowledge (TK) is linked to personal perspectives, intuition, emotions, beliefs, know-how, experiences and values. It is intangible and not easy to articulate, so it tends to be shared between people through discussion, stories and personal interactions.

Knowledge that cannot be written down is defined as Tacit Knowledge while knowledge that can be written down but is not yet put into documents, manuals or databases is defined as implicit knowledge ((Hildreth and Kimble 2002). Implicit knowledge is the subset of tacit knowledge that can be transformed into explicit knowledge. Knowledge, which could be codified if subjected to some type of a mining or translation process, is another view of implicit knowledge.

(Ikujiro Nonaka, Ryoko Toyama et al. 2000) describe that tacit and explicit knowledge are complementary, which means both types of knowledge are essential to knowledge creation. Explicit knowledge without tacit insight quickly loses its meaning. Knowledge is created through interactions between tacit and explicit knowledge and not from either tacit or explicit knowledge alone. In our research study, we focus mainly on the explicit and explicit knowledge.

Experience is often used to define knowledge that is being unconsciously captured and shared. Through routinization of processes and procedures the organizational knowledge is captured unconsciously (Jorge Muniz Jr and Salgado 2013). Experience is something that is built over a number of years in a field of expertise and is accumulated by individuals or groups within an organization. Experience is what enables people to “know what to do” in a given situation. An expert is an individual that embodies a deep knowledge of a subject, and have been tested and trained by experience (Davenport & Prusak, 2000).

Skill definition is defined as the individual’s aptitudes, talents, capacities, capabilities, abilities and personal experiences to use knowledge in an efficient way (Muniz Jr. & Salgado, 2013). Skills allow people to make connections between understanding, knowing and doing. Personal characteristics and psychological traits create the
basis for developing skills to a larger degree than the acquisition of theoretical knowledge. Experience has an important meaning for skill development. (Muniz Jr. & Salgado, 2013)

Throughout the remaining paper, the terms “Tacit” and “Explicit” knowledge will be used to distinguish between two different states of knowledge conversion.

Knowledge Sharing (KS)

Any organization or company that operates within a field of expertise will inevitably accumulate knowledge over the years in some form or another. A senior engineer within the organization may embody knowledge of different types and importance for the organization. Therefore, Knowledge Sharing (KS) is essential for business or manufacturing processes improvement; thus allowing companies to be more competitive.

Davenport (2000) defines knowledge sharing as providing one’s knowledge to others as well receiving knowledge from others. A more pragmatic description of knowledge sharing is “the process through which one unit is affected by the experience of another” (Argote, McEvily et al. 2003).

Knowledge sharing can be as well defined as the process of exchanging knowledge (skills, experience and understanding) among groups of people (Tsui, Chapman, Schnirer, & Steward, 2006).

It can also be defined as something that refers to provision of task information and know-how that can be used in collaboration with others to solve problems, develop new ideas, or to implement policies or procedures. Knowledge sharing can occur in a variety of forms, either through face-to-face interaction with colleagues, peers and other experts, documentation or written correspondence, or by organizing and capturing of knowledge for others (Wang & Noe, 2010). Knowledge sharing and learning mechanisms are highly complex processes to promote in organizations (Allix 2003). Plethora of research studies have investigated the factors facilitating or hampering the knowledge sharing process (ref).

Nonaka & Konno, (1998) introduces the concept of Ba in a sharing environment. Ba is a Japanese concept and can be roughly translated to the English word “place”. Ba can be thought of as a shared space for emerging relationships: it can be physical, virtual, mental, or any combination of these. The authors describe the Ba as a place that differs from the ordinary human interaction because it introduces the concept of knowledge creation.

They further describe the knowledge and its dependence in the Ba, where knowledge that resides in Ba is defined as knowledge, but when the knowledge is separated from the Ba, it turns into information. This information can be communicated independently from the Ba, and it resides in media and networks: the information is tangible. In contrast, knowledge resides in Ba and is intangible.

In essence, the knowledge being tacit or explicit in nature is dependent on the context and situation where it resides. As described in the previous section, the tacit and explicit knowledge in a given situation are both essential for knowledge creation, and the explicit knowledge is codified based on the experiences that reside in the tacit knowledge.

Any organization that successfully shares knowledge both internally and externally can be defined as a learning organization. And therefore it is important to understand the mechanisms of the knowledge sharing that is fostering the learning capability of any organization.

Companies understanding the need to harness knowledge are aware about the decisive issue of creating a work environment that foster knowledge sharing mechanisms and learning capabilities within and across organizations.

Learning Organization (LO)

With the realization of the value of knowledge and learning, organizations have begun looking at how to increase organizational knowledge to gain competitiveness and to become learning organizations (Husted 2002; Gupta 2004). An important ingredient of a rounded knowledge management initiative is the application of the concept of the learning organization. A learning organization is defined as an organization skilled at creating, acquiring and transferring knowledge, and at modifying its behavior to reflect new knowledge and insights (Garvin 1993). Learning can be defined as an accumulation of knowledge, it is very important for organizations to be able to realize what a vital impact an efficient knowledge flow and learning process have on an organization’s innovation and success (Malone 2002).

Garvin, Edmondson, & Gino, (2008) have identified three individual and separate building block defined as follows:

- a supportive learning environment block, describing different characteristics that distinguish an environment that support learning
- A second block encompassing the concrete learning processes, describing the need for structured and concrete processes to handle learning within an organization.
- The third building block, leadership behavior that provides reinforcement, describes the behavior of leaders that reinforces learning.

Garvin, Edmondson, & Gino, (2008) have developed a very interesting questionnaire allowing companies to assess their learning capability in several areas. In addition, this survey instrument proves to be very useful
for companies allowing them to compare themselves against benchmark scores gathered from other firms. Strengths and areas with room for improvement can be identified and corrected if needed. In our research study, in order to evaluate the learning capability of the GAN Company, we have used this survey and the results will be discussed in the following section.

Research methodology

Context of study

GKN Aerospace Norway AS (GAN) is an aerospace engine component manufacturer located in Kongsberg, Norway. The company employs about 480 employees, all located in Norway. The company produces highly technical engine components for major customers in the commercial and military sectors. Today, the company is facing important challenges like for instance producing at a competitive price in a country with high cost such as Norway. Many companies are outsourcing in low cost countries. However, GAN have experience of more than 40 years in producing components with high quality. It is recognized that the company has accumulated long experience, know-how capabilities and top-notch competence in the field of expertise. Thus despite the high cost of Norway; GAN is able to compete with other countries. However, the experience and competence built over the years is embodied in the people that are employed in the company, and is subject to a certain risk of competence loss through retirement, career changes and other influences. In order to preserve their knowledge as one of the most important corporate assets, the company has started few years ago to investigate how to keep their crucial knowledge and to overcome the listed challenges above. Our research study focuses on the identification of factors or processes allowing the company to stay on the competitive edge. To this end, we intend to understand how the company is capturing and reusing the knowledge. In order to achieve our research study we have conducted both qualitative and quantitative methods described in the next sections.

Data collection

In order to optimize predictive effectiveness of the data collected, we used a mixed-mode approach for obtaining insights from the GAN Company. Methods to collect qualitative data include interviews, informal discussion, literature reviews and researcher's participations. In this study, one of the researchers used his own experience while working in the company. The observation of the working routines and more specifically the knowledge sharing process has contributed to enrich the descriptive aspect of the collected data. Most of the interviews were based on unstructured and face to face interactions with open questions, giving the employees full latitude to discuss issues occurring while working.

Literature reviews helped us to refine the questions asked during the interviews related to the learning capability and knowledge reuse.

In order to assess if GAN is a learning organization we have used the survey instrument developed by Garvin, Edmondson, & Gino, (2008). The survey is available online at los.hbs.edu and can be used by any company wishing to know how well the company functions as a learning organization. The survey was distributed to several employees with different job positions and range of experience. Out of the 40 received questionnaires, only 24 were complete and exploitable.

Finding and data analysis

The result of the survey is depicted in figure 1 and figure 2. Strengths and areas with room for improvement within the organization have been identified. The block related to Supportive Learning Environment shows that 4 out of 5 characteristics are placed in the top quartile. For example, the interpretation of the characteristic Psychological Safety, with a score between 87 and 100, indicates that the employees of GAN feel safe to express their ideas in a working environment that they would consider safe and not threatening their own opinions. The second characteristic, Appreciation of Differences shows, that workers acknowledge and respect different opinions in discussions and accept to consider alternative ways to work. The characteristic Time for Reflection indicates that employees perceive that they have enough time to reflect on their work and work routines despite potential workload, while the one related to Openness to New Ideas ranks in the second quartile. This can be interpreted that employees might feel reluctant to new ideas and solutions. However, all in all GAN is providing a Learning environment composite that is confirmed by high score.

With these results in mind, remaining results from the survey were analyzed. The two results that stand out as particularly low ranked are Information Collection and Information Transfer. These two results are the only ones from the survey placed in the bottom quartile, which indicate that information might not be adequately collected/captured and transferred within GAN.

The results are in line with what has been observed by one researcher during the daily work. In addition, during the interviews, some participants agreed that although the company was keen to establish mechanisms facilitating the knowledge or information capture and sharing, there was still the need to improve these activities.

Results of the survey as depicted in figure 2 shows where GAN stand compared to other companies that have taken
defined by (Garvin, Edmondson, & Gino, 2008). The exchange of experience and knowledge during the daily reflection but more important, that although the extensive procedures or incidents at the company, and in many ways is similar to the US-Army’s After Action Review (AAR). The CFI yields some formalized documentation concerning the specific problem or incident and should by definition create some formalized root cause analysis to handle the situation; this however, is not always adequately performed at the CFI. In essence, it is argued that the valuable knowledge gathered during face-to-face interactions are not adequately defined and documented in the informal and formal meeting structure to ensure future reference to enable reuse of knowledge for the company. Therefore, there is a need to suggest appropriate methods or models to capture the knowledge and to codify it for sharing and reuse purpose. This is how GAN can become a Learning organization based on the principles defined by (Garvin, Edmondson, & Gino, 2008).

**DISCUSSION**

The quantitative data collected through the survey provide a general understanding of GAN’s perspectives related to the learning capability. The analysis of the quantitative and qualitative data shows that there is a convergence of the identified issues like for instance that the capture and transfer of the organizational knowledge were not sufficiently addressed.

