



**PROCEEDING OF POSTGRADUATE SEMINAR  
ON SCIENCE AND TECHNOLOGY**

*7<sup>th</sup> November 2016, Universiti Sains Islam Malaysia,  
Nilai, Negeri Sembilan, Malaysia*

***"Flourishing Ummah Well-Being Through Islamic  
Science Exploration"***

**Editors:**

**Farah Wahida Harun  
Farida Ridzuan  
Nurlida Basir  
Yumn Suhaylah Yusoff  
2017**

**PUBLISHED IN 2017**

© Faculty of Science and Technology, Universiti Sains Islam Malaysia

All rights reserved, no part of this publication may be reproduced, stored in retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission from the Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM).

eISBN: 978-967-440-399-7



PROCEEDING OF POSTGRADUATE SEMINAR ON SCIENCE AND TECHNOLOGY

Farah Wahida Harun  
Farida Ridzuan  
Nurlida Basir  
Yumn Suhaylah Yusoff

## Contents

<b>Analysis of Gelatin DNA in Supplement Capsule Shell using Real-Time Polymerase Chain Reaction for <i>Halal</i> Authentication .....</b>	<b>1</b>
<i>Inayatullah, A., Hassan, M. S. and Husaini, M. A. R.</i>	
<b>Comparative Analysis Of Surface Antigens (SAG) Proteins In Three Major Eimeria Species Causing Coccidiosis In Domestic Chickens Using Bioinformatic Approach .....</b>	<b>9</b>
<i>Manaf, Z., Ramly, N. Z., Wan, K. L. and Firdaus Raih, M.</i>	
<b>Morphological and Molecular Identification of Bacteria Associated with <i>Holothuria (Mertensiothuria) Leucospilota</i> from Pangkor Island.....</b>	<b>18</b>
<i>Muhd Lutfi, F. N., Mohamed Rehan, M., Kamarudin, K. R., Che Zahri, M. Z., Mohd Noor, H. and Md Radzi, S.</i>	
<b>Structural Study of Proteases from <i>Bacillus Lehensis</i> G1.....</b>	<b>30</b>
<i>Sulaiman, N. A. and Ramly, N. Z.</i>	
<b>Preliminary Study on Dimensional Stability of Oil Palm Lumber Treated with Rosin-Gum Resin.....</b>	<b>43</b>
<i>Hamid, A. B., Kamarudin, H., Hassan, A. R. and Jalani, B. S.</i>	
<b>Enhance CO<sub>2</sub> Hydrate Formation Inside the HPVA by Vigorous Stirring During Sample Preparation .....</b>	<b>53</b>
<i>Mohd Hafiz, A. H., Snape, C. E. and Stevens, L.</i>	
<b>DFT Calculations of 3C- and 2H-SiC Structural and Electronic Properties and Their Response to a Hydrostatic Pressure .....</b>	<b>73</b>
<i>Enhemed, A. M. M. and Rosli, A. N.</i>	
<b>A Robust Ridge Regression for Multicollinearity Problem in the Presence of Outliers in the Data .....</b>	<b>83</b>
<i>Ferdaos, N. A. and Mohamad Shariff, N. S.</i>	
<b>Modified HPM for Linear Fractional Integro-Differential Equations of Fredholm-Volterra Type .....</b>	<b>91</b>
<i>Hamson, K. M., Eshkuvatov, Z. K. and Taib, B. M.</i>	
<b>Bachok--Hasham Polynomials Approach for Eigenvalues Problem of Logarithmic Singular Integral Equations .....</b>	<b>106</b>
<i>Alhawamda, H., Taib, B. M., Eshkuvatov, Z. K. and Ibrahim, R. I.</i>	
<b>A Review of Security Methods for Mobile Cloud Computing.....</b>	<b>119</b>
<i>Elfurgani, Z. B. M. and Abdullah, I.</i>	
<b>An Approach for Effective Volunteer Placement During Post-Disaster .....</b>	<b>135</b>
<i>Basarang, S. N., Azizan, N., Bahari, M. and Ismail, W.</i>	

<b>A Systematic Review Analysis for Scam Detection and Response Using Apoptosis.....</b>	145
<i>Aliyu, L. Y., Saudi, M. M. and Abdullah, I.</i>	
<b>A Review on Intrusion Prevention System (IPS) Based On Bio-Inspired Algorithm.....</b>	158
<i>Nadir, A. M., Saudi, M. M. and Touhid, B.</i>	
<b>Clustering Algorithm to Improve Network Connectivity In WSNS for Correlated Nodes Behaviour: A Systematic Review.....</b>	167
<i>Marsal, K. A., Azni, A. H. and Ridzuan, F.</i>	
<b>Quality Effect on Students' Satisfaction Towards Student Information Systems (SIS) in a University .....</b>	188
<i>Mohamed, Y. O., Fabil, N., Alwi, N. H. M. and Azizan, N.</i>	
<b>Concept, Practicality and Reality in Malaysia Government Project Payment Certification of Final Account .....</b>	199
<i>Waluyob, S. N, Darus, Z. M. and Ismail, M. N.</i>	

## PREFACE

The Postgraduate Seminar Faculty of Science and Technology 2016 (KoSiST 16), formerly known as Postgraduate Colloquium of Science & Technology is designed as an ideal platform for postgraduate students in the area of science and technology to discuss, interact and revise their research project. The seminar was held on November 7, 2016 at Dewan Tuanku Canselor, Universiti Sains Islam Malaysia (USIM). The seminar has triggered intellectual discussion among students, researchers and lecturers to improve their research, writing and communication skills. The annual seminar with carefully-selected theme, “*Flourishing Ummah Well-Being through Islamic Science Exploration*” aspired competent postgraduate students to present their current research findings.

KoSiST 16 was held in conjunction with The Exposition on Islamic Innovation (i-Inova) 2016. This was to encourage both participants and committees to get involved in innovation program-based organization. The keynote speaker for this seminar was The Honorable Prof. Emeritus Dr. Jalani Bin Sukaimi, who delivered an interesting talk on “*How to Complete Your Masters/PhD on Time?*” which has been highly appreciated by the participants.

KoSiST 16 has received 32 papers where each paper was peer reviewed by the appointed reviewers based on their expertise. This seminar proceeding volume contains 17 written research articles presented during the seminar. The articles cover the area of food biotechnology, pure and applied chemistry, pure and applied physics, pure and applied mathematics, theoretical and applied computer science as well as architecture. The remaining 15 papers were selected to be published in the Malaysian Journal of Science, Health and Technology (MJoSHT), the first edition journal from Faculty of Science and Technology.

We would like to thank all participants for their contributions to the seminar program and to this proceeding. Many thanks to the reviewers for their insightful comments on the papers submitted to KoSiST 2016. We are also indebted to those who have served as chairmen, judges and time keepers during the programme. A very special thanks to the organizing committees for their endless effort and commitment in making this seminar a success.

It is our pleasant duty to acknowledge the financial support from Faculty of Science and Technology, Universiti Sains Islam Malaysia.

Editors:

Farah Wahida Harun

Farida Ridzuan

Nurlida Basir

Yumn Suhaylah Yusoff

## ANALYSIS OF GELATIN DNA IN SUPPLEMENT CAPSULE SHELL USING REAL-TIME POLYMERASE CHAIN REACTION FOR HALAL AUTHENTICATION

Inayatullah, A.,<sup>1</sup> Hassan, M. S.<sup>2</sup> and Husaini, M. A. R.<sup>3</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1,2,3</sup>

Institute of Halal Research and Management, Negeri Sembilan, Malaysia<sup>2</sup>

ahlaminayatullah@gmail.com<sup>1</sup>, mohdsukri@usim.edu.my<sup>2</sup>, seinrahman182@gmail.com<sup>3</sup>

### ABSTRACT

*Halal authentication and verification in foodstuffs are highly required as the ingredients and sources of numerous products available in the market supplied to consumers are not clear. It is an obligation for Muslim to consume and use the halal consumer goods. At present, the most ideal technique to be used for sensitive detection of porcine and bovine DNA in gelatin is Polymerase Chain Reaction (PCR), due to the higher stability of DNA compared to protein. This study was conducted to identify the contamination of pigs in the supplement product that using gelatin as a shell wrapping of drug, with real-time PCR technique. Total of 20 samples supplement that have halal stamp from this and other countries, were collected and tested the authenticity of halal in capsules gelatin shell and was used SYBR green mastermix and specific primer bovine and porcine in the area of mitochondrial cytochrome B. The purpose of this study was to determine the specific animal that was used in gelatin capsule supplement. A total of 20 samples of supplements that have halal logo from different countries were tested using SYBR green mastermix together with specific primer for porcine (pig) and bovine (cow) to detect the present of porcine and bovine DNA in the area of mitochondrial cytochrome b. Six (6) samples were tested positive for porcine using specific primer porcine in real-time PCR, but only four (4) samples were tested positives when primer bovine was used in the same technique (real-time PCR).*

**Keywords:** Gelatin, supplement, real-time PCR, Halal.

### 1. INTRODUCTION

The detection of the pig contaminants becomes essential to ensure the safety and *halal*. Percentage production of gelatin in the world in 2006, is still dominated by pig gelatin, with 45.8% coming from pigt skin and 28.4% came from cowhide. Many studies have been done on the properties, applications and gelatin structure (given the breadth of gelatin uses). However gelatin from different sources can be very similar in terms of physical and chemical properties and it is difficult to distinguish between gelatin from *halal* and *haram* sources (pig) (Nemati et

al., 2004). The analysis based on chemical had been done to analyze the content of gelatin in order to detect of *halalness* with mass spectrometry (Zhang et al., 2009), Fourier transform infrared (FTIR) spectroscopy (Al-Saidi et al., 2012) and sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) that was combined with principal component analysis (PCA) (Nur Azira et al., 2012) and they was considered not sufficiently effective in detecting the source of gelatin from pigs and cows. Analysis of the soft capsule shell in commercial products has been carried out using high performance liquid chromatography (HPLC) based on the amino acid profile with ortho-phthalaldehyde derivatization method (OPA) - 2-mercaptoethanol (MCE) and chemometry technique. The analysis of amino acid profile with chemometry technique could classify soft capsule shell that was made from pig and cow. However, PCA could not classify the products from market. The disadvantage of these methods was caused by protein analysis, where the protein is unstable in heat and pH extreme (Widyaninggar et al., 2012).

Analysis using Real-Time PCR or quantitative PCR (qPCR) is based on a combination of traditional PCR detection, using "end-point" with fluorescent detection technology to record the accumulation of amplification at a time at each amplification cycle (Smith and Osborn, 2008). With this method, we can observe directly the result of amplification or reproduction of DNA fragments. This method can also determine the concentration of DNA contained in a sample by measuring the increase of fluoresce in fluorescence dyes when bound to the double-stranded DNA. Thus, analysis of *halal* gelatin with Real-Time PCR is a method that is simple, reliable and trustworthy (Cai et al., 2011).

In this study, the *halal* examination method in gelatin product was developed using Quantitative Polymerase Chain Reaction (Q-PCR) method with the assist of SYBR green as the fluorescent dye. The *halal* guarantee in food or beverage product is something that must be enforced to provide a sense of security and confidence to the consumer (especially Muslims). At the present *halalness* is not the priority in the pharmaceutical industries, because the company of pharmaceutical industry is still using pig as the raw material in capsule gelatin and the worst thing is they declare that the product is sources from *halal* material (cow) (having *halal* logo). In the next future, people will be more observant in address the *halal* product (food and beverage).

## 2. METHODOLOGY

The extraction and isolation of gelatin capsule used *Agilent porcine detection kits* that consists of nucleic acid binding buffer (25 ml); high salt wash buffer (24 ml), elution buffer (10 mM Tris-HCl, pH 7.5) (12 ml); DNA-binding spin cups and 2-ml receptacle tubes (50 each); 1.5-ml collection tubes (50 each); Proteinase K (2 x 0.5 ml); Proteinase K digestion buffer (2 x 5 ml). The DNA was analyzed at wavelengths of 260 nm and 280 nm. The result of instrument (DUO BioDrop) would be obtained was a DNA concentration data with ng/ $\mu$ l and data of purity DNA with ratio of A260 and A280 (BioDrop, 2012).

The samples were diluted and tested by real time PCR instrument (Applied Biosystems StepOnePlus) with bovine and porcine specific primer (Tanabe et al., 2007), and Mastermix Sso Advanced SYBR Green. All 0.2 ml PCR tubes were spin before placed in a 36-Well Rotor. Amplification was carried out at 95°C for 10 minutes as pre-denaturation stage to ensure all of double-stranded DNA become single-stranded DNA; (40 cycles) 95°C (denaturation step) for 10 seconds and 55°C for 45 seconds (annealing stage). Table 1 shows the formulation used for real time PCR.

**Table 1:** Formulation PCR mix for real time.

Component	Concentration ( $\mu$ M)	Volume ( $\mu$ l)
<i>Master mix</i>	-	10
Primer Forward	0.5	1
Primer Reverse	0.5	1
ddH <sub>2</sub> O	-	3
DNA template	-	5
Total Volume		20

## 3. ANALYSIS OF DATA

In detecting the presence or absence of porcine or bovine DNA in the capsule shell of supplement product, might be seen based on the results of amplification plot and melt curve analysis in real time PCR. If the DNA on a particular sample might be amplified with the primers porcine, it could be concluded that the gelatin on a hard or soft shell capsules are derived from pigs. Likewise, if a particular DNA sample might be amplified with the primers

bovine, it could be concluded that the gelatin on a hard or soft capsule shell is derived from cows. The amplification curve is generated by plotting the number of cycles horizontally and vertically fluorescence value. The curves are generated automatically by real time-PCR. From the curve it could be seen that the value of  $C_p$  (Crossing Point) of each DNA isolate (Vaerman et al., 2004).

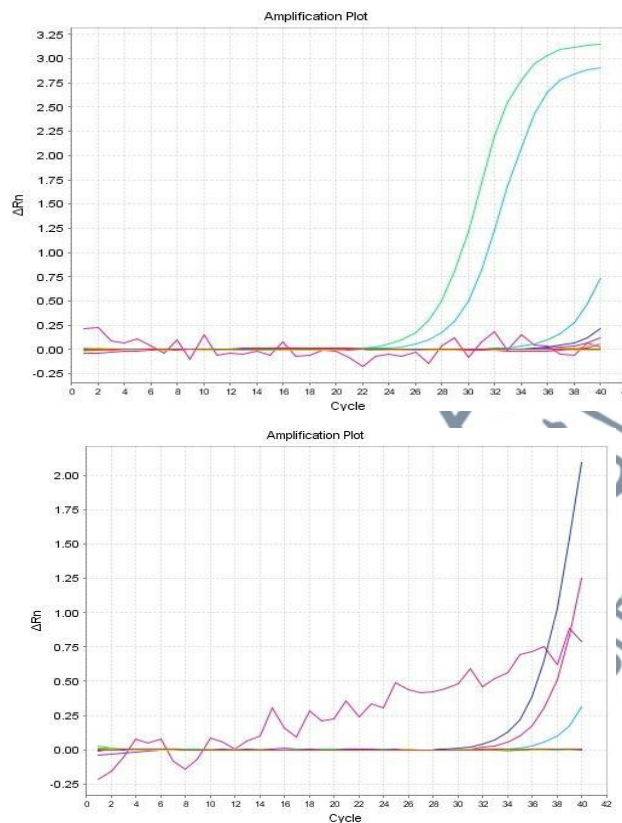
#### 4. RESULT AND DISCUSSION

The amplification of gelatin DNA in supplement had been used Real-Time PCR technique with SYBR Green method. The study used primer concentration, annealing temperature, the concentration of DNA template and the number of cycles that have been optimized in advance so it could obtain the accurate result (Izzah, 2014). Amplification of the isolate DNA samples using bovine primer is done as many as 40 cycles (Rochea, 2008). The increase of amplification curves is marking an increase of DNA concentration that was paired with a primer in PCR (Figure 1). The increase of curve was due to the enhancement of fluorescent in which was bound to double-stranded DNA. Fluorescence that was produced is proportional to the amount of template DNA that was amplified (Dooley et al., 2004).

Amplification product had been indicated that 10 samples were detected for bovine and porcine primer that based on the curve that had been formed. The tenth sample had melt peak curve with the temperature almost equally between one another ( $81^{\circ}\text{C}$ ). The curve of amplification plot increase is shown in three phases: the initial phase, exponential phase or peak and plateau phase or stable (Vaerman et al., 2004). The curve amplification would answer the problem statement whether the shell gelatin from supplement that had *halal* logo is free from porcine DNA or not. As in previous discussions, the increase in amplification curve indicates the positive result for sample.

From the results of the amplification curve is clearly visible only 10 samples had amplification curves log-linear (Table 2). The sample is S1, S2, S4, S5, S6, S12, S13, S14, S17 and S18. In sample S1, S4, S12 and S 18, the amplification curve forms from two types of primer (Bovine and Porcine). In sample S2, S13, S14 and S 17 only forms of bovine primer and sample S5 and S6 is only formed of porcine primer. From the  $C_p$  values that had been obtained, sample S1,

S4 and S12 of both primer tests, had concentration of bovine is higher than porcine DNA. This is because the Cp values of bovine primer were smaller than porcine primer. But, for sample S18, the concentration of porcine is higher than bovine DNA, because the Cp values of porcine primer were smaller than porcine primer.



**Figure 1:** Amplification plot of bovine and porcine primer.

Mechanical detection of animal origin, mainly in food products is very important because it associated with health, economics and religion. A safe and lawful source is a major concern, especially in Malaysia with a majority Muslim population. Methods of detection and identification of animal sources usually uses protein analysis but the method has a drawback and could only be done if the sample have form of raw material and requires quite a lot of samples, and the accuracy is very low in a state of overcooked sample (processed meat). The results of this study indicate that the use of specific primers pigs and cows that was designed by Tanabe et al. (2007) for the identification of the animal origin in supplement capsule shell

was proved to be sensitive to detect the fragments of pigs and cows only on samples that have the level of DNA concentration is 0.0001%.

**Table 2:** The data from spectroscopy UV/VIS and real-time PCR.

Sample	Concentration (ng/ $\mu$ L)	Purity (A260/A280)	Ct Bovine (Mean)	Ct Porcine (Mean)
S1	5.053	1.655	30.02	35.95
S2	2.876	1.533	36.98	-
S3	2.898	1.527	-	-
S4	3.924	1.342	34.11	34.81
S5	5.142	2.400	-	35.03
S6	6.954	1.404	-	37.11
S7	3.543	2.296	-	-
S8	3.499	1.460	-	-
S9	2.673	0.563	-	-
S10	3.234	0.324	-	-
S11	5.146	1.636	-	-
S12	4.872	1.696	22.64	31.36
S13	4.419	1.827	24.52	-
S14	4.664	1.751	35.75	-
S15	2.634	1.435	-	-
S16	1.982	1.576	-	-
S17	1.876	1.675	33.43	-
S18	3.245	1.542	37.08	34.33
S19	4.487	1.342	-	-
S20	3.278	1.309	-	-

## 5. CONCLUSION

SYBR green method as fluorescent dye and specific primer of bovine and porcine that was developed from *cytochrome b* gene could be used to identify the presence of porcine and the mixtures in gelatin capsule shells by using real-time PCR technique. Porcine and bovine primer that had been tested with the commercial supplement, four samples positive for bovine primer, two samples positive for porcine primer, four samples positive for bovine and porcine primer and 10 samples negative for porcine and bovine primer. The research funded by Universiti Sains Islam Malaysia through Research University Grant Scheme is gratefully acknowledged.

## REFERENCES

- Al-Saidi, G. S., Al-Alawi, A., Rahman, M. S., & Guizani, N. (2012). Fourier transform infrared (FTIR) spectroscopic study of extracted gelatin from shaari (*Lithrinus microdon*) skin: effects of extraction conditions. *International Food Research Journal*, 19, 1167-1173.
- BioDrop. (2012). Quick Start Guide [Online]. Available: <http://www.biodrop.co.uk>.
- Cai, H., Gu, X., Scanlan, M. S., Ramatlapeng, D. H., & Lively, C. R. (2012). Real-time PCR assays for detection and quantitation of porcine and bovine DNA in gelatin mixtures and gelatin capsules. *Journal of Food Composition and Analysis*, 25(1), 83-87.
- Dooley, J. J., Paine, K. E., Garrett, S. D., & Brown, H. M. (2004). Detection of meat species using TaqMan real-time PCR assays. *Meat science*, 68(3), 431-438.
- Fernando, N., Loke, S. W., & Rahayu, W. (2013). Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1), 84-106.
- Izzah, N.A. (2014). Perbandingan antara metode SYBR green dan metode hydrolysis probe dalam analisis DNA gelatin sapi dan DNA gelatin babi dengan menggunakan real time PCR," Bachelor. Thesis, Program Studi Farmasi., Fakultas Kedokteran dan Ilmu Kesehatan., Universitas Islam Negeri Syarif Hidayatullah., Jakarta.
- Nemati, M., Oveisi, M. R., Abdollahi, H., & Sabzevari, O. (2004). Differentiation of bovine and porcine gelatins using principal component analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 34(3), 485-492.
- Nur Azira, T., & Amin, I. (2012). Differentiation of bovine and porcine gelatins in processed products via sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and principal component analysis (PCA) techniques. *International Food Research Journal*, 19, 1175-1180.
- Rochea. (2008). The LightCycler® 480 Instrument Operator's Manual [Online]. Available: <http://www.roche-applied-science.com>.
- Smith, C. J. & Osborn, A. M. (2008). Advantages and Limitations of Quantitative PCR (Q-PCR)-based Approaches in Microbial Ecology. Federation of European Microbiological Societies., UK: Blackwell Publishing Ltd.
- Tanabe, S., Miyauchi, E., Muneshige, A., Mio, K., Sato, C., & Sato, M. (2007). PCR method of detecting pork in foods for verifying allergen labeling and for identifying hidden pork ingredients in processed foods. *Bioscience, Biotechnology, and Biochemistry*, 71(7), 1663-1667.
- Vaerman, J. L., Saussoy, P., & Ingargiola, I. (2004). Evaluation of real-time PCR data. *Journal of Biological Regulators and Homeostatic Agents*, 18(2), 212-214.

Widyaninggar, A., Triyana, K., & Rohman, A. (2012). Differentiation between porcine and bovine gelatin in capsule shells based on amino acid profiles and principal component analysis. Indonesian Journal of Pharmacy, 104-109.

Zhang, G., Liu, T., Wang, Q., Chen, L., Lei, J., Luo, J., Ma, G. & Su, Z. (2009). Mass spectrometric detection of marker peptides in tryptic digests of gelatin: a new method to differentiate between bovine and porcine gelatin. Food Hydrocolloids, 23(7), 2001-2007.



# COMPARATIVE ANALYSIS OF SURFACE ANTIGENS (SAG) PROTEINS IN THREE MAJOR EIMERIA SPECIES CAUSING COCCIDIOSIS IN DOMESTIC CHICKENS USING BIOINFORMATIC APPROACH

Manaf, Z.,<sup>1</sup> Ramly, N. Z.,<sup>2</sup> Wan, K. L.,<sup>3</sup> and Firdaus Raih, M.<sup>4</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1,2</sup>

National University of Malaysia (UKM), 43600 UKM Bangi, Selangor, Malaysia<sup>3,4</sup>

za52zm@gmail.com

## ABSTRACT

*Eimeria species are one of well-known member of phylum Apicomplexan that widely studied and cause severe intestinal disease coccidiosis. This infection has been suggested involve in host-parasite interactions including surface antigens (SAGs), Rhoptries (ROP) kinase and micronemes (MICs). This study uses bioinformatics approach to analyse SAGs sequence, to generate three-dimensional structures of SAGs and carried out comparative sequence and structure analysis of SAGs from three major Eimeria species. A total 89, 39 and 16 SAGs sequences were identified expressed in Eimeria tenella, Eimeria maxima and Eimeria acervulina, respectively. BLAST search was carried out for sequences similarity showed more than 40% residue identity of sequences belongs to SAG family member in Eimeria species only with no close homolog in PDB databases. Multiple sequence alignment was carried out using T-Coffee server and Jalview program for manual sequence alignment editing. Then, the phylogenetic tree was constructed for further analysis for evolution homology identification using PHYLIP 3.695. A total 26 SAGs sequence from Eimeria species were selected as representative of three subfamilies suggested presence in all Eimeria species for structure generation using SWISS-MODEL and CABS-fold server. PROCHECK 3.5.4 was used to validate and check the stereochemical quality of protein structures display in Ramachandran plot. More residues cluster at allowed region with high-resolution structure showed the good quality model produced. UCSF Chimera 1.9 was used for model visualization and analysis as well as for structural editing. Overall, Ramachandran plot result indicated the average percentage of favor region SAGs protein in Eimeria species was 73%.*

**Keywords:** Apicomplexa, Eimeria species, surface antigen, protein structure

## 1. INTRODUCTION

### 1.1 Recent Issues About Eimeria

*Eimeria*, *Toxoplasma*, *Plasmodium* and *Cryptosporidium* species are well-known member of phylum Apicomplexa that widely studied and cause several severe diseases to humankind and animals such as coccidiosis, toxoplasmosis and malaria (Frölich et. al., 2011). Apicomplexan parasites can be very difficult to remove once attached to the host cell. They can develop and manipulate host defense immunity to protect themselves from immune attack (Boulanger et. al., 2010). Seven *Eimeria* species include *Eimeria tenella*, *Eimeria necatrix*, *Eimeria acervulina*, *Eimeria maxima*, *Eimeria praecox*, *Eimeria brunetti* and *Eimeria mitis* infected specific region of intestinal organ in chickens and lead to serious bleeding, weight loss and even death (Barta et. al., 1997; Blake & Tomley, 2014). Within all these, seven *Eimeria*, *Eimeria tenella*, *Eimeria maxima* and *Eimeria acervulina* have been identified responsible in the most infection (Shirley et. al., 2004). Many studies have been suggested that the host-cell invasions involve the surface antigens (SAGs), Rhoptries (ROP) kinase and micronemes (MICs). SAGs and SAG-related sequence proteins (SRS) encoded single domain, membrane-bound proteins tethered by glycosylphosphatidylinositol (GPI)-anchors on *Eimeria tenella* and *Toxoplasma gondii* on surface of invasive sporozoites and merozoites stage, respectively (Tabarés et. al., 2004; Lekutis et. al., 2001; Jung et. al., 2004; Kasper et. al., 1984; Tomavo et. al., 1991). Range of GPI sequences in SAGs of *Eimeria tenella* are about 20 to 36 amino acids. There are three subfamilies of sag genes in *Eimeria* species which sagA for all species, sagB for *Eimeria tenella* and *Eimeria necatrix* whilst sagC for the other species include *Eimeria acervulina*, *Eimeria maxima*, *Eimeria praecox*, *Eimeria brunetti* and *Eimeria mitis*. The presence of the sagA in all *Eimeria* species suggested that sagA involve in core function of SAGs protein while the sagB and sagC may be involved in specific function for each species (Reid et. al., 2014). Total of SAGs protein identified so far varies for each species indirectly have different level of pathogenicity (Long et. al., 1976).