**Building a house of Sharing and Learning**

The results of data collected allow us to delineate a model based on LEAN house concept that we call the house of Sharing and Learning Organization (SLO) (Vision Lean, 2008). The figure 3 illustrates this model and describes the different elements composing the house.

The house is built from the ground up where the foundation represents the pre-requisites for any organization. The “Body” of the building represents the means and ingredients of the organization and the roof represents the goal or the vision of the organization (Vision Lean, 2008). With the SLO - model it is possible to visualize a storyline, with a bottom up approach “People, with different Culture and Attitude, can through Methods and Processes obtain Experience, Knowledge and Skills, and with the support of Technological Framework help create and maintain a Sharing and Learning Organization.” One important pillar is Knowledge that needs to be captured and transferred or shared. At the company, it was important to be able to capture whenever it is possible the Tacit knowledge and to provide means to convert this knowledge into explicit knowledge in order to be reused. The next section describes how the SECI model can contribute to the conversion of tacit into explicit knowledge.

**A Modified SECI – Model**

Based on the literary review, the SECI – Model (SECI) as defined by (Nonaka & Konno, 1998) is considered as the most promising approach. The SECI is depicted in Figure 4 and shows a spiraling process illustrating steps that knowledge goes through by being converted from tacit knowledge to explicit knowledge, and back again to tacit knowledge.

Furthermore, the SECI introduces further concepts when considering its usage in practice, such as the barriers between the quadrants in the model, the demand for consistent and clear inputs and outputs, and the need to fully complete the “rotation” of the SECI for complete knowledge conversion. In order to identify how GAN is processing the knowledge conversion, we have used this model and try to link it with the current work practices at the company. The first interpretation indicates that *socialization*, as defined in the first quadrant is probably the most relevant since employees usually have time and ressourcers to discuss issues and solutions whenever needed. A concrete example would be that product responsible engineer usually interact face to face with an operator of a machine in the manufacturing process. Knowledge is shared on a very regular basis. However, the externalisation as described in the second quadrant is not a current practice. In testimony, the comments from people mentioning that the results of the shared experiences or knowledge during face to face interactions are not systematically translated into formal procedure or documentation. One of the reasons was related that some employees were not fully aware of the need of structured for formalized structures or documentation in order to convert TK to EK. Although the CFI process yields a certain amount of documentation about a problem or issue, but in a majority of cases, this documentation lacks a clear structure and format that enable reuse and sharing of crucial knowledge within the company. The output from a CFI is often not sufficient to serve as an input to the third quadrant of the SECI that seeks to handle
The third quadrant of the SECI illustrates the place where documented knowledge or information is combined in order to generate again more documented knowledge with a structured format. Output of this quadrant is well codified knowledge and stored in different knowledge based systems. Hence the reuse by others employees in facilitated. Today the company uses different systems ranging from a simple excel document to supplicated product data management systems or network systems describing the operational processes. Today, it is agreed that the set of the currents systems used by worked are not sufficiently aligned with the business or manufacturing process. A layer facilitating the knowledge sharing should be further developed.

Socialization at GAN

This quadrant as described in figure 5, indicate that knowledge and experiences are shared though face to face interaction between participants. In order to support the shared knowledge that is embodied in employees, we suggest several routines or processes such as internal, formal or informal discussions, face to face interaction, mentoring, morning meetings. What is the most important is to allow people to interact by giving them not only time but as well structure facilitating the exchanges in the office setting; coffee break room might be a good option. In shop floor, we can designate some spaces only for this kind of interaction like for instance when the dayshift and the nightshift changes, the operator can describe any encountered issues to the employee taking over the machine. Today, at GAN it is done naturally and everyone recognise the value of this shared experience.

Externalization at GAN

The shared knowledge as output will serve as input in the next quadrant, where Tacit Knowledge (TK) is converted into explicit Knowledge (EK). This process of externalisation requires a set of specified and formalized procedures. Today, due to the lack of clear standard in capturing the knowledge generated by the sharing experience, employees need more time to search the relevant information in order to perform their work routines. The figure 6 illustrates the second quadrant of the SECI. A set of procedures allowing the knowledge capture and externalisation into some formalised output can include for instance; collaborative tools, formalised forums, morning meetings, root cause analysis, and might therefore generate Best practices or lessons learned documents with a specific format allowing the easily reuse of the generated formalized documents.

Combination at GAN

The third quadrant of the SECI illustrates the place where documented knowledge or information is combined in order to generate again more documented knowledge with a structured format. Output of this quadrant is well codified knowledge and stored in different knowledge based systems. Hence the reuse by others employees in facilitated. Today the company uses different systems ranging from a simple excel document to supplicated product data management systems or network systems describing the operational processes. Today, it is agreed that the set of the currents systems used by worked are not sufficiently aligned with the business or manufacturing process. A layer facilitating the knowledge sharing should be further developed.

Internalization at GAN

The fourth quadrant described by (Nonaka & Konno, 1998) as the “Exercising Ba” where an individual or group seeks knowledge and uses it in their daily operations. The knowledge is transferred by such activities as training, exercising or courses, etc. In this quadrant the knowledge completes the full circle and is converted from EK to TK. This internalisation process enables the organizational learning. Therefore factors facilitating or impeding the learning capability should be further understood.

Upon completion of this quadrant, the goal of the SECI is to enable a new rotation of the knowledge, where knowledge triggers a new experience, perhaps in a different setting and context, and restart the SECI rotation.

Framework supporting the learning capability

In order to trigger the spiral of knowledge conversion from TK to TK, we suggest that the technological framework as developed in (Arntzen, Ribiere, 2009) could be used and adapted to the GAN requirements. This model is depicted in figure 9.

CONCLUSION

The paper discusses how an engineering companies could be become a learning organisation by looking closely on how to capture and share knowledge generated during the work routines. The model of SECI has been primary used in order to understand the knowledge conversion from the knowledge embodied in the employees into some formalised procedures, thus allowing the reuse of the knowledge. This SECI model has been extended by adding some further perspectives such as set of processes supporting the transfer from different state of the knowledge.

We suggest that tools supporting the whole conversion should be further analysed and aligned with the manufacturing and productions process.

A Sharing and Learning Organisation model have be delineated and shows that the foundation rely mainly on
the people attitude and culture of the organisation. The pillars such experience, Knowledge and skills based on people and on right set of methods and process, need a technological framework layers in order to foster the sharing and learning capability of an organization.

ACKNOWLEDGMENT

We would like to thanks all the participants of our investigations and a special thanks to Gerrit Muller for his valuable input.

REFERENCES


" Long Range Planning 33(1): 5-34.


Jorge Muniz Jr and A. M. P. Salgado (2013). KNOWLEDGE, EXPERIENCE AND SKILL: AN EXPLORATORY STUDY IN A BRAZILIAN GLASS MACHINE TOOLING PLANT. Organizational Learning, Knowledge and Capabilities conference.

Joseph G. Davis, Eswaran Subrahmanian, et al. (2002). "Knowledge Sharing and Management in Large, Multi-National Corporations.".


Fig. 1 - Result of the LO survey

Fig. 2 - Benchmark of the GAN

Fig. 3 The House of a Sharing and Learning Organization (SLO)

Fig. 4 SECI – Model (Nonaka & Konno, 1998)

Fig. 5 Modified SECI - Model, Socialization Quadrant

Fig. 6 Modified SECI - Model, Externalization Quadrant
Fig. 7 Modified SECI - Model, Combination Quadrant

Fig. 8 Modified SECI - Model, Internalization Quadrant

Fig. 9 A Knowledge Type Driven Framework of KM technologies
TOWARDS AN UNOBSTRUSIVE COTS EEG-BASED ASSISTIVE TECHNOLOGY FOR PATIENTS WITH MOVEMENT DISABILITIES: A REVIEW

Sian Lun Lau, Afzal Ahmed, Ruslan Zhunussov
Department of Computer Science and Networked Systems, Faculty of Science and Technology
Sunway University
Bandar Sunway, 47500 Malaysia

ABSTRACT

People with movement disabilities will not be able to move around as convenient and easy as other people. One of the solutions that can help them is assistive technology based on electroencephalography (EEG) signals. Brain computer interface (BCI) systems record EEG signals and perform interpretations that can capture the “thoughts” of their users. Two possible application domains are mobility and emotion detection. Besides this, it is also essential to develop these assistive technology to be unobtrusive and intuitive to use. Therefore, we envision the use of commercial-off-the-shelf (COTS) EEG devices as a feasible and affordable solution. In this paper, recent work, which utilises COTS EEG-based devices to provide solutions for the two selected application domains, is presented and discussed.

INTRODUCTION

Motivation for the proposal

For a person with disabilities, particularly those who are paralyzed or cannot move independently, it is a challenge to carry out everyday activities that are seemingly simple for normal healthy people. For example, it is not possible for them to drive or move independently. They will face difficulties in giving input to devices, such as a computer or audio/visual devices. It is also a challenge to them to express their feeling or emotion to other people, when speech and expressions are obviously not as easy as people without similar disabilities.

Traditionally, among techniques applied and investigated in the past decades, newer approaches aim to enable hands-free control using various human-machine interfaces (HMI). Examples of investigated HMI are such as techniques using muscle signals (Electromyography - EMG) (Felzer & Freisleben 2002), eye movements (Electrooculography - EOG) (Barea et al. 2002) or video-based eye gaze tracking (Murata 2006), human body movements (limited limb movements, head gestures, facial expression etc.) (Ju et al. 2008), voice command (Harada et al. 2008) and brain signals (Electroencephalography - EEG) (Iturrate et al. 2009).

To provide a feasible and affordable solution to the targeted users, unobtrusive techniques based on commercial-off-the-shelf (COTS) devices can be an attractive option. The availability and cost of these products will be an advantage as compared to specialized, high-end and/or medical-grade devices. They are usually simpler and more intuitive than the high-end devices in terms of daily usage. Common shortcomings are the trade-off of accuracy and sometimes efficiency. However, there are situations where a fair trade-off is expected but patients may still accept the unobtrusive COTS devices, as long as the errors and loss of efficiency are not life threatening. For example, for self-initiated vital information monitoring and measurements at home, away from health care institutions and professionals, may be worth the above mentioned trade-off, because patients do not need to travel to their assigned health care institutions to perform periodical check-up (Lau et al. 2010). Patients may opt for in-home recuperation, as long as the monitoring procedures and reporting can be performed by medical experts with acceptable accuracy and efficiency.