In the present study, we have used bioinformatics and molecular modeling approaches to analyse SAGs sequence, to generate three-dimensional structures of SAGs and carried out comparative sequence and structure analysis of SAGs from three major *Eimeria* species, which are *Eimeria tenella*, *Eimeria maxima* and *Eimeria acervulina*. This method can be the key to identify the entire family *Eimeria* species and give better understanding of the role of surface antigen in *Eimeria*.

## 2. METHODOLOGY

### 2.1 Retrieval of the Protein Sequence

All the protein sequences were retrieved from the European Nucleotide Archive: *Eimeria tenella* Houghton, *Eimeria maxima* Weybridge, *Eimeria acervulina* Houghton. A total 89, 39 and 16 SAGs sequences in *Eimeria tenella*, *Eimeria maxima* and *Eimeria acervulina*, respectively were opted. Average length for all SAGs sequence is 231 amino acids.

### 2.2 Sequence Similarity and 3-Dimensional (3D) Template Search

Each of SAG protein sequences was search against the National Center for Biotechnology Information (NCBI) protein database using protein blast (blastp) for sequence similarity and 3D template search. BLAST search showed more than 40% residue identity of sequences belongs to SAG family member in *Eimeria* species only with no close homolog in PDB databases.

### 2.3 Multiple Sequence Alignment and Phylogenetic Tree Construction

Multiple sequence alignment (Msa) and phylogenetic tree construction are fundamental tools in molecular biology. Sequence alignments of all SAG proteins were constructed using T-Coffee (Notredame et. al., 2000) and manually edited using Jalview version 2.9.Ob2 (Waterhouse et. al., 2009). Then, the neighbor-joining phylogenetic tree constructed using PHYLIP version 3.695 (Felsenstein & Churchill, 1996) to determine relationship between surface antigen (SAG) sequences within *E.tenella*, *E.acervulina* and *E.maxima*.

### 2.4 Three-Dimensional Structure Prediction

From the refined edited alignment and neighbor-joining tree constructed, a total 26 SAG proteins were selected as representative of three subfamilies that presence in all *Eimeria* species as stated earlier for structure generation using SWISS-MODEL and CABS-fold server.

Another reason for SAG proteins selection is because there are missing residues that supposed to be conserved across the family and to see how the missing residues affect the 3D structure. For SagA in *Eimeria tenella*, seven SAG proteins were selected from sixty members whilst five for SagB subfamilies from 28 members. As for *Eimeria acervulina*, three were selected from thirteen members as representative for SagA subfamily, one (eah\_44370) for SagB and two (eah\_59950, eah\_59200) for SagC. Last but not least, five (emwey\_59210, emwey\_59710, emwey\_52810, emwey\_52830, emwey\_5490, emwey\_43380) were selected for SagA in *Eimeria maxima* and two (emwey\_58360, emwey\_35570) members as representative for SagC subfamily. All selected SAG proteins have been summarized in the Table 1.

## 2.5 Model Validation and Visualization

PROCHECK 3.5.4 was used to validate and check the stereochemical quality of protein structures display in Ramachandran plot. More residues cluster at allowed region with high-resolution structure showed the good quality model produced. UCSF Chimera 1.9 was used for model visualization and analysis as well as for structural editing (Pettersen et. al., 2004).

**Table 1:** List of representative of each subfamily for modeling in three *Eimeria* species.

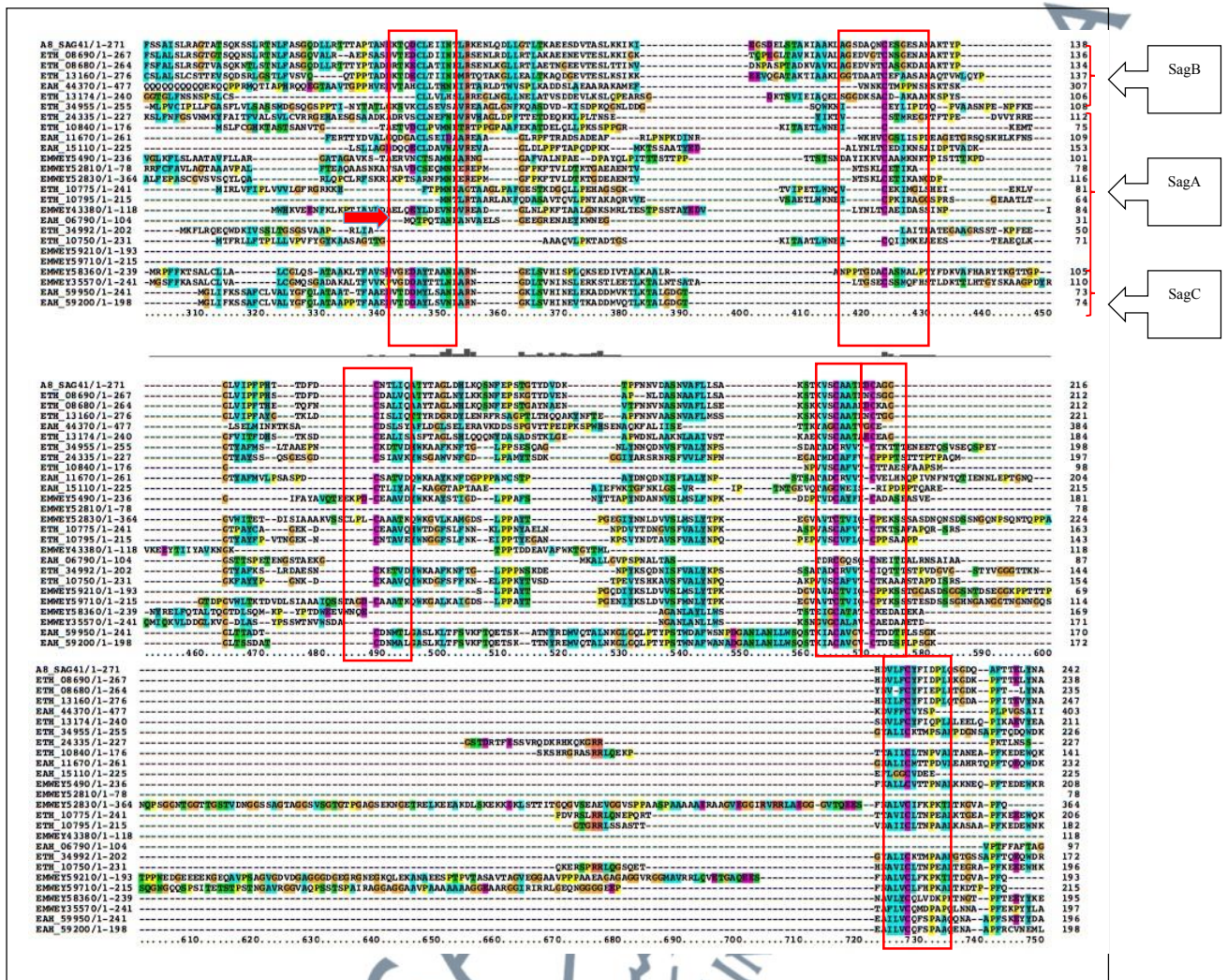
<i>Eimeria</i> species	SagA	SagB	SagC
<i>E.tenella</i>	eth_34955, eth_10775, eth_10795, eth_34992, eth_24335, eth_10840, eth_10750	eth_8680, eth_8690, eth_13174, eth_13160, SAG41	-
<i>E.maxima</i>	emwey_59210, emwey_59710, emwey_52810, emwey_52830, emwey_5490, emwey_43380	-	emwey_58360, emwey_35570
<i>E.acervulina</i>	eah_6790, eah_15110, eah_11670	eah_44370	eah_59950, eah_59200

### 3. RESULTS AND DISCUSSION

#### 3.1 Sequences Analysis

In this section, we evaluate the amount of cysteine residues presence in 26 sequences of *Eimeria* species, as cysteine is unique amino acid that can form disulfide bond. The cysteine residues in proteins play an important role in protein folding, stability and function (Hogg, 2003; Cheek et. al., 2006; Schmidt & Hogg 2007; Thangudu et. al., 2008; Marques et. al., 2010; Borloo et. al., 2013).

As mentioned previously, there are three subfamilies presence in *Eimeria* SAG protein. Recent genome analysis stated SagA and SagB consist six conserved cysteines whilst SagC consist of four conserved cysteines (Reid et. al., 2014). As shown in Figure 1, SagB and SagC subfamily consist all six and four conserved cysteines, respectively. Within all three subfamilies, only SagA subfamily has missing some of conserved cysteines indicated in red arrow as shown in Figure 1. However, the cause of missing cysteines still remains unknown. Hopefully, the models generated could be used to suggest any possibility and able to provide significance beneficial information.



**Figure 1:** Multiple sequences alignment of 26 selected SAG proteins from *Eimeria* species. Red boxes indicated all conserved cysteine residues located in the whole sequences. SagA, SagB and SagC consist 16, 6 and 4 SAG proteins as shown above.

### 3.2 Structure Validation and Visualization

In the absence of template available in PDB library, fold recognition and *de novo* approaches have been used in this study. In general, we analyzed each subfamily from every *Eimeria* species and summarized all the data shown in Table 2. Average for Ramachandran plot result for SagA, SagB and SagC subfamily was 72.7%, 86.0% and 77.9%, respectively. As in *Eimeria tenella*, the highest percentage of Ramachandran plot for SagA subfamily was eth\_10775 with 78.6% whilst SagB subfamily was SAG41 with 87.9%. Eth\_10775 model consist of 2 helices

with one missing cysteine residue whilst SAG41 model consist seven helixes and four strands with no missing cysteine residues. While, in *Eimeria maxima*, emwey\_5490 had the highest percentage of Ramachandran plot that is 76.4% for SagA subfamily and emwey\_58360 had 81.4% for SagC subfamily. Emwey\_5490 model consist 8 helixes whilst emwey\_58360 model consist 6 helixes and 2 strands. Then, in *Eimeria acervulina*, eah\_15110, eah\_44370 and eah\_59200 had 80.3%, 82.7% and 77.3% of Ramachandran plot result, respectively. As for eah\_15110, eah\_44370 and eah\_59200 model consist 3 helixes, 7 helixes and 4 strands, 5 helixes and 2 strands, respectively.

All SAG models in both *Eimeria maxima* and *Eimeria acervulina* above had no missing cysteine. Overall, Ramachandran plot for all models is 73.0%. This result can be the starting point to visualize 3-Dimensional structure of SAG proteins as well as to identify their specific function in host-invasion.