One specific technology listed above is the usage of EEG devices to allow “mind-control”-based approaches to help patients to express themselves, control and move around. Particularly for patients with amyotrophic lateral sclerosis (ALS), cerebral palsy or spinal cord injury, this technology enables them to capture their “thoughts” through the translation of their measured brain signals into usable commands or machine-comprehendible thoughts.

This paper intends to present a survey of related work that applies EEG-devices for patients with movement disabilities. Focus will be given to the COTS devices and the techniques used for the respective solutions.

ASSISTIVE TECHNOLOGY FOR PATIENTS WITH MOVEMENT DISABILITIES

Electroencephalography - EEG
The existence of human EEG signal was first recorded by German neurologist named Hans Berger (1873-1941) (Berger 1931). He was the first person who proved that an electrical signal can be recorded from the human scalp without opening the skull. This discovery enabled the investigation and usage of EEG signals for different purposes.

Early multi-electrode EEG system was mostly used in clinics and laboratories to diagnose epilepsy and monitor coma patients. Continuing research and development on EEG systems over the past decades has brought the opportunity to take this technology from the clinics and laboratories into informal environment such as homes and schools. Common approach to perform EEG-based research and development investigations is to apply an EEG cap that wraps the skull with electrodes at fixed positions. Electrodes are commonly connected to the measurement system via cables, hence making the EEG system non-flexible and non-practical for all day usages, especially when it comes to different factors such as mobility, outdoor and convenience (Stopczynski et al. 2014), (De Vos et al. 2014). For laboratory applications, such set up and approaches are still acceptable, for one is more interested in high quality signal and analysis of the intended EEG signals.

Today, EEG systems offer multiple applications across a variety of fields from medical to non-medical. Some of them have even become popular in the consumer market. Recently, companies, such as Emotiv Systems Incorporation (Emotiv 2014a) have presented new low-cost EEG-based Brain-Computer Interface (BCI) portable devices (e.g., see Figure 1 and Figure 2) as well as software development kits to support the expansion of tools and games available from their application markets. Originally implemented as a cognitive game interface, they have been applied as brain computer interfaces to directly manipulate robotic arms or to drive a car.

Using EEG as a technique to enable communication or interaction between human and machines/computers is an active research area. The BCI technology enables people to interact with the device by analysing and processing brain wave signals. In the research, number of publications have increased over the years (Jonathan Wolpaw 2012) due to the advances in EEG signal processing as well as the addition of newer BCI devices. Particularly in the area of assistive technology, BCI is an attractive approach. Users can communicate with their devices by “thinking”, making controls and usage of technology more intuitive and less obtrusive.

In the following subsections, related and recent work using BCI for two categories of applications will be presented. Focus will be given to the interaction with BCI as sole technology, or as additional technology apart from another technology. Another focus is given to understand users’ emotion using EEG signals via BCI technology.

Majority of the selected investigations will have one thing in common – the BCI technology used are portable and mobile. We reviewed some recent work that focuses on accessing EEG signal in an unobtrusive manner.

To enable and to apply EEG devices as an unobtrusive assistive technology, four requirements need to be fulfilled. Firstly, the EEG device should be mobile and wireless. In other words, users are not restricted to a fixed area to perform their activities if they wish to use EEG. Secondly, the preparation and training of the system should take place outside of the laboratory environment. Traditional studies were usually performed in well-controlled laboratory environments, and this can produce solutions that will not work as well as anticipated in the real-world scenario (Wong et al. 2014). By having the approaches tested alongside the suitable devices in less or non-controlled environments, the chances, where the solution being applicable in everyday activities, are higher. Thirdly, techniques identified and investigation should put users’ needs first. Trade-offs between accuracy and recognition speed will be expected. Last but not least, the solutions, including the hardware as well as the software, should also be intuitive and unobtrusive in handling. From the preparation to usage, there is no need for expert users to assist or intercept. The users themselves should be provided with the appropriate tools to perform these tasks independently.

Idealistically speaking, a good approach should have first two factors considered while balancing the right trade-offs to provide the best usage scenario for its users. Design of the tools and applications should consider the usability of all factors, so that the developed solution can be operated by the users themselves.

**AVAILABLE COTS EEG DEVICES**

Currently there are a few options for COTS EEG devices that are available in the market. Often applied and used in different investigations are the Emotiv EPOC and EEG device (Emotiv 2014a). The Emotiv EPOC is one of the first commercially available BCI system offered to the general public. It is priced from $299. Originally designed as a gaming peripheral, it is often also used for research investigations. The Emotiv EEG costs $750, and records raw EEG signal (the EPOC does not record raw EEG signal). Both devices have 16 wet sensors, where 14 are the EEG sensors/electrodes. Connectivity is provided wirelessly to a computer via the provided USB dongle. Gyroscope sensor is also built on the devices, so that head movements can also be detected. The provided SDK can be used to create new applications that are controlled or influenced by brain wave and its changes.

Taylor and Schmidt performed evaluation on the Emotiv EPOC to investigate its ability to detect and classify 6 different mental actions accurately (Taylor & Schmidt 2012). The authors concluded that the EPOC is a
capable device and can be attractive as a low-cost BCI system. Other people who have performed different evaluations using Emotiv devices are (Carrino et al. 2012; Vokorokos et al. 2012; Navuluri et al. 2011).

Based on the outcome and investigations it has achieved, the company has planned the Emotiv Insight to be offered to the consumer by end of 2014 (Emotiv 2014b). The Emotiv Insight offers a more modern and sleek design. It records 5 channels of brain waves and translate them into information that can be understood by systems and users. All 5 sensors are dry sensors. It uses Bluetooth to communicate with other devices. The device also includes motion sensors - accelerometer and gyroscope. Again, SDK will be provided. We expect both developers and researchers will be adapting this product for future research work.

Neurosky devices are also often seen in research investigations (Neurosky 2014). Due to its simplicity in design and features, it is not as popular as the Emotiv devices. The research investigations performed using Neurosky headsets include (Wong et al. 2014; Mostow et al. 2011; Chen et al. 2012). The Neurosky Mindwave consists of a single channel for the measurement of raw EEG signal. Transmission of signal is performed via radio frequency through a dongle. Another variation of the Mindwave is the Neurosky Mindwave Mobile. It communicates with a device, such as a mobile phone as well as a desktop computer, via Bluetooth 2.1. The Neurosky devices are positioned in the market as an introductory BCI device. It is less expensive when compared to the Emotiv devices. Nevertheless, recognition supported by the Neurosky devices includes eye blinks, attention and meditation.

Another upcoming product that has been proposed in Kickstarter in 2013 is the Melon headband (Melon 2014). It has three electrodes (similar to Neurosky Mindwave) and comes with an SDK for iOS and Android. It allows connection via Bluetooth to a smartphone. Expected applications include focus, attention, and emotion recognition.

Muse is another product that utilizes 7 sensors that are capable of reading 4 channels of data (Muse 2014). It uses Bluetooth to communicate with other devices. The device carries also a 3-axis accelerometer, which detection of motion of the user’s head for gaming purposes. A Muse headband costs currently $299.

One more product that normal users can buy off the shelf is the iFocusBand (IFocusBand 2014). It works as a sport headband, and brain waves are measured using 3 sensors. The iFocusBand works in standalone mode, where it does not require another device for data interpretation. One can connect the headband to a mobile device or to upload data through the mobile device to the cloud. This enables social sharing of brain wave based information between friends.

A summary of the above mentioned devices is found in Table 1.

**CURRENT SOLUTIONS FOR PATIENTS WITH MOVEMENT DISABILITIES USING COTS EEG DEVICES**

As mentioned earlier, patients with movement disabilities such as ALS or spinal cord injury will require most likely hands free control if they wish to move around or to interact with a computer/device (Barea et al. 2002; Cipresso et al. 2011). Current solutions include the usage of EMG and tongue movement to control the wheelchair (Felzer & Freisleben 2002; Iturrate et al. 2009), (Rechý-Ramírez et al. 2012). Another application is to enable input for computer usage by spelling. Commonly this is achieved by using eye or gaze tracking and many approaches rely on video-based techniques (Ju et al. 2008), (Nguyen & Jo 2012).

BCI approaches are comparative more intuitive, as it requires only input obtained through brain wave signals. The emergence of mobile and portable devices has made it easier to fulfil the first requirement. However, majority of the more mobile solutions tend to be more specific in its applications and capabilities. Hence, the number of channels/electrodes is relatively smaller as compared to the laboratory grade devices.

Since the past years, there are several investigations looking into using the Emotiv EPOC/EEG neuro headset (Emotiv 2014a), sometimes coupled with a portable device (e.g. tablet, smartphone), as a potential solution for patients with movement disabilities to be mobile again. Carrino et al. have proposed to use an Emotiv EPOC to help patients to control an electric wheel-chair in a self-paced manner (Carrino et al. 2012). The investigation tested the EPOC using motor-imagery technique to perform wheel-chair control. The best results obtain was not higher than 60%. Considered that the use of professional (or laboratory) grade device gave results up to 80%, the results from the use of the Emotiv EPOC was as promising as expected. However, the authors mentioned an increase in the number of subjects may lead to better recognition result.

The authors Vourvopoulos and Liarokapis have investigated both Emotiv EPOC and Neurosky devices (Mindset and Mindwave) to evaluate their suitability in navigation (Vourvopoulos & Liarokapis 2014). Instead of navigating a wheelchair, the users have been requested to control a robot in a maze using these COTS EEG devices. At the end of the experiments, users are given a questionnaire to provide their feedback. From this investigation, it is concluded that most users find the EEG devices acceptable respectively. As for Neurosky devices, users have more problem controlling the robots. However, the responsiveness of the system built based on Neurosky was good. Users also reported about the trouble of the
need to have full concentration. The users gave slightly better feedback on the Emotiv EPOC, though the mastery of control took some time (up to 10 minutes) to be better achieved. Latency between the control (with mind) and action is also another issue raised by the users.

Kaysa and Suprijanto have performed similar studies in integrating the Emotiv EPOC on a BCI platform to control an electrical wheelchair (Kaysa & Widyotriatmo 2013). In their experiments, EEG signals have been classified to control movement of the wheelchair. The outcome has shown that the accuracy of the system still has room for improvements. The shortcoming of their approach is the limit of the simultaneous process. Currently, only one process is allowed.

The Emotiv system also comes with other sensors such as a gyroscope sensor. Rechy-Ramirez et al. used the Emotiv EPOC to control a wheelchair but chose not to use the EEG signals (Rechy-Ramirez et al. 2012). Instead, the investigation only considered the gyroscope values. The assumption is that, if the user can still perform simple head movements, the device can be used to help the control of a wheelchair. The outcome was acceptable.

There are other investigations that looked into the features of an Emotiv device. Vokorokos et al. investigated recognition of facial expressions in EEG signals using an Emotiv EPOC (Vokorokos et al. 2012). They reported intentions of the subjects to push, pull, move left, right, up and down, as well as to turn around were detectable. Navuluri et al. looked into prediction of drivers’ intentions while driving using the Emotiv EEG device (Navuluri et al. 2011). Astaras et al. used an Emotiv EPOC to control a 6-degree-of-freedom robotic arm (Astaras et al. 2013). The preliminary tests have verified that simple pick and place tasks can be performed by an operator after a relatively short learning period.

**Current Solutions for emotion recognition using COTS EEG devices**

Another application based on EEG devices is emotion detection. With the use of such modern EEG devices and the software tools, it is now possible to study mental processes associated with human behaviour, such as user emotions.

Sourina’s team have developed EEG-based emotion recognition algorithms that allow recognition up to eight emotional states (satisfied, happy, surprised, protected, sad, unconcerned, angry, and fear) (Liu et al. 2011; Hou & Sourina 2013). By using an Emotiv EPOC device, the researchers were able to detect selected emotions (six emotions: fear, frustrated, sad, happy, pleasant and satisfied) with fewer electrodes as compared to related work (Liu et al. 2011). Two prototypes were built, which are an EEG-based music therapy and an EEG-based music player. Further work focused on the combination of EEG with haptics technology (Hou & Sourina 2013). The authors proposed that serious games can be supported by EEG and haptics technology. Patients may use these games for post stroke rehabilitation.

Another area similar to the emotion is the recognition of attention. Li et al. have designed a real-time EEG-based system that can recognize attention level in a ubiquitous environment (Li et al. 2011). The system combines an EEG device that measures only from one electrode (the Fpz electrode) and a smartphone. The k nearest neighbour (KNN) classifier has been applied to investigate its suitability for attention recognition. The outcome has shown that a recognition accuracy close to 70% has been achieved.

Berka et al. performed an investigation on real time analysis of EEG indexes of alertness, cognition and memory acquired (Berka et al. 2004). The EEG device used is produced by B-Alert and has the capability to acquire six channels of EEG signals. Three experimental tasks were evaluated – the warship commander, the three level cognitive as well as the image learning and recognition memory. The outcome has been positive. The B-Alert EEG can detect specific changes in the EEG signals and can associate them to the levels of cognitive workload.

Fattouh et al. applied emotion detection using an Emotiv EPOC to control the wheelchair (Fattouh et al. 2013). The Emotiv API was used to detect frustration of the user. The intention is to stop the wheelchair from moving as soon as the system detects its user’s frustration. The authors observed that user concentration time is shorter if his emotional state is considered.

Mostow et al. has tested a Neurosky Mindset to investigate whether a low-cost EEG device can enable to detection of mental state in recognizing reading difficulties (Mostow et al. 2011). While subjects (adults and children) are reading sentences, the system aims to detect whether a read sentence is perceived as easy or difficult. Best accuracy achieved were up to 62%. Such solution can be deployed in schools to monitor reading and learning progress as well as to improve teaching efficiency.

**Other potential applications/usages of EEG devices**

There are also some other applications using a mobile and portable EEG device. For example, Askamp and van Putten looked into possibilities to detect epilepsy symptoms by using mobile EEG approach (Askamp & van Putten 2014). Li et al. developed a wireless EEG monitor system by building it into a body sensor network (BSN) node (Li et al. 2010). Chen et al. performed experiments to investigate how EEG can help to improve automatic speech recognition (ASR) accuracy (Chen et al. 2012). Miranda et al. created an assistive prototype that makes music based on measured EEG signals (Miranda et al. 2011). Hu et al. proposed to use the unique EEG profile of an individual as a mean to provide EEG-based
Biometric identification and validation of individuals (Hu et al. 2011).

From the list of various application of mobile and portable EEG devices, it can be seen as a trend where more exciting and interesting EEG-based solutions will be investigated, tested and proposed in the near future. Particularly when newer, cheaper and more specific devices are made available to the consumer, systems and applications will also continue to grow.

**TOWARDS AN UNOBSHRUSSIVE COTS EEG-BASED ASSISTIVE TECHNOLOGY FOR PATIENTS WITH MOVEMENT DISABILITIES**

In recent years, there is an increase in available COTS EEG devices being offered to common consumers. The research results of EEG investigations in the past decades have enabled the development of these devices. The first batch of devices, released into the market focus on simple game control and recognition of mental states, such as attention and emotion. The recent devices, such as Melon and Muse, utilize minimal number of sensors to be more portable and unobtrusive. This characteristic will allow day to day use without feeling unnatural or attracting attention from other people.

From the reviewed devices, the Emotiv EPOC and EEG devices are suitable for more exploratory investigations. One can explore the combination of channels that may be used for newer areas of application. This include investigations of specific channel(s) and newer algorithms suitable to interpret desired brain signals patterns. Once identified, one can then move towards a simpler device, similar to the devices like Melon and Muse, for more dedicated purposes as assistive technology.

The simpler types of COTS EEG devices, such as Mindwave, Melon, Muse and iFocusBand are nevertheless still useful. One can investigate and develop applications that adapt or react based on the recognition performed by these devices. New possibility can still be explored to extend the current recognition capabilities. This will also enable the addition of newer applications for the patients with disabilities.

Based on the reviewed recent work, we expect more portable and mobile COTS EEG devices to be made available in the near future. Hence, more applications based on these devices are also expected to be offered. Costs for these devices will continue to be relatively affordable. Looking at the few devices being offered currently from $79 to $230, the future COTS EEG devices should be priced below $150. Compared to medical grade devices, these devices will be considered as affordable.

With mobile devices being more ubiquitous, connectivity with mobile devices will be common. This can be seen with some of the devices supporting Bluetooth 4.0 LE (refer to Table 1). This is a strategic trend, since modern mobile devices have relatively fast computing power, and EEG pattern recognition can be performed directly on these devices.

Ultimately, the logical move is to embed the EEG sensors into headwear. A good example is shown in iFocusBand, where athletes wear it to evaluate their performance results. Being wearable and seamless, the patients can use these EEG-enabled devices to assist them in communication and control. This will also encourage researchers to discover newer usage possibilities other than the current capabilities.

**CONCLUSION**

In this paper, we have presented some of the recent investigations that utilize EEG-based techniques to provide solutions for patients who have movement disabilities. Particularly if we want to develop better assistive technology for the intended group of users, we need to identify approaches that are unobtrusive and low-cost. The emergence of COTS EEG-based devices has become an opportunity as well as an attractive technology to be considered for the intended application domain. From the reviewed work, we can see devices that cost below $300 have been used to help patients with movement disabilities to move or to express their emotions. Looking ahead, we anticipate that more devices will be out in the market in the near future. The upcoming EEG devices will no longer cost too much, and by applying the right techniques as well as algorithms, newer and innovative solutions can be investigated and developed as assistive technology.

**REFERENCES**


Carrino, F. et al., 2012. A self-paced BCI system to control an electric wheelchair: Evaluation of a


**FIGURES AND TABLES**

**Fig. 1** The Emotiv EPOC system (Emotiv 2014a)

**Fig. 2** The Neurosky Mindwave (Neurosky 2014)

<table>
<thead>
<tr>
<th>Device, URL</th>
<th>Sensors</th>
<th>Communication</th>
<th>Battery run time</th>
<th>Price (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotiv EPOC, <a href="http://emotiv.com/epoc/">http://emotiv.com/epoc/</a></td>
<td>14 EEG sensors: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, Gyroscope</td>
<td>Proprietary wireless, 2.4GHz band</td>
<td>Est. 12 hours</td>
<td>299</td>
</tr>
<tr>
<td>Emotiv EEG, <a href="http://emotiv.com/eg/">http://emotiv.com/eg/</a></td>
<td>14 EEG sensors: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, Gyroscope</td>
<td>Proprietary wireless, 2.4GHz band</td>
<td>Est. 12 hours</td>
<td>750</td>
</tr>
<tr>
<td>Emotiv Insight, <a href="http://emotiv.co/">http://emotiv.co/</a></td>
<td>5 channels: AF3, AF4, T7, T8, Pz, 6-axis inertial sensor</td>
<td>Bluetooth 4.0 LE</td>
<td>4 hours minimum</td>
<td>229</td>
</tr>
<tr>
<td>iFocusBand, <a href="http://www.ifocusband.com/">http://www.ifocusband.com/</a></td>
<td>3 electrodes</td>
<td>Bluetooth</td>
<td>6 hours</td>
<td>495</td>
</tr>
<tr>
<td>Melon, <a href="http://www.thinkmelon.com">http://www.thinkmelon.com</a></td>
<td>3 electrodes, including primary electrode FP1</td>
<td>Bluetooth 4.0 LE</td>
<td>Approximately 8 hours</td>
<td>149</td>
</tr>
<tr>
<td>Muse, <a href="http://www.choosemuse.com">http://www.choosemuse.com</a></td>
<td>7 EEG sensors: AF3, AF4, TP9, TP10 and 3 electrode electrical reference CMS/DRL, 3-axis accelerometer</td>
<td>Bluetooth 2.1 + Enhanced Data Rate (EDR)</td>
<td>Up to 5 hours</td>
<td>299</td>
</tr>
<tr>
<td>Neurosky Mindwave Mobile, <a href="http://store.neurosky.com/products/mindwave-mobile">http://store.neurosky.com/products/mindwave-mobile</a></td>
<td>1 channel: FP1</td>
<td>Bluetooth 2.1</td>
<td>8 hours runtime</td>
<td>129</td>
</tr>
</tbody>
</table>
STUDY ON POTENTIAL USE OF GEOPOLYMER USING SARAWAK FLY ASH

Hsiao-Yun Leong¹, Ek-Leong Dominic Ong¹, Jay Sanjayan², Brian Lynch²
¹Swinburne Sarawak Research Centre for Sustainable Technologies,
Swinburne University of Technology Sarawak Campus, 93350, Kuching,
Sarawak, Malaysia
²Centre of Sustainable Infrastructure,
Swinburne University of Technology Melbourne Campus,
Hawthorn VIC 3122, Australia

ABSTRACT
This paper reports the characterization of Sarawak Fly Ash. The main focus is to identify the mechanical properties of geopolymer using Sarawak Fly Ash. The studies show that Sarawak Fly Ash can be an attractive construction material with respect to the high compressive strength obtained from Unconfined Compression Test. A series of tests are performed in this research such as scanning electron microscope, compressive strength test, x-ray fluorescence and particle size distribution. The results obtained show that SFA is predominantly smaller than 40μm, the compressive strength of sample curing for 1 day and 7 days are 42MPa and 55MPa respectively. The morphology of the Sarawak Fly Ash has also been studied. The strength development in geopolymer using Sarawak Fly Ash was mainly influenced by its particle size.