**Table 2:** Ramachandran plot result for all *Eimeria* with each subfamily.

<i>Eimeria</i> species	SagA	SagB	SagC
<i>Eimeria tenella</i>	Eth_10775 = 78.6 %	SAG41= 87.9%	-
<i>Eimeria maxima</i>	Emwey_5490 = 76.4%	-	Emwey_58360 = 81.4%
<i>Eimeria acervulina</i>	Eah_15110 = 80.3%	Eah_44370 = 82.7%	Eah_59200 = 77.3%
Average percentage of subfamily from all species (%)	72.7	86.0	77.9

#### 4. CONCLUSION

In summary, we have generated models and reviewed several SAG proteins from different *Eimeria* species. These models could be used to identify and predict their specific involvement in host-invasion. Although the conserved cysteine present in most of the SAG proteins, the contribution of these feature still remain unclear.

## ACKNOWLEDGMENT

The author would like to thank the Ministry of Higher Education of Malaysia for financial support and thank NZR, MFR and KLW for knowledge and advises support.

## REFERENCES

- Barta, J. R., Martin, D. S., Liberator, P. A., Dashkevicz, M., Anderson, J. W., Feighner, S. D., Elbrecht, A., Perkins-Barrow, A., Jenkins, M.C., Danforth, H.D. & Ruff, M. D. (1997). Phylogenetic relationships among eight *Eimeria* species infecting domestic fowl inferred using complete small subunit ribosomal DNA sequences. *The Journal of Parasitology*, 262-271.
- Blake, D. P., & Tomley, F. M. (2014). Securing poultry production from the ever-present *Eimeria* challenge. *Trends in Parasitology*, 30(1), 12-19.
- Boulanger, M. J., Tonkin, M. L., & Crawford, J. (2010). Apicomplexan parasite adhesins: novel strategies for targeting host cell carbohydrates. *Current Opinion in Structural Biology*, 20(5), 551-559.
- Borloo, J., Geldhof, P., Peelaers, I., Van Meulder, F., Ameloot, P., Callewaert, N., Vercruyse, J., Claerebout, E., Strelkov, S.V. & Weeks, S. D. (2013). Structure of *Ostertagia ostertagi* ASP-1: insights into disulfide-mediated cyclization and dimerization. *Acta Crystallographica Section D: Biological Crystallography*, 69(4), 493-503.
- Cheek, S., Krishna, S. S., & Grishin, N. V. (2006). Structural classification of small, disulfide-rich protein domains. *Journal of Molecular Biology*, 359(1), 215-237.
- Frölich, S., Entzeroth, R., & Wallach, M. (2011). Comparison of protective immune responses to apicomplexan parasites. *Journal of Parasitology Research*, 2012.
- Felsenstein, J., & Churchill, G. A. (1996). A Hidden Markov Model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution*, 13(1), 93-104.
- Hogg, P. J. (2003). Disulfide bonds as switches for protein function. *Trends in Biochemical Sciences*, 28(4), 210-214.
- Jung, C., Lee, C. Y. F., & Grigg, M. E. (2004). The SRS superfamily of *Toxoplasma* surface proteins. *International Journal for Parasitology*, 34(3), 285-296.
- Kasper, L. H., Bradley, M. S., & Pfefferkorn, E. R. (1984). Identification of stage-specific sporozoite antigens of *Toxoplasma gondii* by monoclonal antibodies. *The Journal of Immunology*, 132(1), 443-449.
- Lekutis, C., Ferguson, D. J., Grigg, M. E., Camps, M., & Boothroyd, J. C. (2001). Surface antigens of *Toxoplasma gondii*: variations on a theme. *International Journal for Parasitology*, 31(12), 1285-1292.

- Long, P. L., Millard, B. J., Joyner, L. P., & Norton, C. C. (1976). A guide to laboratory techniques used in the study and diagnosis of avian coccidiosis. *Folia Veterinaria Latina*, 6(3), 201-217.
- Marques, J. R. F., Da Fonseca, R. R., Drury, B., & Melo, A. (2010). Conformational characterization of disulfide bonds: A tool for protein classification. *Journal of Theoretical Biology*, 267(3), 388-395.
- Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302(1), 205-217.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612.
- Reid, A. J., Blake, D. P., Ansari, H. R., Billington, K., Browne, H. P., Bryant, J., Dunn, M., Hung, S.S., Kawahara, F., Miranda-Saavedra, D. & Malas, T. B. (2014). Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Research*, 24(10), 1676-1685.
- Schmidt, B., & Hogg, P. J. (2007). Search for allosteric disulfide bonds in NMR structures. *BMC Structural Biology*, 7(1), 49.
- Shirley, M. W., Ivens, A., Gruber, A., Madeira, A. M., Wan, K. L., Dear, P. H., & Tomley, F. M. (2004). The *Eimeria* genome projects: a sequence of events. *Trends in Parasitology*, 20(5), 199-201.
- Tabarés, E., Ferguson, D., Clark, J., Soon, P. E., Wan, K. L., & Tomley, F. (2004). *Eimeria tenella* sporozoites and merozoites differentially express glycosylphosphatidylinositol-anchored variant surface proteins. *Molecular and Biochemical Parasitology*, 135(1), 123-132.
- Thangudu, R. R., Manoharan, M., Srinivasan, N., Cadet, F., Sowdhamini, R., & Offmann, B. (2008). Analysis on conservation of disulphide bonds and their structural features in homologous protein domain families. *BMC Structural Biology*, 8(1), 55.
- Tomavo, S., Fortier, B., Soete, M., Ansel, C., Camus, D., & Dubremetz, J. F. (1991). Characterization of bradyzoite-specific antigens of *Toxoplasma gondii*. *Infection and Immunity*, 59(10), 3750-3753.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191.

## MORPHOLOGICAL AND MOLECULAR IDENTIFICATION OF BACTERIA ASSOCIATED WITH *HOLOTHURIA (MERTENSIOTHURIA) LEUCOSPILOTA* FROM PANGKOR ISLAND

Muhd Lutfi, F. N.<sup>1</sup>, Mohamed Rehan, M.<sup>2</sup>, Kamarudin, K. R.<sup>3</sup>, Che Zahri, M. Z.<sup>4</sup>, Mohd Noor, H.<sup>5</sup> and Md Radzi, S.<sup>6</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1,2,3,4,5,6</sup>

najihahlutfi@gmail.com

### ABSTRACT

*Holothuria (Mertensiothuria) leucospilota* is the most abundant sea cucumber species in Malaysia. Previous studies showed the presence of bacteria associated with this sea cucumber species from the marine environment. However, there is still a lack of information about the bacterial community associated with different body parts of the white threads fish. Thus, this study aimed to identify bacteria isolated from the external and internal body parts of *H. leucospilota*. Fresh specimens of *H. leucospilota* were collected from the coastal water of Pangkor Island, Perak, Malaysia. The bacterial samples were isolated using the streak plate method on Luria-Bertani agar. A total of 26 strains were successfully isolated from eight body parts of two specimens of *H. leucospilota* and from the surrounding sediments and seawater, with only 21 out of the total strains were included in subsequent analyses. Identification of the isolates were based on microscopic examination, 16S ribosomal DNA amplification and phylogenetic analysis using the neighbour-joining method. Three genera of bacteria were identified based on blastn analysis namely *Vibrio*, *Bacillus*, and *Acinetobacter*. Genus *Vibrio* was found to be the main bacterial group associated with the *H. leucospilota* specimens from Pangkor Island. On the basis of our results, of 21 bacterial isolates identified, 18 were from genus *Vibrio*. The different body parts showed higher number of bacterial isolates and species as compared to the environmental sources. In addition, three primary clusters were formed in the neighbor-joining tree consisting of family *Vibrionaceae*, family *Moraxellaceae* and family *Bacillaceae*.

**Keywords:** 16S rDNA sequencing, associated bacteria, *Holothuria leucospilota*, Pangkor Island, *Vibrio*.

### 1. INTRODUCTION

The *holothurians*, or sea cucumbers, can be found in great numbers from the marine environment throughout the world (Ridzwan, 2007). These soft-bodied animals often make up the majority of marine animal biomass. Sea cucumbers, in particular, are prized for its health beneficial values in traditional and modern medicine. They are also consumed as food by local

Malaysians. Besides benefiting human, sea cucumbers serve a useful role in the marine ecosystem when they live in close association with microbes. Sea cucumbers help to recycle nutrients, decompose detritus and other organic matters after which bacteria can continue the degradation process. Microbes, meiofauna, decaying organic debris, inorganic components, and dissolved organic matters that were accessible on surface sediments were ingested by sea cucumbers (Roberts et al., 2001; Gao et al., 2010).

Researchers reported significant roles played by the bacteria in the digestive tracts of many animal species. The bacteria aided in nutrient digestion and absorption, immunisation and biological antagonism of the hosts (Amaro et al., 2009; Hess et al., 2011; Amaro et al., 2012). Numerous marine animals live symbiotically with microbes and many of them host diverse populations of bacteria. The bacteria in the sediment are considered either as a direct food source or indirect source that offers the sea cucumbers with essential nutrients that are otherwise not available (Amaro et al., 2009; Gao et al., 2010).

In this study, *Holothuria (Mertensiothuria) leucospilota* was chosen as it was regarded as the most dominant sea cucumber species in the marine environment of Malaysia and seen to be present in most tropical localities all over the globe (Purcell et al., 2012). It is believed that the species, known as white threads fish in English and *bat puntil* among the Malaysians (Kamarudin et al., 2015), may contain indigenous microbes that help it to adapt and exist in the marine environment (Lukman et al., 2014). Previous study regarding microbes inhabiting the coelomic fluid of the sea cucumber species showed that more diverse microbial population present in *H. leucospilota* as compared to *Stichopus chloronotus* (locally known as *gamat* species) (Lukman et al., 2014). The members of family *Stichopodidae* are called *gamat* while the others including *H. leucospilota* is known as *timun laut* (Kamarudin et al., 2009; 2015). Environmental factors such as penetration of light, feeding behaviour, and level of antimicrobial properties in coelomic fluid of sea cucumbers are believed to influence the presence of microbial population (Lukman et al., 2014).

To date, despite the diversity and abundance of sea cucumber species, only a few comprehensive studies regarding association of microorganisms with Malaysian sea cucumbers were conducted. Little information about the bacterial community in the internal

body parts of the white threads fish were available. Therefore, the objective of this study was to identify bacteria associated with *H. leucospilota* from Pangkor Island, Perak, Malaysia. To address this issue, bacterial samples from eight body parts of the specimens, surrounding sediment and seawater were collected. Microscopic examination, 16S ribosomal DNA amplification and sequence analyses were used to characterise the bacterial community composition. A phylogenetic analysis using the distance-based neighbour-joining (NJ) method was also performed to examine the genetic relationships between the bacteria that associated with *H. leucospilota* from Pangkor Island, Perak, Malaysia.

## 2. MATERIALS AND METHOD

### 2.1 Specimen Collection and Bacterial Isolation

Two specimens of *H. leucospilota* were collected from Giam Island (Global Positioning System (GPS) position- 4° 14' 09.5"N 100° 32' 22.4"E) and Teluk Nipah Beach (GPS position- 4° 14' 03.3"N 100° 32' 41.4"E), Pangkor Island, Perak, Malaysia on March 2016. The specimens were labelled as HL and HL1, respectively. Non-protein-coding 12S and 16S mitochondrial ribosomal RNA (rRNA) gene sequences of the HL1 specimen were registered with the GenBank, National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine (accession no.: KX768273 - KX768274). The seawater temperature was 26°C during the sampling. The bacterial samples were isolated from eight body parts of the specimens i.e. cuticle, coelomic fluid, cloaca, gastrointestinal, respiratory tree, tentacle, cuvierian tubules, and polian vesicles, and from the surrounding sediments and seawater. Sterile cotton swabs were used for isolating the bacterial samples from the different body parts, whereas surrounding sediments and seawater samples were taken and placed in Falcon tubes. The samples were streaked across Luria-Bertani (LB) agar (Merck, Germany) in a zig-zag pattern. After overnight incubation at 26°C, bacterial colonies with different morphologies were restreaked onto new LB agar until pure cultures or single bacterial colonies were obtained. The cultures were also streaked on nutrient agar (NA) media (Merck, Germany). After Gram staining, the single colonies were observed under compound microscope with magnifications of 1000x and oil immersion.

## 2.2 DNA Extraction and Polymerase Chain Reaction (PCR)

A pure bacterial colony from each LB plate was scraped using a sterile micropipette tip and transferred into a 200  $\mu$ L PCR tube containing 20  $\mu$ L of 0.02 M NaOH. The mixture was incubated at 95°C for 10 min. Subsequently, the sample was placed on ice for 5 min, followed by centrifugation (Eppendorf, Germany) at maximum speed i.e. 13400 rpm for 2 min. The supernatant was used as a template DNA for the PCR. The universal primer set, PB36 Forward: 5'- AGAGTTTGATCCTGGCTCAG -3' and PB38 Reverse: 5'- GGTACCTTGTTACGACTT -3', were used to amplify the 16S rDNA region. All PCR reactions were conducted in 25  $\mu$ L volume containing 12.5  $\mu$ L of Promega PCR Master Mix, 1.3  $\mu$ L of each universal primer (10  $\mu$ M), 7.9  $\mu$ L of sterilised dH<sub>2</sub>O, and 2  $\mu$ L of template DNA. The thermocycling steps were as follows: 95°C for 2 min, followed by 35 cycles at 95°C for 40 sec, 55°C for 40 sec, 72°C for 90 sec, and a final extension step at 72°C for 10 min. All PCR products were analysed on 1% (w/v) agarose gel stained with FloroSafe DNA Stain, and 1 kb DNA ladder was included as the standard DNA marker. Unpurified PCR products were sent for PCR clean up (i.e. PCR product purification) and DNA sequencing at the First BASE Laboratories Sdn Bhd, Seri Kembangan, Selangor, Malaysia.