NOMENCLATURE
LOI = Loss on Ignition
PSD = Particle Size Distribution
SEM = Scanning Electron Microscope
SFA = Sarawak Fly Ash
SSD = Saturated Surface Dry
XRF = X-ray Fluorescence
Cu = Coefficient of uniformity
Cc = Coefficient of curvature
d = Diameter (μm)

INTRODUCTION
Fly ash is a by-product from the coal-fired power stations. In general, fly ash is spherical in shape and it may consist of some inorganic matter (Pan 2012; Geliga & Ismail 2010). The large scale of production of fly ash from the combustion of coal has triggered some environmental concerns, especially in terms of its disposal method. If dumped in landfills or ash ponds without any proper handling, it may pose a threat to the groundwater system, fertility of agricultural land and air quality. Thus, research and useful applications of fly ash are viewed as important to eliminate such environmental threat.

In recent years, the research works in geopolymer technology has been steadily increasing. Since Davidovits (2008) introduced the name ‘geopolymer’ in 1979, geopolymer has widely been researched and even used in real life works. Geopolymer is a ceramic-like-material which comprises alumino-silicate materials such as fly ash in alkaline environment (Concrete Pavement Technology Program, 2010). The geopolymerisation process is basically activation of alumino-silicate source by the alkaline solution such sodium hydroxide, sodium silicate, potassium hydroxide, potassium silicate etc.

As Ordinary Portland Cement (OPC) generates high emission levels of CO₂ during the manufacturing process, the use of OPC is viewed negatively since global carbon emissions is now a major international issue. According to Carbon Dioxide Information Analysis Centre (2013), the top three world largest CO₂ emission countries are China, USA and India. Malaysia is ranked at 26th, with total CO₂ emission of 59,423 thousand metric tons in 2010. It was reported that the CO₂ emission was 187 million tons in 2006, equivalent to 7.2 tons of CO₂ produced from each Malaysian (SinChew 2009). The rate of CO₂ was 342 parts per million (ppm) in 1980 but rose up to 365ppm in 2000, it had further increased to 396ppm in 2012. By 2020, the CO₂ emission is targeted to reduce by 40% in comparison to the 2005 level (Mahadi 2012). These emissions are partially emitted from the manufacture of cement. There are 11 integrated cement plants with total capacity of 24.5Mt/yr in Malaysia (Edwards 2012); the annual production of cement will increase the environmental pollution.

The low carbon footprint of geopolymer, which is approximately 80% lower than OPC (Concrete Institute of Australia, 2011) has been shown to be an attractive, viable alternative construction material to Ordinary Portland Cement (OPC) if its mechanical behaviour can be fully
and consistently understood. Hence, research and development works play an important role here.

Objective

The main objective of this paper is to study and identify the mechanical properties of Sarawak Fly Ash (SFA). This research includes the studies on SFA chemical composition, particle size distribution, microstructure from the observation through scanning electron microscope and compressive strength of geopolymer using SFA. The in-depth understanding of SFA can enhance the development of geopolymer technology using SFA.

EXPERIMENTAL WORKS

Materials

Fly ash sourced from Sejingkat Power Station was used for this research, which is denoted as Sarawak Fly Ash (SFA) in this paper. Sejingkat Power Station is one of the main power stations in Sarawak that is used to generate electricity. It is located in Kuching, the capital of Sarawak State in Malaysia. It is estimated that 1,400 tons of fly ash is produced daily and 1 million tons of coal is used for the combustion annually. The combustion of coal is carried out in 2 boilers operating independently. The first or older boiler has a capacity to generate 100MW (2x50MW) of electricity, while the second or newer boiler has a capacity to generate slightly more electricity at 110MW (2x55MW). The by-product of the power plant, fly ash, is disposed into two ash ponds nearby.

The chemical composition of the SFA was studied using X-ray Fluorescence (XRF) and the results are presented in Table 1. SFA was classified as Class F type fly ash in accordance with ASTM C618-05. According to ASTM C618-05, fly ash can be classified as Class C or Class F. If the combined percentage of SiO₂+Al₂O₃+Fe₂O₃ is greater than 70%, it is classified as Class F, while greater than 50%, it is classified as Class C. For SO₃ content, moisture content and Loss on Ignition (LOI) content, the requirements for both Class F and Class C type fly ashes are identical, where the contents should lower than 5%, 3% and 6% respectively. The former type of fly ash is generally produced by the combustion of anthracite or bituminous coal while the latter type is produced by lignite or sub-bituminous coal. SFA is classified as Class F.

The other materials used for this geopolymer research were washed sand, sodium hydroxide (NaOH) and sodium silicate (Na₂SiO₃). The sand was prepared to saturated surface dry (SSD) condition. NaOH and Na₂SiO₃ were the main activators used to trigger the geopolymerisation process.

Sample Preparation

Fly ash and sand were pre-mixed for about 1 minute using a mortar mixer. Sodium hydroxide (NaOH) and sodium silicate (Na₂SiO₃) were added into the mixture with continuous mixing for 5 minutes to allow the mixture to blend evenly. The ratio of fly ash to sand and the ratio of Na₂SiO₃/ NaOH for this experiment were 1:2 and 2.5, respectively. The details of mixing ratio of geopolymer using SFA are tabulated in Table 2.

The workability of the mixture was measured using the flow table. The conical mould was filled with the ready mixed mixture. Having reached the top surface of the cone, the tamping rod was used to tamp the mixture in order to ensure the cone was completely filled. After removing the conical mould, 20 counts of electrical taping were applied and the length of the flow was then measured and recorded.

The flow of the geopolymer sample for this research was 170±5% as shown in Table 2. The amount of Na₂SiO₃ and NaOH added into the mixture were 160.7g and 64.3g respectively. It was found that the mixture would become very viscous and difficult to flow when the amount of the alkaline solution was less than the amount as mentioned.

The mixture was poured into the cube mould with dimensions of 5cm x 5cm x 5cm. The mould was vibrated for 1 minute on the vibrating table in order to allow for dissipation of air bubbles within the plastic sheet was covered over the mould and placed into the curing box for heat curing in the oven at 60°C for 24 hours. The sample was demoulded after 24 hours of curing. It was placed at room temperature for further curing until testing.

Compressive Strength Test

The test was carried out using the unified compressive strength test machine (YAW 3000). The cured cube samples were tested. The load applied to the samples was 0.2kN/s. 3 samples were tested in order to evaluate an average strength value.

Particle Size Distribution

Particle Size Distribution (PSD) test was performed using the Particle Size Analyser Cilas 1190. The test procedure involved adding of distilled water and fly ash into the machine. Background measurement and sample measurement were conducted before starting the test. While the test was carried out, sonicator was turned on. The diameter of the particles at 10% (d₁₀), 50% (d₅₀) and 90% (d₉₀) can be obtained from the test results. Also, the mean diameter of the particles can be achieved from the test.

Scanning Electron Microscope

The morphology of the fly ash was studied using the Scanning Electron Microscope (SEM). The equipment
allowed the magnification up to 2μm. For this research, magnification of up to 20μm and 100μm were selected.

RESULTS AND DISCUSSIONS

Chemical Composition

Table 1 shows the chemical composition of SFA and other fly ashes from different countries. The details of fly ash from other countries were sourced from the literature which included the fly ash from Australia, Europe, US, China and India. The chemical composition of fly ash was varied depending on its origins. However, it is obvious that the major constituents of fly ash are SiO₂, Al₂O₃, Fe₂O₃ and CaO. According to ASTM C618-05, Class F type fly ash is classified based on its summation of SiO₂+Al₂O₃+Fe₂O₃, which is greater than 70%. In addition, it is known as low calcium fly ash as it contains lower CaO content.

In comparison, the chemical compositions of SFA are rather similar to those observed from other countries. Overall, the chemical components of SFA were in the range of other fly ashes. The chemical composition of SiO₂, Al₂O₃, Fe₂O₃ and CaO on SFA was 43.8, 18.1, 7.7 and 3.9 respectively. The research and application of the fly ash sourced from these countries have long been developed. SFA which had similar chemical content to these fly ashes showed the potential of having similar attractive properties as being a potential constituent for geopolymer research.

For the LOI content, SFA had very low unburned carbon content, which was 0.49%. According to ASTM C618-05, the LOI of Class F type fly ash should less than 6%. Thus, the LOI content of SFA was in the range as required in ASTM.

Davidovits (2008) stated that the optimal binding properties could be achieved when CaO content was low, Fe₂O₃ and unburned carbon content less than 10% and 5%, the vitreous phase was high, reactive silica ranged from 40–50% and 80–90% of the particle size less than 45μm. With this in mind, SFA which was in line with the findings had shown the potential to be used as an environmental friendly and sustainable construction material.

Particle Size Distribution

The results of the Particle Size Distribution (PSD) on SFA were presented in Table 3 and Figure 1. SFA had a broad particle size distribution with particle predominantly (d₉₀) smaller than 40μm and mean diameter of 16μm.

The percentage of SFA steadily increased with the particle size in the range less than 60 μm. The distribution can be divided into 4 groups, (i) 0.04μm to 0.6μm (4%), (ii) 0.6μm to 6μm (30%), (iii) 6μm to 20μm (31%) and (iv) 20μm to 70μm (35%). The particles were distributed evenly in the range of 0.6μm to 70μm, indicated as (ii), (iii) and (iv). The diameter of SFA particle at 10% (d₁₀), 50% (d₅₀) and 90% (d₉₀) were 1.75μm, 11.26μm and 40.63μm, respectively. Mehta (1985) concluded that the reactivity of fly ash was proportional to the particle size smaller than 10μm, which gave higher compressive strength. Inversely, the particle sizes greater than 45μm, would lead to slower rate of strength development. With this in mind, SFA contained 46% of particle size smaller than 10μm and 7% of the particle size greater than 45μm, and thus for the strength development on geopolymer this can be considered favourable.

The Coefficient of uniformity (C_u) and Coefficient of Curvature (C_c) were used as an indication to describe the particle size distribution of SFA. It was found that the particle size distribution of SFA was well-graded. The C_c of SFA was 1.32 and the C_u was 6.29. This implied that the SFA contained wider ranges of particle sizes and the particles are closely packed together and therefore having lesser voids.

Scanning Electron Microscope

The morphology of SFA can be studied through Figure 2. It is observed that SFA has less smooth spherical spheres. It consisted of some irregular grains which were greater in size than those spherical particles. Few broken spheres have been observed on SFA. However, the majority of the SFA particles were complete spheres. Thus, it was expected that SFA can react well with the alkaline solution for the geopolymerisation process.

Compressive Strength

The compressive strength test on SFA was conducted after 1 day and 7 days as presented in Table 2. The test results showed that the compressive strength of SFA was initially 42MPa at day 1 but increased to 55MPa after 7 days. It may be due to some unreacted fly ash particles, which did not undergo geopolymerisation during heat curing in oven but continued to react with the alkaline solution when curing at room temperature. Se et al (2012); Kumar et al. (2009) reported that the strength gained with the curing age was probably due to the pozzolanic reaction on the reactive fine particles of fly ash to form the aluminate silicate gel which consequently improved the bonding in geopolymer development. Similar results were reported in Chindaprasirt et al. (2011); Fernandez-Jimenez and Palomo (2003); Se et al. (2012).

Kumar et al. (2007) stated that fine particles with higher specific surface would enhance the reactivity of the particle. The results obtained from this research were in agreement with the results presented by Chindaprasirt et al. (2011) mentioned earlier, the smaller particles would lead to higher compressive strength. The results obtained from this research have shown that SFA, which has 90% of the particle sizes in overall smaller than 40μm, and thus most likely attributed to the good strength development.
CONCLUSIONS

Sarawak Fly Ash which was collected from Sejingkat Power Station showed good potential of being use as an attractive construction material. As its chemical composition was rather similar to other fly ashes details from Australia, Europe, US, China and India, it is believed that the fly ash behaviour in geopolymer would be similar to these fly ashes. As the research and application of the fly ash sourced from these countries have long been developed, thus, it is believed that SFA has tremendous benefits to the development of geopolymer in Sarawak.

From the morphology studies, SFA was observed to consist of spherical particles but of less smooth surfaces. Some irregular grains were found as greater in size than the spherical particles. Majority of the particle sizes of SFA, which represented 90% of overall, was smaller than 40μm. For SFA, the particle size smaller than 10μm was 46% and greater than 45μm was 7%. Due to the presence of much higher portion of fine particles, generation of higher geopolymer compressive strength seems highly possible. The compressive strength of geopolymer using SFA after 1 day and 7 days were assessed to be 42MPa and 55MPa, respectively. The strength improved with the curing age was most probably due to the further geopolymerisation.

ACKNOWLEDGEMENT

Financial support from Melbourne Sarawak Research Collaboration Scheme for carrying out this collaboration project is gratefully acknowledged. The author expresses her sincere gratitude to the Centre for Sustainable Infrastructure (Hawthorn Campus) and Swinburne Sarawak Research Centre for Sustainable Technologies for her giving the opportunity to travel to Melbourne to conduct the experiment works. The author would like to acknowledge the assistance, guidance and support rendered by Ms. Chandani during the collaboration project are gratefully appreciated.

REFERENCES

ASTM C618-05, Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete, ASTM International.
Carbon Dioxide Information Analysis Centre 2013, Ranking of the world's countries by 2010 total CO₂ emissions, <http://cdiac.ornl.gov/trends/emis/top2010.tot>
Concrete Institute of Australia 2011, Recommended Practice Geopolymer Concrete, viewed 17 January 2013.
Concrete Pavement Technology Program 2010, Geopolymer Concrete, viewed 20 January 2013.
Mahadi, N 2012, 40 percent CO₂ emission target within sight — Chin, Borneo Post Online, <http://www.theborneopost.com/2012/06/19/40-per-cent-co2-emission-target-within-sight-chin/>
### Figures and Tables

#### Table 1: Chemical Composition of Fly Ash

<table>
<thead>
<tr>
<th>Elements (%)</th>
<th>Australia</th>
<th>Europe</th>
<th>US</th>
<th>China</th>
<th>India</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>43.8</td>
<td>48.8-66</td>
<td>28.5-59.7</td>
<td>37.8-59.5</td>
<td>35.6-57.2</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>18.1</td>
<td>17-27.8</td>
<td>12.5-35.6</td>
<td>19.1-28.6</td>
<td>18.8-55</td>
</tr>
<tr>
<td>Fe$_2$O$_3$</td>
<td>7.7</td>
<td>1.1-13.9</td>
<td>2.6-21.2</td>
<td>6.8-25.5</td>
<td>2.3-19.3</td>
</tr>
<tr>
<td>CaO</td>
<td>3.9</td>
<td>2.9-5.3</td>
<td>0.5-28.9</td>
<td>1.4-22.4</td>
<td>1.1-0.6</td>
</tr>
<tr>
<td>MgO</td>
<td>0.45</td>
<td>0.3-2</td>
<td>0.6-3.8</td>
<td>0.7-4.8</td>
<td>0.7-4.8</td>
</tr>
<tr>
<td>K$_2$O</td>
<td>2.04</td>
<td>1.1-2.9</td>
<td>0.4-4</td>
<td>0.9-2.6</td>
<td>0.8-0.9</td>
</tr>
<tr>
<td>Na$_2$O</td>
<td>0.27</td>
<td>0.2-1.3</td>
<td>0.1-1.9</td>
<td>0.3-1.8</td>
<td>0.6-1.3</td>
</tr>
<tr>
<td>SO$_3$</td>
<td>0.08</td>
<td>0.1-0.6</td>
<td>0.1-1.2</td>
<td>0.1-2.1</td>
<td>1-2.9</td>
</tr>
<tr>
<td>TiO$_2$</td>
<td>0.61</td>
<td>1.3-3.7</td>
<td>0.5-2.6</td>
<td>1.1-1.6</td>
<td>0.2-0.7</td>
</tr>
<tr>
<td>P$_2$O$_5$</td>
<td>0.078</td>
<td>0.2-3.9</td>
<td>0.1-1.7</td>
<td>0.1-0.3</td>
<td>1.1-1.5</td>
</tr>
<tr>
<td>BaO</td>
<td>0.025</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LOI</td>
<td>0.49</td>
<td>-</td>
<td>0.8-32.8</td>
<td>0.2-11</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Table 2: Details of geopolymer samples

<table>
<thead>
<tr>
<th>Ash/sand ratio</th>
<th>Na$_2$SiO$_3$ (g)</th>
<th>NaOH (g)</th>
<th>Na$_2$SiO$_3$/NaOH ratio</th>
<th>Liquid/ash ratio</th>
<th>Flow (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:2</td>
<td>160.7</td>
<td>64.3</td>
<td>2.5</td>
<td>0.45</td>
<td>170±5</td>
</tr>
</tbody>
</table>

Compressive Strength (1 day) 42MPa  
Compressive Strength (7 days) 55MPa

#### Table 3: Test results from PSD

<table>
<thead>
<tr>
<th>Particle diameter</th>
<th>SFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Diameter at 10% (d$_{10}$)</td>
<td>1.8μm</td>
</tr>
<tr>
<td>(b) Diameter at 50% (d$_{50}$)</td>
<td>11μm</td>
</tr>
<tr>
<td>(c) Diameter at 90% (d$_{90}$)</td>
<td>40μm</td>
</tr>
<tr>
<td>(d) Mean diameter</td>
<td>16μm</td>
</tr>
</tbody>
</table>

**Particle size distribution (group range)**

- (i) 0.04 – 0.6μm 4%
- (ii) 0.6 – 6μm 30%
- (iii) 6 – 20μm 31%
- (iv) 20 – 70μm 35%

- **Total** 100%

<table>
<thead>
<tr>
<th>Particle size</th>
<th>SFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle size &lt; 10μm</td>
<td>46%</td>
</tr>
<tr>
<td>Particle size &gt; 45μm</td>
<td>7%</td>
</tr>
</tbody>
</table>

---

**Figure 1:** Particle size distribution of SFA

(a) SFA with magnification of 100μm  
(b) SFA with magnification of 20μm
(c) SFA with magnification of 20µm

Figure 2: SEM on SFA
Practical Management of Access Permission in the Inter-organizational Communication

Masahito Gotaishi, Hiroshi Yamaguchi, Shigeo Tsujii
R & D Initiative, Chuo University
1-13-27 Kasuga, Tokyo, Japan 112-8551
gotaishi@tamacc.chuo-u.ac.jp

ABSTRACT: Information security of the communication between multiple organizations is discussed. Both the communication between organizations enables the leader of the organization receiving the confidential information to assign the access permission to appropriate members. The procedure of managing the information security varies according to the kind of business and structure of the team.

Technical solutions to enable management of the access permission is also described.

Keywords—Cryptosystem, Modern Cryptosystem, Access Control, Organization Management, Network-centric organization

I. Introduction

As we know from experience, most of the business, military, medical, or political activities are done by using some kind of classified information. Although it is necessary to protect the confidentiality, the business never proceeds without making effective use of the information. Therefore the majority of the effort of managing the information security is spent on maintaining the confidentiality of the information. Or, it could safely be said that the essential technology of information security is "access control," protecting the information from unauthorized access while assuring the legitimate user to access it.

Although it has been always considerably difficult to manage the access permission, it is becoming far more difficult these days, when cooperation and inter-organizational activities are increasing both in business and various other activities. If all activities of the whole operation is completed by a single organization, it would be easier to manage the information security, because the whole organization structure is disclosed to all stakeholders including the leader. However, these days significant amount of the activities are outsourced. Additionally, the activities left in-house are usually done in cooperation with other organizations and the knowledge needs to be shared among the participants.

As discussed above, when confidential information is exchanged between different organizations, the information should be protected not only from the outside but also the unauthorized members within. The classified information should be disclosed exclusively to the people in the role to use the information. But the member of the company sending the information does not know the organization structure of the company receiving it. Therefore the sender cannot assign access permission to the information being sent.

For example, the important information such as medical record of patients or information about the M & A should be disclosed exclusively to the members included in the activities using the information.