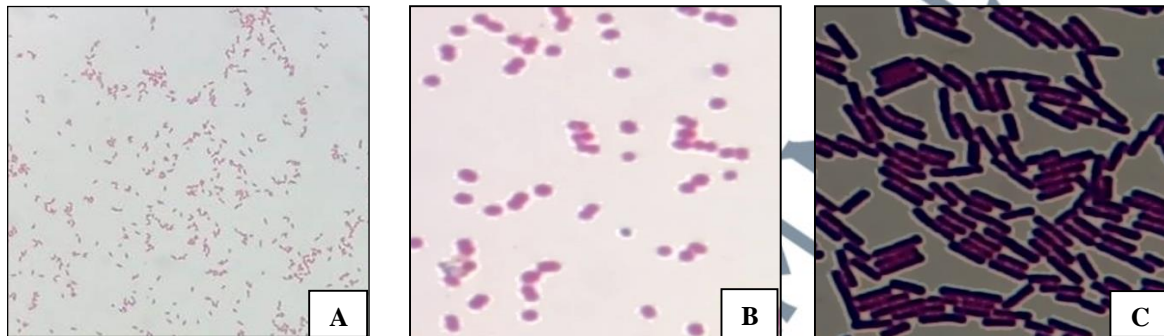
## 2.3 16S rDNA Sequence Analysis

Once the DNA sequencing results were obtained, Basic Local Alignment Search Tool program (BLAST) was used to detect the sequence similarity in a non-redundant sequence database (Altschul et al., 1997). Multiple alignments of the DNA sequences were done using the ClustalX Version 1.81 and following that, the alignments were checked by naked eyes. A phylogenetic tree based on the distance-based neighbour-joining (NJ) method was constructed using the Molecular Evolutionary Genetics Analysis version 6.0 (MEGA6) software (Tamura et al., 2013).

# 3. RESULTS AND DISCUSSION

## 3.1 Morphological Characteristics of Bacterial Isolates

Bacterial isolates were Gram-stained to help in confirmation of their bacterial groups or classes (Lukman et al., 2014). Gram-negative bacteria which belonged to the genera *Vibrio* and *Acinetobacter* made up to 96% of the bacterial isolates, while *Bacillus* was the only representative for Gram-positive bacterial group (4%). Under the microscopic examination, the representative species were observed to be short curved rods for *Vibrio* sp., round and double forms for *Acinetobacter* sp., and long curved rod for *Bacillus* sp. (Fig. 1).



**Figure 1:** Morphological characteristics of representative strain of every genus as observed under the compound microscope. (A) *Vibrio*, (B) *Acinetobacter*, and (C) *Bacillus*.

### 3.2 Genetic Characterization

A total of 26 PCR products of 16S rDNA region representing 26 isolates, each with a fragment length of approximately 1500 base pairs (bp), were successfully obtained. In terms of DNA sequencing, only 21 out of the 26 PCR products showed good DNA sequencing results. Five 16S rDNA sequences showed bad sequencing results due to the presence of noisy data in their nucleotide sequences.

Twelve isolates were successfully identified by the BLAST for nucleotide (blastn) as *Vibrio parahaemolyticus*, five isolates as *Vibrio harveyi*, two isolates as *Acinetobacter schindleri*, one isolate as *Vibrio maritimus* and another one isolate as *Bacillus thuringiensis* (Table 1). Three different bacterial species were isolated from various body parts of HL specimen and two bacterial species were isolated from its surrounding sediment and seawater, whereby *V. maritimus* was not isolated from the environment sources. On the other hand, four bacterial species were isolated from various body parts of HL1 specimen and two bacterial species were isolated from its surrounding sediment and seawater, whereby only *A. schindleri* and *V. parahaemolyticus* were isolated from both the body parts and the environment sources. *V.*

*parahaemolyticus* was isolated from both the surrounding sediments of HL specimen and HL1 specimen. Furthermore, *V. harveyi* and *V. parahaemolyticus*, were isolated from both HL specimen and HL1 specimen. *V. maritimus* was present in HL specimen but absent in HL1 specimen; and *A. schindleri* and *B. thuringiensis* were present in HL1 specimen but absent in HL specimen. The presence of *A. schindleri* in HL1 specimen was due to the presence of the bacterial species in the surrounding seawater. The results suggested that more bacterial species were found in different body parts of the sea cucumber specimens as compared to their environment sources i.e. the surrounding sediments and seawater.

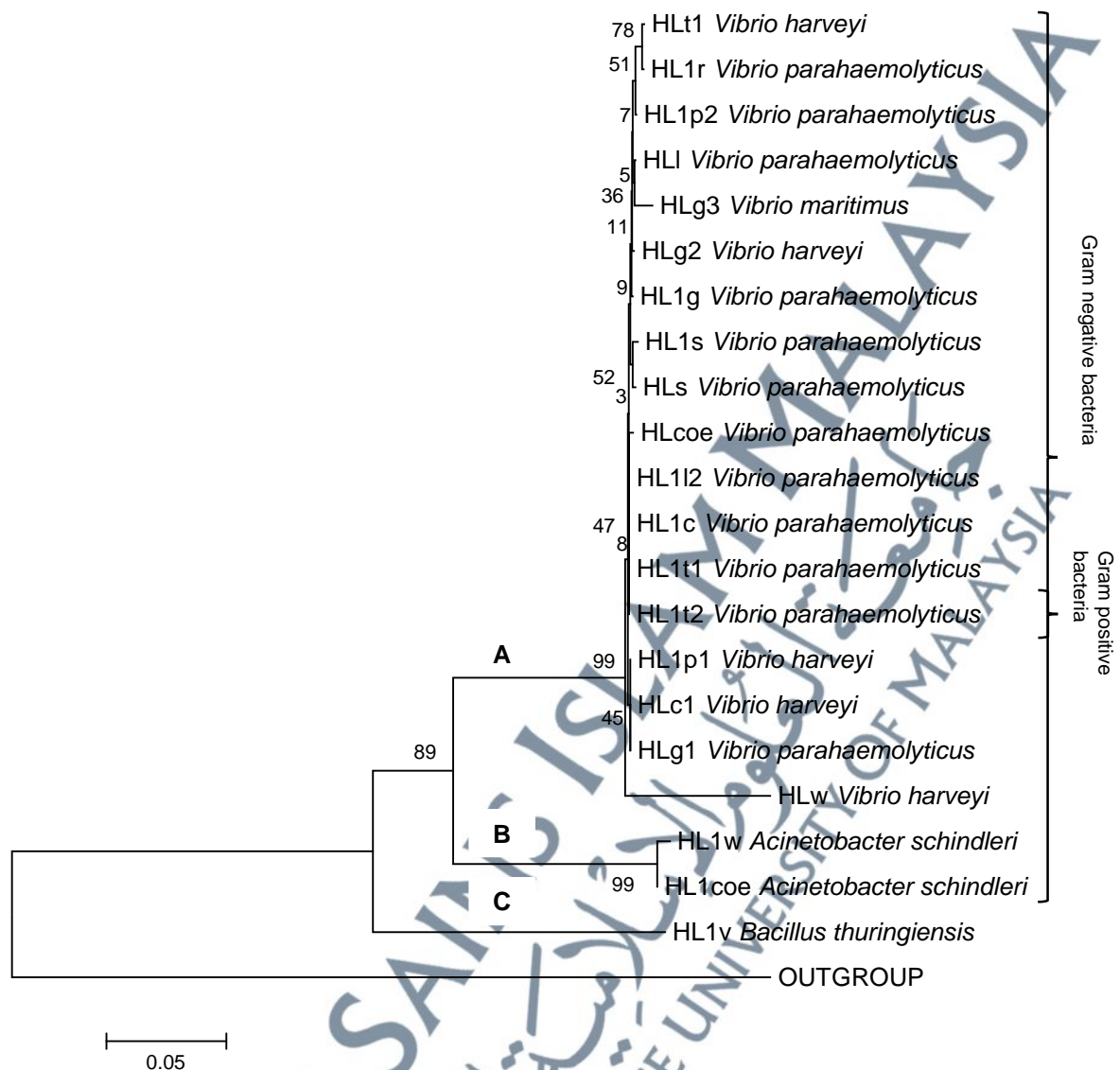
In total, 22 partial sequences of 16S rDNA region were included in the phylogenetic analysis including a 16S rDNA sequence of *Methanosarcina barkeri* (GenBank accession no.: CP009526.1) as the outgroup to root the distance-based NJ tree. There were 625 nucleotide bases in the final dataset after the multiple alignment. A phylogenetic tree constructed using the NJ method (Saitou and Nei, 1987) was summarized in Figure 2. The optimal NJ tree with the sum of branch length = 0.87157370 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (i.e. 1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980) and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd. All positions containing gaps and missing data were eliminated.

Three primary clusters are shown in Fig. 2. Cluster A consisted of family Vibrionaceae, cluster B represented family Moraxellaceae, and cluster C comprised family Bacillaceae. 16S rDNA sequences of *V. parahaemolyticus*, *V. harveyi*, and *V. maritimus* were grouped together in cluster A with 99% bootstrap value. Meanwhile, cluster B represented *A. schindleri* species, and the cluster was grouped together with cluster A with 89% bootstrap support. Furthermore, cluster C comprised the representative 16S rDNA sequence of genus *Bacillus*, *B. thuringiensis*. Moreover, the NJ tree (Fig. 2) also shows two main clusters of Gram-negative bacteria (the clusters A and B) and the Gram-positive bacterium (cluster C).

**Table 1:** Identification of bacterial isolates associated with the *Holothuria (Merthensiothuria) leucospilota* specimens from Pangkor Island, Perak, Malaysia based on BLAST analyses of 16S rDNA sequences.

Note: ID – based on Identities score (Ident)

Bacterial isolates	Sources of isolates	Probable species	% ID
<u><i>HL specimen</i></u>			
HLw	Water	<i>Vibrio harveyi</i>	92%
HLs	Sediment	<i>Vibrio parahaemolyticus</i>	99%
HLc1	Cuticle	<i>Vibrio harveyi</i>	99%
HLcoe	Coelomic fluid	<i>Vibrio parahaemolyticus</i>	99%
HLl	Cloaca	<i>Vibrio parahaemolyticus</i>	99%
HLg1	Gastrointestine	<i>Vibrio parahaemolyticus</i>	99%
HLg2	Gastrointestine	<i>Vibrio harveyi</i>	99%
HLg3	Gastrointestine	<i>Vibrio maritimus</i>	99%
HLt1	Tentacle	<i>Vibrio harveyi</i>	99%
<u><i>HLI specimen</i></u>			
HL1w	Water	<i>Acinetobacter schindleri</i>	99%
HL1s	Sediment	<i>Vibrio parahaemolyticus</i>	99%
HL1c	Cuticle	<i>Vibrio parahaemolyticus</i>	99%
HL1coe	Coelomic fluid	<i>Acinetobacter schindleri</i>	99%
HL1v	Cuverian tubules	<i>Bacillus thuringiensis</i>	99%
HL1l2	Cloaca	<i>Vibrio parahaemolyticus</i>	99%
HL1g	Gastrointestine	<i>Vibrio parahaemolyticus</i>	99%
HL1r	Respiratory tree	<i>Vibrio parahaemolyticus</i>	99%
HL1t1	Tentacle	<i>Vibrio parahaemolyticus</i>	99%
HL1t2	Tentacle	<i>Vibrio parahaemolyticus</i>	99%
HL1p1	Polian vesicle	<i>Vibrio harveyi</i>	99%
HL1p2	Polian vesicle	<i>Vibrio parahaemolyticus</i>	99%



**Figure 2:** Neighbour-joining (NJ) tree of 21 bacteria isolates associated with the *Holothuria (Mertensiothuria) leucospilota* specimens from Pangkor Island, Perak, Malaysia inferred from 16S rDNA gene sequences. Note: OUTGROUP - 16S rDNA sequence of *Methanosarcina barkeri* (GenBank accession no.: CP009526.1). Cluster A - family *Vibrionaceae*, cluster B - family *Moraxellaceae*, and cluster C - family *Bacillaceae*.