As an application of public key cryptosystems to the management of access permission in the organizations, Attribute-Based Encryption (or Functional Encryption) is extensively developed. In these systems, a sending organization must identify the adequate or qualified receiver in the receiving organization by embedding the capacity of receiving encryption message or key. For example, it is easy to embed the permission of viewing charged television for broadcast companies using the broadcast cryptosystem. However, it is sometimes difficult for the sender, who is out of the organization to identify the qualified receiver. In this case, the authors propose a new cryptosystem, with which the leader of the receiver organization can assign the access permission without decrypting the cipher. We call the new system "Organization Cryptosystem."

II. Background

A. Trend of Collaboration and Inter-organizational Activities

One of the most symbolic example of the prevalence of collaborative activities would be the military organization. The military organization has long been working separately from other organizations including police. Nonetheless, as Alberts et al. pointed out [1], the military organizations have been engaged in the operation as allied forces in the Gulf War. Additionally, the current military operation did not end at the end of the war. The allied forces had to work on the liberalization and they had to restore the order in the area.

The commanders in the field had to cooperate with the NGOs and other not-for-profit organizations.
It is pointed out that the counterterrorism operations must be done in cooperation with other organizations including FBI.

This trend has been accelerated by several breakthroughs of ICT technologies including cloud computing and various other applications including social networking services (SNS). Currently it is quite easy to build a dedicated server for a given activity such as some kind of collaboration among multiple companies.

B. Sharing Information among Different Organizations

With the substantial improvement of the computing power and high-speed internet connection, more advanced information security function is required and therefore the conventional security protection system is becoming insufficient in coping with the emerging information security problem, because of the following changes in the society:

- The explosive increase of the data and complication of the organization
  As widely known, progress of the computer system and telecommunication led to the increase of both the amount and variety of information processed and exchanged on the internet, consequently causing the trend of ‘big data.’ Collaboration among multiple organizations are commonly planned and often earned success.

- Multi-Stakeholder
  Therefore there are numerous stakeholders, who are included in the activity from various aspects, in a single business.

- Urgent need of maintaining confidentiality of the information
  As the ‘identity theft’ increases and becomes serious these days, it has become important to protect the personal data as well as delicate information kept in administrative organizations, companies, nursing care suppliers, or hospitals.

- Drastic Change of the IT Environment
  Besides Cloud computing, there are various kind of progress in the IT society. As the numerous service or strategic movements utilizing ‘Big Data’ emerge, it became possible to reveal activities of individuals by analyzing various kinds of data. With the employment of ‘national code number citizen identification’ act in most nations, personal information must be processed with the greatest possible care.

One of the good examples would be the data used in taxation. Several managers of the local governments worry about the access management of personal data such as the income or the value of the fixed property.

If the business is done by a single organization, there is no need of ‘Organization Cryptosystem.’ However, as the high-speed telecommunication system is developed and accordingly the ‘Ubiquitous Society’ is realized, most of the current businesses need collaboration with other companies, not-for-profit, or public sectors. Consequently, lots of stakeholders are included and the both amount and variety of information increase.

For example, as cooperation between hospitals and home care service providers progressed, telecommunication came to take the place of oral direction or paperwork. Simultaneously sharing information is required to comply with the regulation and keep the rationality of the logic.

One of the most desirable safeguards would be cryptosystem. However, although existing cryptosystems can protect information from the eavesdroppers, they can not regulate the access to the information within the organization. For that purpose, the ciphertext must be decrypted to classify each part of the information. Subsequently it should be re-encrypted according to the classification level. It is important to note that the plaintext is revealed during this procedure. Although it is of little problem when the system is running on conventional systems, it might bring about security breach in the cloud system, where the people operating the system should not be trusted. Besides, if the decrypted data is written on a temporary file, can often become a foothold of attack.

Therefore, it would be preferable assigning permission to access a specific part of the document to relevant members without decryption. The actual operation would be done in the following procedure:

1. The receiver organization sends the public key to the sender. He/She has the secret key corresponding to the public key.
2. The sender encrypts the information with the public key to send it to the receiver.
3. The representative of the receiver classifies each part according to its content.
4. The representative of the receiver decides which member should be allowed to access which parts of the information.
5. The representative of the receiver generates the secret key which can decrypt only the parts which he/she is permitted to access.

In implementing ICT systems/networks in the operations using, very delicate information such as medical/care service, customers are often keen on the inappropriate use of the data or insufficient management of the data access.
C. The properties required of the Inter-organizational confidential communication

The background described above requires following properties of the inter-organizational communication [2]:

1. Property of the Transmitted Information
   The information should be appropriate for the organization structure and division of role (considerably in large volume, such as clinical history, condition, or medication history)

2. Confidentiality of the Transmitted Information
   The transmitted information should not be disclosed to anyone except the sender and receiver.

3. Distribution of the Information should be decided by the Receiver
   The sender can leave the decision of access permission to the receiver.

4. Optimum distribution of the information
   The communication system should allow optimum distribution of the information corresponding to the division of the role, properties of the members, or situation of the organization.

Most of the clients of the care service have disease such as cancer or lifestyle disease, or some disabilities. Therefore, caretakers must have access to the clients’ medical record, otherwise they cannot supply even a simple meal assistance. Besides, since the clients are typically living on their own, even their address should be treated carefully. On the other hand, most of the helpers, who give the care service, do not belong to the care service provider but are sent from temp agency and the assignment changes very frequently. It rarely occurs that a helper takes charge of a client for a long time. Hence the manager of the care service provider must quickly decide which information is necessary in giving service to a given client. Additionally, it often occurs that the care service provider changes. So the manager assigns Ms. Jones to take care of her. Subsequently he needs to supply the above information to Ms. Jones, without disclosing it to other members, or, desirably, even to system administrator of the cloud system.

In the situations like this, Organization Cryptosystem would be very effective.
scribing the information and its confidentiality to each information unit.

Then the director of the receiving organization reads the metadata to find that it is about the project PJ13A21. Hence he forwards the information of the project leader. The leader assigns the access permission to members, as illustrated in the Fig. 3. The document about the business proposal should be assigned to the sales representative, and a draft of the agreement be assigned to the legal, etc.

**B. Flat Organizations: Medical and Care Services**

After we made several interviews to ASP service providers for care services and several medical practitioners, it was found that the medical services giving treatments to a given patient is operated by a flat organization, without any manager or leader.

As shown in the Figure 4, the network of the professionals expands on the judgment of each participant. Here Mr. Smith suffering from diabetes is taken care of by a physician. She is worrying about the influence of the diabetes on his eyes. So she makes contact to the ophthalmologist of Mr. Smith. Or other members taking care of Mr. Smith would expand the network by inviting other people. There is no person managing or leading the movement of the expansion. Each member decides who to invite to the network or what kind of information should be told to whom as a professional. This is entirely a flat organization.

**IV. Technology of Cryptography enabling the Inter-organizational Access Control**

In assigning the decryption key to the legitimate users, some technical solutions have been proposed:

**A. Multivariate Public Key Cryptosystem: For Hierarchical Organization Model**

The access control system for the hierarchical organization is based on the Multivariate Public Key Cryptosystem (MPKC). [3][4] Multivariate Public-Key Cryptosystems (MPKC) is a public-key cryptosystem which uses the one-wayness that "it is easy to substitute values in multivariate polynomials, but it is difficult to solve the equation system." The MPKCs are faster than RSA or elliptic curve cryptosystem both in encryption and decryption, because MPKC does not include any exponentiation. Moreover, typically the degree of the MPKC public keys is as low as two. As a lightweight cryptosystem, they were initially expected to be used for RFIDs or IC cards. After that, the variants of MPKCs were actively proposed worldwide around the early 1990s. Nonetheless, in spite of the numerous proposals published in papers, provably secure, efficient systems have not been proposed yet.

Original purpose of the development of MPKC in the 1980s in Japan was to develop a public key cryptosystem faster than RSA. But in 1994, when it was shown that the problems of number theory such as prime factorization or discrete logarithm could be computed in polynomial time if quantum computers are put into practical use [5], MPKCs were focused as one of the likely candidates of the 'Post-Quantum Cryptosystem.' If these problems of number theory were to be computed efficiently, most of the existing public-key cryptosystems including RSA and ElGamal would be threatened. Since solving the system of algebraic equations is proved to be NP-complete and there has not been an efficient algorithm for quantum computers to solve them is not discovered yet, MPKCs has been expected to remain secure even in the post-quantum society. Consequently MPKCs has been actively developed worldwide. Currently MPKC is regarded as one of the likely candidates of post-quantum cryptosystems.
and therefore the textbook of post-quantum cryptosystem [6] spends considerable length of chapters on them. A dedicated textbook illustrating the existing MPKCs and cryptanalysis was published in 2006 [7].

Hence currently MPKC is attracting attention as a lightweight public key cryptosystem, with quick encryption and decryption, and a post-quantum cryptosystem.

The MPKC used for the organization cryptosystem is Triangular Secret Key (TSK), one of the basic trapdoors. Tsujii proposed Sequential Solution Method [8], which was compiled in a paper in 1986. However, it was cryptanalyzed by Kaneko et al. Tsujii et al. improved the Sequential Solution Method and its reinforced version was proposed in 1989. This method, rational map, was cryptanalyzed by Ding et al., [9] about 10 years after the creation. Adi Shamir also proposed the Sequential Solution Method for signature (birational permutation), [10] independently of Tsujii. His system was cryptanalyzed by Coppersmith, et al. After the year 2000, Kasahara et al. presented some generalized structures of sequential solution method [11].

Generally MPKCs are structured as shown in Figure 5. The equation system of the central map $w = G(u)$ is easily solved by exploiting the trapdoor. Hence the structure of the central map is disguised by the two affine transformations $S, T$. The polynomial system $E(x)$ is published as the public key. The secret key is constituted by the central map $G(u)$, and the inverse of the two affine transformations $S^{-1}, T^{-1}$.

Encryption is done by just assigning the value of $x_1, \ldots, x_n$ to the polynomials of the public key $E(x)$ to compute the cipher text $y$. Decryption is done in the following way:

1. Affine transformation $T$ is inverted: $w := T^{-1}(y)$
2. The equation $G(u) = w$ is solved.
3. Affine transformation $S$ is inverted: $x := S^{-1}(u)$

B. Other Ways to Realize the Access Control

The above cryptosystem has been proposed for hierarchical organizations. The teams with flat structure would need different kind of technology. Proxy re-encryption methods, a kind of public-key cryptosystem, which have been proposed by Mambo et al. [12] enables the person with the secret key to transform the ciphertext into another ciphertext decryptable by another person’s secret key.

We have designed a security system utilizing elliptic-curve ElGamal cryptosystems which allows quick transformation of the ciphertext into the one decryptable by different secret key. It is being developed to implement in a security systems for ASP services for care service providers.

V. Conclusion

A. Need of the New Cryptosystem for Access Control

Multivariate Public Key Cryptosystem is a lightweight cryptosystem, which allows quick encryption and decryption. Additionally, since its key has a structure analogous to the ordinary organization, it is quite fitted for application to the access control within an organization. As discussed in the chapter of Background, it is quite important to keep the stored data by encryption, especially when they are stored in the cloud environment. Therefore we propose the security system to protect the information while its access permission is defined.

B. Future Work

With its security confirmed, the next step would be to confirm the performance. We are going to confirm the advantage of fast encryption and decryption with the experiment with computers. Currently implementation on the production-scale network to test its operability is also planned.

ACKNOWLEDGEMENT

This study is supported by the Project entitled “Development of Public-key Cryptosystem for confidential communication among Organizations (17201)” of the National Institute of Information and Communications Technology (NICT), an independent administrative corporation of Japanese Government.
REFERENCES


WORKSHOPS
The workshop opened with a short “setting-the-scene” introduction by Bernd Krämer in which he

- Questioned the traditional text book approach to teaching, which locks educational ideas in closed formats and closed copyrights and puts the student in the position of a mere consumer,
- Presented the advantages of open education systems with short authoring, editing, quality control, publishing, and distribution with constant feedback from peers and students who can now also become learning content producers,
- Addressed related research questions incl. the visibility and sustainability of digital learning materials, the heterogeneity of end systems (learning management systems [LMSs], learning environments and tools, authoring systems and more), adaptation needs, or the need for codifying and sharing codified educational practices and patterns
- Reviewed the model of open educational resources\(^1\), which does not only refer to educational content (e.g., courses, learning objects, reference works) but also to development and provision components (such as LMSes, social software, development tools), implementation resources and guidelines (such as standards, quality and design principles, licensing models and tools),
- Sketched the functionality and architecture of the open source learning content repository network edu-sharing\(^2\), which his group has developed since 2005, and finally,
- Compared claims and reality of a new kid on the block, MOOCs.

The workshop organizers Pratap Chillakanti and Bernd Krämer then raised a series of questions that were first discussed in 6 smaller brainstorming groups, then with the audience at large. Below we list the questions together with the contributions from the group. The organizers promised to provide a summary of the workshop (with the help of two kind Swinburne students who took discussion notes). The workshop participants, who were overly engaged, raised their interest in furthering the discussion online and collaboratively working towards the objective to present substantial advances at the next SDPS conference.

1. **What are the tangible objectives to promote online education through SDPS?**

   - Leverage awareness of the society

\(^1\) OECD (2007): Giving Knowledge for Free – The emergence of OER

\(^2\) [http://www.edu-sharing.net](http://www.edu-sharing.net)
• Achieve leadership in disseminating expert knowledge about trans-disciplinary research and education to end up in a system of satellites multiplying the knowledge diffusion process
• Publish conference papers, journal contributions, research data, which should be linked to the SDPS journal
• Give access to domain expert knowledge in the focus of the society (this would be a place for people to leave their intellectual footprints)
• Provide responsive technology/framework (e.g., Wordpress)
• Offer a course/content broker to help people find content matching their needs (not necessarily content maintained on the society website)
• Publish practical problems illustrating the need for trans-disciplinary approaches
• Low-barrier provision to reach learners living in rural areas
• Enhance visibility of SDPS in different social media

In this context the question was raised how one can define quality in relation to diverse audiences. One subgroup claimed that courses would no be in focus here. In response, Bernd Krämer referred to continuing education courses that members of the society offered in the past to engineers in the US. The group would like to know whether this know-how can be made available.

2. How can you work with us to meet these objectives?
   a. Think outside the box as to what kind of collaboration models you would use to meet the objective?
   b. Think of this question as what role you can play to move this forward.

• Collaborative interactive environment
• Student contests
• Marketing
• Easy access to upload of content
• Collect metadata from useful resources from everywhere, paraphrase, filter, and make accessible to content broker

3. How would you use the technology provided by SDPS to create content for public consumption?

• Invite experts from other fields than just engineering and science
• Publish not just papers but also (raw) research data
• Develop training materials addressing fields like social science or environmental economics

4. How can we form collaborative partnership with the institutions to take this forward (meaning SDPS is opening up for future growth of membership)?

An immediate question raised was: What is the relationship to be established? SDPS should establish an OE chapter
Forum for exchanging and discussing ideas (like threads in ResearchGate)
Workshop “Open Education”

Organizers: Bernd Krämer, FernUniversität in Hagen; Murat Tanik, UAB; ...

Open Education describes both a political movement and a methodological approach to making education freely available. The attribute open refers to teaching materials, including textbooks, courseware and other resources, that are freely shared through reuse, redistribution, revision and remixing. Closely related concepts like Open Content, Open Educational Resources (OER), Open Access, or open textbooks pursue the same objective from different perspectives, namely sharing and giving away for free. OER form the basis of this development. They comprise freely licensed teaching materials, educational scenarios, best practice reports, and software supporting teaching and learning. However, open education goes beyond OER by supporting collaborative learning capabilities and possibilities to test and certify acquired knowledge and competences. Open education is not the same as e-learning, but it can use the same technical means, including content repositories, OER, learning environments, and social software.

The workshop will initiate a broad discussion on Open Education with two or three short keynotes that will lead into the subject, raise challenges, and sketch promising solution.
Effective end-to-end management of dynamic manufacturing networks is consistently touted as a top priority for manufacturing enterprises that need to strive to improve their efficiency, adaptability and sustainability of their production systems. Moreover, it is a crucial prerequisite for the emerging powerful new model of production based on community, collaboration, self-organisation, and openness rather than on hierarchy and centralised control.

This workshop will be centered around the outcomes of the EU-sponsored Integrated Project IMAGINE, which addresses the need of modern manufacturing enterprises for a new end-to-end management of dynamic manufacturing networks and will develop a multi-party collaboration platform for innovative, responsive manufacturing that encompasses globally distributed partners, suppliers & production facilities (SMEs and/or OEMs) that jointly conduct multi-party manufacturing. The project, which is now its third and final year, is currently implement a unique comprehensive methodology for the management of dynamic manufacturing networks. The methodology and an adaptable generic service platform provide a consolidated and coordinated view of information from various manufacturing sources and systems and enable service-enhanced product and production lifecycle and responsive manufacturing processes throughout the value chain. Living Labs in major industrial sectors drive the implementation, testing, evaluation and dissemination of the IMAGINE methodology and supporting ICT platform. The IMAGINE manufacturing model is an innovative plug and produce approach that implements an end-to-end manufacturing interoperability solution.

The workshop will be opened with a presentation of key findings of the IMAGINE project and will then open the floor for discussion on the presentation topics and related issues brought forward by the participants.
AUTHORS

Abdel-Rahman Tawil ........................................... 107
Abu Hassan Md Isa .................................. 423
Afzal Ahmed ................................................ 457
Ali H. Dogru ............................................. 251
Ali Shakil Khan ........................................... 13
Ana Sinkovic .............................................. 293
Anatoli Vakhguelt ........................................ 125
Anna Abelló Pla ........................................... 33
Anna Andreu Povar ...................................... 33
Anuar Ahmad .............................................. 267
Arie Nakhmani ............................................ 181
Armstrong Ighodalo Omorogie .................. 261
Arul Arulrajah ............................................. 233
Aurilla Aurelie Arntzen Bechina ............... 447
Ayako Fukunaga ........................................ 177
Basil T. Wong ............................................. 125
Brian Lynch .............................................. 465
Brikat Al-Mouz ............................................ 7
Bu-Yeon Jo ................................................. 345
Byung H. Moon ........................................... 311
Chandana Gunathilaka ................................ 431
Chin-Hong Puah ....................................... 389, 417
Chong Eng Tan ......................................... 141
Choonhee Yang ......................................... 331
Chua Hong Siang ....................................... 149
David Preston ............................................. 107
Daw Tin Hla ............................................. 423, 431
Djamshid Tavangarian .................................. 57
Dolores Rexachs ........................................ 33
Dominic Ek-Leong Ong ......................... 233, 261
Dong-Su Yoo ............................................. 341
Duaa Fatima S. Khan .................................. 37
Edwin Mit ............................................... 275
Eiman Almami ........................................... 301
Ek-Leong Dominic Ong ............................ 465
Emilio Luque ........................................... 33
Eric Gampher ............................................ 181
Farouq Twaiq .......................................... 7
Fayi Zhou ............................................... 87
Fung-Thai Thien ......................................... 417
Gu-Hyung Jung ......................................... 345
Hamda Binghubash Almarri ..................... 281
Hamid Bagha ............................................ 319
Heinz Wörn ............................................. 133
Hemanshu Roy Pota .................................. 211, 219, 227
Heon-Yong Jung ....................................... 349
Hiroshi Yamaguchi ..................................... 471
Hong Yi .................................................... 165
Ho-Soung Kong ......................................... 439
Hsiao-Yun Leong ....................................... 465
Hussain Al Balushi ..................................... 7
Hwang Kyoo-il .......................................... 335
Hyeonkoo Cho .......................................... 157
Ichikawa Masahiro ..................................... 189
Igor Tesanovic ......................................... 293
Ingolf Waßmann ....................................... 57
Inhwa Kim ................................................. 293
Jane Labadin ............................................. 51
Jarl Simonsen ......................................... 447
Jay Sanjayan ............................................. 465
Jeong T. Ryu ............................................. 311
Jinoh Kim ................................................ 157
Johannes Schneider ................................... 133
Jonathan Sidi ............................................ 275
Jordi Esquirol Caussa ............................. 33
Justin Tharp ............................................. 157
Kai-Hung Wong ....................................... 389
Kartinah Zen ........................................... 241
Vanessa Bayo Tallón ............................... 33
W. H. Hii ............................................. 355
Wah C. R. J. ............................................ 375
WaiShiang Cheah ................................. 275
Weixiang Shen ................................. 211, 219, 227
Wilhelm Stork ...................................... 133
Xiaoguang Deng ................................. 87, 165
Xuan Sun .......................................... 87
Yasuaki Nishimoto .............................. 47
Yew Wong ........................................... 233
Yong Zeng ......................................... 87, 165
Yongman Kwon .................................... 443
Yujiro Ogata ........................................ 177