Genus *Vibrio* was identified to be the main bacterial group associated with the *H. leucospilota* specimens from Pangkor Island. According to other studies, *Vibrio* presence in association with sea cucumber species have also been found (Lukman et al., 2014). *Vibrio* bacteria were commonly found in the marine, estuarine, and freshwater environment worldwide (Romalde et al., 2014). *V. parahaemolyticus* in particular, a rod-shaped Gram-negative bacterium, formed the major bacterial strain in this study with percentage of 57%. Sediments, plankton, fish, crustaceans, and bivalve molluscs were the places that this bacterial species was frequently

found (Iida et al., 2006). Iida et al. (2006) further suggested that water temperature and salinity played important roles in the growth of this species. During the summer season in Europe and United States, as the temperature raised around 25°C and above, this species was frequently found; and interestingly, in the Southeast Asia the *Vibrio* species could be found all year round (Zulkifli et al., 2009). In Malaysia, a high marine temperature between 25 and 35°C favor the outbreaks of *V. parahaemolyticus* (Zulkifli et al., 2009).

*V. harveyi*, the second major bacterial strain associated with the *H. leucospilota* specimens from Pangkor Island, was extensively investigated as luminescent bacterium (Hui and Sherkat, 2005). A symbiotic relationship existed between this bacterium and its host organism, where a nutrient rich environment was provided for the growth of the bacterium by the host and as a return, the luminescent properties provided protection for the host from its predator (Lin et al., 2001). However, there was little information about *V. maritimus*. This species was isolated from *Palythoa caribaeorum* a coral species in Brazil (Chimetto et al., 2011). Whereas for genera *Acinetobacter* and *Bacillus*, these bacterial strains were found to be associated with sea cucumbers (Yasoda et al., 2006, Gao et al., 2014).

For future research, a higher number of *H. leucospilota* specimen from different locations all over Malaysia is required in order to obtain more reliable results on bacterial communities in a specific population. In addition, different molecular techniques such as Total Genomic DNA-Restriction Fragment Length Polymorphism (TgDNA-RFLP) and PCR-RFLP can also be used for better species identification and validation of bacteria associated with *H. leucospilota*.

#### 4. CONCLUSION

In summary, a higher number of bacterial isolates and species from different body parts of *H. leucospilota* specimens from Pangkor Island, Perak, Malaysia were observed as compared to their environmental sources. The blastn and phylogenetic analysis of 16S rDNA sequences showed that three genera of bacteria namely *Vibrio*, *Acinetobacter*, and *Bacillus* that can be further classified under three families i.e. family Vibrionaceae, family Moraxellaceae, and family Bacillaceae were identified from 21 bacterial isolates associated with the *H. leucospilota* specimens.

## ACKNOWLEDGEMENT

We are grateful to all reviewers of this paper; and all members of Food Biotechnology program, Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia (USIM), Nilai, Negeri Sembilan. Special thanks to Ms. Sarina Irma Binti Saidon, Mr. Muhammad Izzat B. Redzuan, and Mrs. Siti Nabilah Bt Mohd Rusly for all the technical assistance. This research was funded by the Fundamental Research Grant Scheme (FRGS) Phase 1/2015 from the Department of Higher Education, MOE – Ref: USIM/FRGS/FST/32/51515.

## REFERENCES

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389-3402.
- Amaro, T., Luna, G. M., Danovaro, R., Billett, D. S., & Cunha, M. R. (2012). High prokaryotic biodiversity associated with gut contents of the *holothurian* *Molpadia musculus* from the Nazaré Canyon (NE Atlantic). *Deep Sea Research Part I: Oceanographic Research Papers*, 63, 82-90.
- Amaro, T., Witte, H., Herndl, G. J., Cunha, M. R., & Billett, D. S. (2009). Deep-sea bacterial communities in sediments and guts of deposit-feeding holothurians in Portuguese canyons (NE Atlantic). *Deep Sea Research Part I: Oceanographic Research Papers*, 56(10), 1834-1843.
- Chimetto, L. A., Cleenwerck, I., Moreira, A. P. B., Brocchi, M., Willems, A., De Vos, P., & Thompson, F. L. (2011). *Vibrio variabilis* sp. nov. and *Vibrio maritimus* sp. nov., isolated from *Palythoa caribaeorum*. *International Journal of Systematic and Evolutionary Microbiology*, 61(12), 3009-3015.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4), 783-791.
- Gao, F., Sun, H., Xu, Q., Tan, J., Yan, J. & Wang, Q. (2010) PCR-DGGE analysis of bacterial community composition in the gut contents of *Apostichopus japonicus*. *Journal of Fishery Sciences of China*, 17: 672–680.
- Gao, F., Tan, J., Sun, H., & Yan, J. (2014). Bacterial diversity of gut content in sea cucumber (*Apostichopus japonicus*) and its habitat surface sediment. *Journal of Ocean University of China*, 13(2), 303-310.
- Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., Luo, S., Clark, D. S., Chen, F., Zhang, T., Mackie, R. I., Pennacchio, L. A., Tringe, S. G., Visel, A., Woyke, T.,

- Wang, Z. & Rubin, E. M. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331(6016), 463-467.
- Hui, Y. H. & Sherkat, F. (2005). *Handbook of Food Science, Technology and Engineering*, Volume 4. London, New York: CRC Press.
- Iida, T., Park, K. S. & Honda, T. (2006). *Vibrio parahaemolyticus*. *The Biology of Virus*, pp. 340-348.
- Kamarudin, K. R., Mohamed Rehan, A., Lukman, A. L., Ahmad, H. F., Anua, M., Nordin, N. F. H., Hashim, R., Hussin, R. & Usup, G. (2009). Coral reef sea cucumbers in Malaysia. *Malaysian Journal of Science*, 28(2), 171-186.
- Kamarudin, K. R., Usup, G., Hashim, R., & Rehan, M. M. (2015). Sea Cucumber (*Echinodermata: Holothuroidea*) Species Richness at Selected Localities in Malaysia. *Pertanika Journal of Tropical Agricultural Science*, 38(1), 7-32.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2), 111-120.
- Lin, L. Y. C., Sulea, T., Szittner, R., Vassilyev, V., Purisima, E. O., & Meighen, E. A. (2001). Modeling of the bacterial luciferase-flavin mononucleotide complex combining flexible docking with structure-activity data. *Protein Science*, 10(8), 1563-1571.
- Lukman, A. L., Nordin, N. F. H., & Kamarudin, K. R. (2014). Microbial population in the coelomic fluid of *Stichopus chloronotus* and *Holothuria (Mertensiothuria) leucospilota* collected from Malaysian waters. *Sains Malaysiana*, 43(7), 1013-1021.
- Purcell, S., Samyn, Y. & Conand, C. (2012). *Comercially Important Sea Cucumbers of the World*. FAO Species Catalogue, Italy.
- Ridzwan, B. H. (2007). *Sea cucumber: the Malaysian heritage*. Malaysia: Research Management Centre, International Islamic University Malaysia (IIUM).
- Roberts, D., Moore, H. M., Berges, J., Patching, J. W., Carton, M. W., & Eardly, D. F. (2001). Sediment distribution, hydrolytic enzyme profiles and bacterial activities in the guts of *Oneirophanta mutabilis*, *Psychropotes longicauda* and *Pseudostichopus villosus*: what do they tell us about digestive strategies of abyssal holothurians? *Progress in Oceanography*, 50(1), 443-458.
- Romalde, J. L., Diéguez, A. L., Lasa, A., & Balboa, S. (2013). New *Vibrio* species associated to molluscan microbiota: A review. *Frontiers in Microbiology*, 4, 413-424.
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725-2729.

Yasoda, H. N., Chi, Z., & Zhu, K. (2006). Probiotics and sea cucumber farming. *SPC Bech Info Bull*, 24, 45-8.

Zulkifli, Y., Alitheen, N. B., Son, R., Yeap, S. K., Lesley, M. B., & Raha, A. R. (2009). Identification of *Vibrio parahaemolyticus* isolates by PCR targeted to the *toxR* gene and detection of virulence genes. *International Food Research Journal*, 16, 289-296.



## STRUCTURAL STUDY OF PROTEASES FROM *BACILLUS LEHENSIS* G1

Sulaiman, N. A.<sup>1</sup> and Ramly, N. Z.<sup>2</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1,2</sup>

noorulaini@gmail.com

### ABSTRACT

*Protease is an enzyme that catalyses the hydrolysis of peptide bond in polypeptide chain and possess a wide range of applications in industry. The aim of this study is to clone and express several genes encoding proteases from alkalitolerant bacteria Bacillus lehensis strain G1. A total of 13 genes encoding proteases have been selected using bioinformatics approach and they were further processed to remove the signal peptide region. These genes were then amplified using polymerase chain reaction (PCR) method. Then the PCR product was cloned into cloning vector pGEM<sup>®</sup>T and was subsequently transformed into competent cell E. coli DH5a. The transformants were further verified by sequencing. The positive cloned were subcloned into the expression vector and were then expressed in Luria Bertani medium in the present of IPTG using E. coli BL21. The expressions of recombinant proteases were optimized for several hours at different temperatures, 16-37 °C. Furthermore, structural prediction has been carried using Modeller v9.16. Each generated model was verified for overall completeness and bias, using PROCHECK, ERRAT, and Verify 3D. The overall quality of the model was relatively good with percentage of Ramachandran plot is 96.3%, PROCHECK is 86.2% and ERRAT score is 95%, respectively.*

**Keywords:** *Bacillus lehensis, protease, recombinant enzyme, homology modelling.*

### 1. INTRODUCTION

Proteases are an enzyme that cleaves other proteins or themselves and carries diverse functions from protein recycling, nutrient digestion, and immune system. By understanding these biological roles; proteases have been manipulated to be used in industry, therapeutics and academia (Li et al., 2013). There are several methods have been used to improve the proteases yield and activity such as optimization in fermentation, molecular recombinant and site directed mutagenesis (Gupta et al., 2002). A significant amount of studies have been carried out on the application of microbial protease in detergent, food, pharmaceutical, leather, dehairing agent, biowaste disposal, silk degumming, and silver recovery (Haddar et al., 2010a; Bajaj et al.,

2013; Sundararajan et al., 2011; Lagzian & Asoodeh, 2012; Sinha & Khare, 2013; Joshi & Satyanarayana, 2013; Wang & Yeh, 2006).

Proteolytic enzymes from alkaliphile bacteria are commonly used in many industries, thus finding a new alkaliphile bacteria strain with high catalytic activity and thermostability is always essential, valuable and useful yet challenging. Proteases is found from a number of sources, such as animals, bacteria, fungi, and plant; though, in general, microbial proteases are widely in used. Among the bacteria, *Bacillus* species strains are often used as source of proteases due to high amount of active enzyme with pH, temperature and solvent stabilities; such as *B. korieensis*, *B. circulans*, *B. mojavensis* and *B. pumilus* (Anbu, 2013; Benkiar et al., 2013; Haddar et al., 2010b; Bajaj & Jamwal, 2013).

*Bacillus lehensis* is a Gram-positive, endospore-forming, with the ability to grow under extreme conditions as the bacterium are predominantly isolated from natural alkaline environments, which include soda lakes, deserts and arid soils (Blanco et al., 2012; Ghosh et al., 2007). Recently, genome of a novel isolated *B. lehensis* G1 from rubber plantation in Johor, Malaysia (Illias et al., 2002), has been successfully sequenced. This may offer a platform for their protein to be extensively study through bioinformatics' process. *B. lehensis* G1 genome contain 4017 protein-coding sequences with approximately 70% assigned biological functions (Noor at el., 2014). The availability of this bacterial genome may facilitate researcher to find new potential area to explore, as there are few articles published on *B. lehensis* G1 such as cyclodextrin glucanotransferase (Sian et al., 2005; Illias et al., 2002; Jonet et al., 2012), membrane protein (Liew et al., 2007), signal peptide (Low et al., 2012; Low et al., 2013) and Sco protein (Tan et al., 2014). In this present study, we will describe ongoing works on recombinant proteases isolated from *B. lehensis* G1 together with the molecular structure prediction that has never been reported. This research could provide new opportunities for the discovery, production and modification of novel protease to be use in highly alkaline conditions.

## 2. MATERIALS AND METHOD

### 2.1 *Bioinformatics Selection*

The amino acids from *B. lehensis* genome (NCBI Accession no. CP003923) were multiple aligned with the Uniprot and MEROPS with the purpose of finding potential proteases. The presence of a putative secretion signal peptide at N-terminus is predicted using SignalP 4.0 server. Whilst transmembrane are calculated using TMHMM server v 2.0 with coiled coils and low complexity are predicted using DisEMBL and SMART.embl-heidelberg server, respectively.

### 2.2 *Strains, Media And Vectors*

DNA genome was extracted from *B. lehensis* G1 using Wizard<sup>®</sup> Genomic DNA Purification Kit (Promega, USA). The genome was used as a template for amplification of alkaline protease genes. *Escherichia coli* DH5 $\alpha$  was used for the construction and propagation of vector molecules. Whilst *E. coli* BL21 (DE3), Origami 2 (DE3) and Rosetta-gami<sup>™</sup>2 (DE3)(Novagen, USA) were used as expression host in Luria Bertani (LB) medium (10 g bacto-tryptone, 5 g bacto-yeast extract, and 10 g sodium chloride in 1L). LB medium supplemented with Ampicillin (100  $\mu$ g/ml) and kanamycin (100  $\mu$ g/ml) was used to screen the *E. coli* transformants clones. Vector pGEM<sup>®</sup>-T Easy (Promega, USA) was used for cloning whilst pET28b(+) (Novagen, USA), pET32 Ek/LIC (Novagen, USA) and pET200 (Invitrogen, USA) were used as expression vectors.

### 2.3 *DNA Manipulation and Protein Expression*

Standard recombinant DNA routines were carried out according to the methods described by Sambrook et al. (1989). Polymerase Chain Reaction (PCR) amplification genes were carried out using gene-specific primers with the use of *pfu* DNA polymerase (Promega, USA) according to the manufacturer's instructions. Restriction sites *Nde*1 and *Xho*1 were added at the N-terminus and at the C-terminus, respectively, to ensure correct orientation with the expression vector. The cloning vectors were later digested, purified (QIAquick gel purification

kit, Qiagen) and ligated into the multiple cloning site of pET28b(+). Meanwhile, PCR amplicon were directly cloned into pET200 and pET32 Ek/LIC according to the manufacturer's instruction. *B. lehensis* G1 recombinant proteins were expressed in LB medium at different temperatures (16-37°C), 150-250 rpm and at different concentration of IPTG (Isopropyl β-D-1-thiogalactopyranoside) (0.01-1mM) induction for four hours, six hours and overnight. Extracted protein was obtained using BugBuster Protein Extraction Reagent (Novagen CA, USA) and sonication.

#### 2.4 Polyacrylamide Gel Electrophoresis and Western Blot

Protein concentration was measured using the Quick Start™ Bradford Protein Assay kit according to the manufacturer's instructions (BioRad, USA). The molecular mass of the recombinant proteases were verified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) according to Laemmli method. Protein bands were visualized via staining with Coomassie brilliant blue R-250 and further destained in 20% methanol in 10% acetic acid solution. Western blot was carried out to verify the correct protein bands that were expressed together with His tag protein (Mahmood & Yang, 2012).

#### 2.5 Homology Modelling

Homology modelling was used to construct the tertiary structure of one target protease. Protein Data Bank (PDB) template was selected based on sequence similarity using SWISS-MODEL workspace server and three dimensional structures was built using MODELLER (version 9.16) with PyMOL software for viewing the generated model. Quality of the predicted model were assess with Verify 3D (Bowie et al., 1991), PROCHECK (Laskowski, 1993) and ERRAT (Colovos & Yeates, 1993).

### 3. RESULTS AND DISCUSSION

To date, there are 2818 proteins with assigned biological roles in *B. lehensis* G1 (Noor et al., 2014), however, after alignment with proteases from Uniprot and MEROPS database, there are approximately 300 protein candidates available. Following that, only 13 genes were selected

after the removal of predicted transmembranes proteins. Furthermore, prediction of signal sequence using SignalP 4.0 server showed two genes; BleG1\_3207 and BleG1\_1058 contained signal peptide. From 13 genes of *B. lehensis* G1, only 12 genes were successfully amplified. All proteases with sizes ranging from 525 to 1137 bp with the start and a stop codon at the termini were cloned in pGEM<sup>®</sup>-T Easy vector. BLAST analysis of the pGEM<sup>®</sup>-T Easy vector sequences revealed that all proteases have 100% matches with genome database *B. lehensis* G1. Table 1 shows the nucleotide sequences, estimate the molecular mass and pI value using Expsy web, NCBI accession number and conserved domain database results. The clones were later digested with *Nde*IF and *Xho*IR, and subsequently ligated into pET28b (+). Protease in pET28b (+) vector was initially transformed into *E. coli* DH5 $\alpha$  to select the positive clones and subsequently into *E. coli* BL21 (DE3) for recombinant expression. Expression was carried out in LB medium containing kanamycin (100  $\mu$ g/mL) under various conditions.

SDS-PAGE and western blot (results not shown) have demonstrated that seven out of 12 recombinant proteases were insoluble form, whilst another five enzymes were not expressed. Even though it is common for some proteases to be insoluble, inactive or be in the cytoplasm, for example; serine protease from *B. lehensis* MTCC7633 (Joshi & Satyanarayana 2013), subtilisin from *Bacillus amyloliquefaciens* DC-4 and *Bacillus* sp. DJ-4 (Zhang et al. 2005; Choi et al. 2004), nattokinase and pronattokinase from *B. subtilis* (Chiang et al. 2005) and subtilase from *Virgibacillus* sp. SK37 (Phrommao et al., 2011); *B. lehensis* proteases' remained difficult to express recombinantly.

There are few ways in improving protein solubility, such as reducing the rate of protein synthesis, denaturing and refolding, changing growth condition and adding chaperones or fusion partner. Protein expressions were undergo several modifications and optimizations; such as changing the expression vector, host expression strain and expression condition including reducing temperature and IPTG concentration. All 12 genes were then expressed again in two vectors pET32 Ek/LIC and pET200 with several *E. coli* strains as a host cell. Recombinant plasmid pET32 Ek/LIC and pET200 from the selected clone was isolated and transformed into *E. coli* BL21 (DE3), BL21 star<sup>™</sup> (DE3), Origami 2 (DE3) and Rosetta-gami<sup>™</sup> 2 (DE3). The recombinants were compared in the *E. coli* strains with respect to obtain soluble proteases. pET32 Ek/LIC was used as it contains Trx•Tag<sup>™</sup> (thioredoxin protein) that allows the

formation of disulfide bonds in the *E. coli* cytoplasm which enhances the solubility of target proteins (LaVallie et al., 1993). Whereas, pET200 was expressed in BL21 Star™, with the lack of *rne131* gene that might improve mRNA stability and consequently enhance the protein expression (Kido et al., 1996) for non-expressed proteases such as BleG1\_3207, BleG1\_1058, BleG1\_3006, BleG1\_1047 and BleG1\_1242.

**Table 1:** Summary of thirteen proteases gene with assigned NCBI accession number, calculated protein size, isoelectric point, and conserved domain. (GAT\_1) indicates glutamine amidotransferase, (APP\_MetAP) family of aminopeptidase P, aminopeptidase M, and prolidase/metalloprotease\_M24, and (DAP\_dppA) D-Aminopeptidase dipeptide-binding protein/peptidase\_M55.

Name	Size (bp)	kDa	pI	NCBI Accession No.	Domain
BleG1_3041	663	19.1	5.35	AIC95658	GAT_1
BleG1_1242	783	29.0	5.86	AIC93863	Zinc_peptidase_like
BleG1_3207	1137	38.8	4.72	AIC95824	Peptidase_S8_S53
BleG1_3006	618	24.3	4.58	WP0384B2701	Peptidase_C15
BleG1_2536	693	22.8	5.66	WP038481565	GAT_1
BleG1_3196	819	29.9	5.36	WP038483149	DAP_dppA
BleG1_3228	768	27.8	6.58	WP038483224	APP_MetAP
BleG1_3423	594	20.0	4.36	WP051667653	Collagen
BleG1_0055	747	27.1	5.61	WP038475814	APP_MetAP
BleG1_0791	864	33.0	5.30	WP051667334	Peptidase_M23
BleG1_1047	741	26.9	5.64	WP038478032	APP_MetAP
BleG1_1979	525	19.1	5.35	WP038480120	GAT_1
BleG1_1058	918	33.9	4.33	WP051667390	Sortase

*E. coli* heterologous expression of *B. lehensis* G1 proteases seems to be impossible with all protein remains insoluble after countless screening of expression conditions. Proteases are challenging target for soluble expression in *E. coli* and this was suggested by finding that changing the parameters (vector, host strains and expression conditions) did not alter the formation of protein aggregates. Furthermore, the formation of inclusion body seems not

caused by the presence of rare codon, since expression in *E. coli* Rosetta-gami 2 (DE3) did not improve the solubility. Whilst, *E. coli* Origami 2 (DE3) and Rosetta-gami 2 (DE3) which carried two mutation in thioredoxin reductase (*trx-B*) and glutathione reductase (*gor*) for efficient folding of proteins was again failed to aid in *B. lehensis* G1 protease expression. Other methods including conventional IPTG induction and expression using auto-induction were also failed to give any effect on *B. lehensis* G1 protease activity or solubility. Initial attempt of denaturing and refolding for BleG1\_2536 and BleG1\_0055 did not show a promising result. This is contrast with refolded subtilisin from *Bacillus* sp DJ-4 and *B. amyloliquefaciens* DC-4 that retained its enzymatic activity after denaturing-refolding step using urea with overnight dialyzed (Choi et al., 2004; Zhang et al., 2005). In this respect, *B. lehensis* G1 proteases show to behave differently from recombinant proteins that were reported to be effortlessly refolded preceding to *E. coli* expression.

Computational homology modelling is another approach to obtain 3D structure from DNA sequence aside from experimental techniques through protein crystallisation. The aim of this step is to generate 3D model that is comparable to the best results achieved experimentally. The multistep processes of homology modelling consist of sequence alignment, template recognition, backbone generation, loop and side chain modelling followed by model optimization and validation (Krieger et al., 2003). BleG1\_1979 was selected for homology modelling with open reading frame of 525 bp and 19.1 kDa calculated molecular mass. Upon BLAST analysis, the amino acid sequences of BleG1\_1979 *B. lehensis* G1 demonstrated high similarity with *B. plakortidis* glutamine amidotransferase (NCBI Acession no. WP\_055737683). Sequence alignment of BleG1\_1979 with known 3D model in PDB showed 40% similarity with stress response from *Deinococcus radiourans* and intracellular protease from *Pyrococcus horikoshii* and *Thermococcus onnurineus*. The alignment of these sequences is shown in Figure 1. The catalytic residues in the active site are all conserved in these enzymes.

```

D.radiodurans    MTKAKDLTGKKIAILAADGVVEIELTSPRAAI EAAGGTTFLISLEPGEIQSMKGDIEPQE
B.lehensis      ----MRLKGGKIIAAVDEEFDELWYPIRLRQEEGAVDLVGHKGGTYIGKYGVP---
T.onnurineus    -----ASMKVLFLSADGFEDLELIYPLHRIKEEGHEVYVASFQRGK-ITGKHGYS---
P.horikoshii    -----XKVLFLTANEFEDVELIYPYHRLKEEGHEVYIASFERGT-ITGKHGYS---
                * :      : . * : * * * * * : : * . : . : * * *

D.radiodurans    KYRVDHVVSVEVQVSDYDGLLLPGGTVNPDKLRLEEGAMKFRVDMYDAGKPIAAIXHGPWS
B.lehensis      -ATAAYEFDEVQASHYDGLIIPGGW-APDKLRRYETLLSIVKEMDQAQKPIGQICHAGWV
T.onnurineus    -VNVDLTFEEVDPDEFDALVLPGGK-APEIVRLNEKAVMITRRMFEDDKPVASICHGPQI
P.horikoshii    -VKVDLTFDKVNPPEFDALVLPGGR-APERVRLNEKAVSIARKXKFSEGRPVASICHGPQI
                . . . * : . . * : * : * * * * * * * * * * * * * * * * *

D.radiodurans    LSETGIAQGLKMTSWSLKRLETLAQAQWVDEECVTDKGVVTSRKPDDLPAPFNKKIVEEF
B.lehensis      LISANILQGNVTSTPGIKDDMVNAGATWHDTAVIDGHIISRRPPDLPPYAKAFADKL
T.onnurineus    LISAKVLKRRGTSTITIRDDVINAGAEWIDAEEVVVDGNWVSSRHFGDLYAWMREFVKLL
P.horikoshii    LISAGVLRGRKGTSTYPIKDDXINAGVEWVDAEEVVVDGNWVSSRVPADLYAWXREFVKLL
                * . : : * . * * * * : : : * * * * * : * : * * * * * : : : . . .

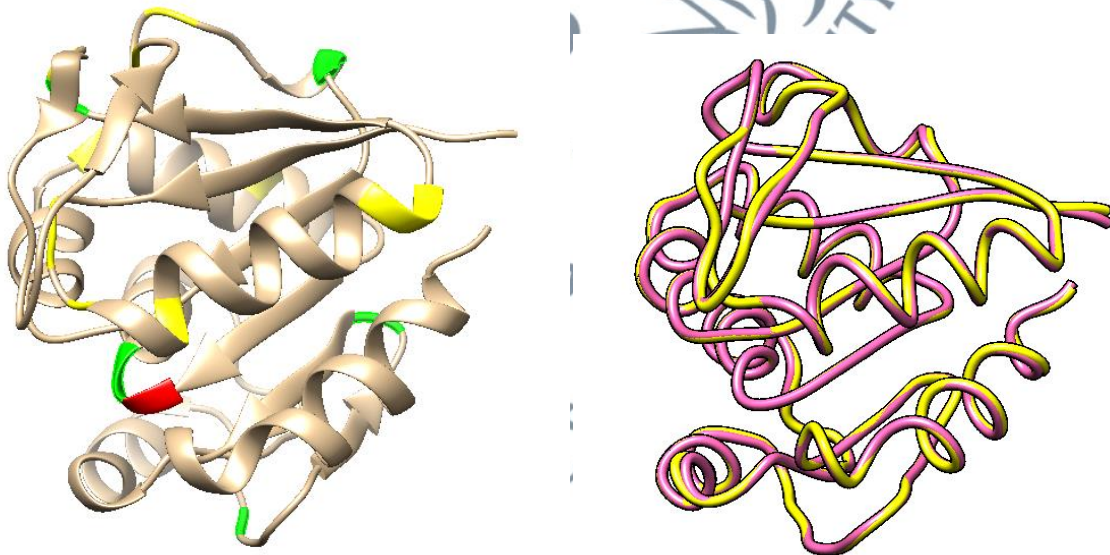
D.radiodurans    AEGDHSSRRK
B.lehensis      AEQ-----
T.onnurineus    H-----
P.horikoshii    K-----

```

**Figure 1:** Alignment of the BleG1\_1979 protease sequence (*B. lehensis* G1) with those of related structure (40%) found in the protein databases; stress response from *D. radiodurans*, and intracellular protease from *T. onnurineus* and *P. horikoshii*. (\*) indicates conserved residues, (:) conservative substitutions and (.) semi-conservative substitutions in all sequences.

The predicted structures were generated using Modeller v9.16 in PDB format and sorted according to DOPE (discrete optimized protein energy) and GA341. The quality of the generated BleG1\_1979 model was evaluated using PROCHECK, ERRAT, and Verify 3D. The backbone conformation was evaluated by examining the  $\phi/\psi$  degrees in Ramachandran plot obtained from the PROCHECK analysis (result not shown). Results showed that only one residue was located in the disallowed region (Cys103); 86.2 % of the residues were located in the most favoured region, whilst the residual residues were found to reside in the additional and generously allowed regions. Model assessment with ERRAT demonstrated the BleG1\_1979 model was found to be 95% which indicated a high quality since the accepted range for ERRAT score is >50 %. Verify 3D analysis showed that 92.9% of the BleG1\_1979 residues possess an average 3D-1D score greater than 0.2, suggesting that the predicted model is satisfying quality with score greater than 80%. Overall, the values obtained from the evaluation programs appeared reasonable for the predicted structure.

Figure 2 shows a representation of the BleG1\_1979 predicted 3D model and the superimposition of the model with the structure of *T. onnurineus* intracellular protease. The proposed model is represented as lines and superimposed well with *T. onnurineus*. There are eight  $\beta$  strands and seven helices with the central  $\beta$  sheet consist of six  $\beta$  strands with two strands on one side. Conserved Cys103 was located in the nucleophile elbow of sharp turn between  $\beta$  strand and  $\alpha$  helix. This is a distinctive strand-nucleophile-helix motif and catalytic triad in the  $\alpha/\beta$  hydrolases. Besides that, the  $\phi/\psi$  angles for Cys103 fall in an unfavorable region in the Ramachandran plot which demonstrates the characteristic of the nucleophile in a nucleophile elbow. All these features are similar to other members of the class DJ-1 superfamily (DJ-1/ThiJ/PfpI superfamily) for which crystal structures has been identified with a highly diverse range of cellular functions, including chaperone and protease activity (Jung et al. 2012). The conservation of the catalytic residues and the identical structure in the generated 3D model suggest that BleG1\_1979 model is an accurate representation of the actual protein's structure indicating a good overall structure.



**Figure 2:** General overview of ribbon and lines diagram of BleG1\_1979 (Chimera v1.10). On the left, the catalytic triad; glutamic acid (yellow), histidine (green) and conserved residue cysteine (red) that is located at the sharp turn between  $\beta$  strand and  $\alpha$  helix. On the right, the superimposition of the BleG1\_1979 of *B. lehensis* G1 (pink) and that of *T. onnurineus* (yellow).

#### 4. CONCLUSION

In conclusion, cloning and expression of *B. lehensis* G1 protease genes was successfully performed for the first time in *E. coli*. However, the recombinant enzymes require more optimization and investigation in order to obtain soluble protein that suitable for crystallization and later for industrial use. Currently, we attempt to express *B. lehensis* G1 protease genes with *E. coli* chaperones (Cpn10 and Cpn60) and yeast expression system, *Pichia pastoris*.

#### ACKNOWLEDGMENT

We would like to thank Science Fund from the Malaysia Ministry of Science, Technology and Innovation (MOSTI) (Grant Number: USIM/SF-UKM/FST/30/40214) for financial support.

#### REFERENCES

- Anbu, P. (2013). Characterization of solvent stable extracellular protease from *Bacillus koreensis* (BK-P21A). *International Journal of Biological Macromolecules*, 56, 162-168.
- Bajaj, B. K., & Jamwal, G. (2013). Thermostable alkaline protease production from *Bacillus pumilus* D-6 by using agro-residues as substrates. *Advances in Enzyme Research*, 1(2), 30-36.
- Bajaj, B. K., Sharma, N., & Singh, S. (2013). Enhanced production of fibrinolytic protease from *Bacillus cereus* NS-2 using cotton seed cake as nitrogen source. *Biocatalysis and Agricultural Biotechnology*, 2(3), 204-209.
- Benkiar, A., Nadia, Z. J., Badis, A., Rebzani, F., Soraya, B. T., Rekik, H., Naili, B., Ferradji, F.Z., Bejar, S. & Jaouadi, B. (2013). Biochemical and molecular characterization of a thermo- and detergent-stable alkaline serine keratinolytic protease from *Bacillus circulans* strain DZ100 for detergent formulations and feather-biodegradation process. *International Biodeterioration & Biodegradation*, 83, 129-138.
- Blanco, K. C., de Lima, C. J., Monti, R., Martins, J., Bernardi, N. S., & Contiero, J. (2012). *Bacillus lehensis* - an alkali-tolerant bacterium isolated from cassava starch wastewater: optimization of parameters for cyclodextrin glycosyltransferase production. *Annals of Microbiology*, 62(1), 329-337.
- Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 164-170.

- Chiang, C. J., Chen, H. C., Chao, Y. P., & Tzen, J. T. (2005). Efficient system of artificial oil bodies for functional expression and purification of recombinant nattokinase in *Escherichia coli*. *Journal of Agricultural and Food Chemistry*, 53(12), 4799-4804.
- Choi, N. S., Chang, K. T., Jae Maeng, P., & Kim, S. H. (2004). Cloning, expression, and fibrin (ogen)olytic properties of a subtilisin DJ-4 gene from *Bacillus* sp. DJ-4. *FEMS Microbiology Letters*, 236(2), 325-331.
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, 2(9), 1511-1519.
- Ghosh, A., Bhardwaj, M., Satyanarayana, T., Khurana, M., Mayilraj, S., & Jain, R. K. (2007). *Bacillus lehensis* sp. nov., an alkalitolerant bacterium isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, 57(2), 238-242.
- Gupta, R., Beg, Q., Khan, S., & Chauhan, B. (2002). An overview on fermentation, downstream processing and properties of microbial alkaline proteases. *Applied Microbiology and Biotechnology*, 60(4), 381-395.
- Haddar, A., Fakhfakh-Zouari, N., Hmidet, N., Frikha, F., Nasri, M., & Kamoun, A. S. (2010a). Low-cost fermentation medium for alkaline protease production by *Bacillus mojavensis* A21 using hulled grain of wheat and sardinella peptone. *Journal of Bioscience and Bioengineering*, 110(3), 288-294.
- Haddar, A., Sellami-Kamoun, A., Fakhfakh-Zouari, N., Hmidet, N., & Nasri, M. (2010b). Characterization of detergent stable and feather degrading, serine proteases from *Bacillus mojavensis* A21. *Biochemical Engineering Journal*, 51(1), 53-63.
- Illias, R.M., Fen, T.S., Abdulrashid, N.A., Yusoff, W.M.W., Hamid, A.A., Hassan, O., & Kamaruddin, K. (2002). Cyclodextrin glucanotransferase producing alkalophilic bacillus sp. G1: Its cultural condition and partial characterization of the enzyme. *Pakistan Journal of Biological Sciences*, 5, 688-692.
- Jonet, M. A., Mahadi, N. M., Murad, A. M. A., Rabu, A., Bakar, F. D. A., Rahim, R. A., Low, K.O. & Illias, R. M. (2012). Optimization of a heterologous signal peptide by site-directed mutagenesis for improved secretion of recombinant proteins in *Escherichia coli*. *Journal of Molecular Microbiology and Biotechnology*, 22(1), 48-58.
- Joshi, S., & Satyanarayana, T. (2013). Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. *Bioresource Technology*, 131, 76-85.
- Jung, H. J., Kim, S., Kim, Y. J., Kim, M. K., Kang, S. G., Lee, J. H., Kim, W. & Cha, S. S. (2012). Dissection of the dimerization modes in the DJ-1 superfamily. *Molecules and cells*, 33(2), 163-171.

- Kido, M., Yamanaka, K., Mitani, T., Niki, H., Ogura, T., & Hiraga, S. (1996). RNase E polypeptides lacking a carboxyl-terminal half suppress a *mukB* mutation in *Escherichia coli*. *Journal of Bacteriology*, 178(13), 3917-3925.
- Krieger, E., Nabuurs, S. B., & Vriend, G. (2003). Homology modelling. *Structural Bioinformatics*, 5, 507-521.
- Lagzian, M., & Asoodeh, A. (2012). An extremely thermotolerant, alkaliphilic subtilisin-like protease from hyperthermophilic *Bacillus sp.* MLA64. *International Journal of Biological Macromolecules*, 51(5), 960-967.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283-291.
- LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., & McCoy, J. M. (1993). A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Nature Biotechnology*, 11(2), 187-193.
- Li, Q., Yi, L., Marek, P., & Iverson, B. L. (2013). Commercial proteases: Present and future. *FEBS Letters*, 587(8), 1155-1163.
- Liew, C. W., Illias, R. M., Muhammad Mahadi, N., & Najimudin, N. (2007). Expression of the Na<sup>+</sup>/H<sup>+</sup> antiporter gene (*G1-Nhac*) of alkaliphilic *Bacillus sp.* G1 in *Escherichia coli*. *FEMS Microbiology Letters*, 276(1), 114-122.
- Low, K. O., Jonet, M. A., Ismail, N. F., & Illias, R. M. (2012). Optimization of a *Bacillus sp* signal peptide for improved recombinant protein secretion and cell viability in *Escherichia coli*: Is there an optimal signal peptide design? *Bioengineered*, 3(6), 334-338.
- Low, K. O., Mahadi, N. M., & Illias, R. M. (2013). Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Applied Microbiology and Biotechnology*, 97(9), 3811-3826.
- Mahmood, T., & Yang, P. C. (2012). Western blot: technique, theory, and trouble shooting. *North American Journal of Medical Sciences*, 4(9), 429-434.
- Noor, Y. M., Samsulrizal, N. H., Jema'on, N. A., Low, K. O., Ramli, A. N. M., Alias, N. I., Damis, S.I., Fuzi, S.F., Isa, M.N., Murad, A.M., & Raih, M. F. M. (2014). A comparative genomic analysis of the alkalitolerant soil bacterium *Bacillus lehensis* G1. *Gene*, 545(2), 253-261.
- Phrommao, E., Yongsawatdigul, J., Rodtong, S., & Yamabhai, M. (2011). A novel subtilase with NaCl-activated and oxidant-stable activity from *Virgibacillus sp.* SK37. *BMC Biotechnology*, 11(1), 65.
- Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.*

- Sian, H. K., Said, M., Hassan, O., Kamaruddin, K., Ismail, A. F., Rahman, R. A., Mahmood, N.A.N., & Illias, R. M. (2005). Purification and characterization of cyclodextrin glucanotransferase from alkalophilic *Bacillus sp.* G1. *Process Biochemistry*, 40(3), 1101-1111.
- Sinha, R., & Khare, S. K. (2013). Characterization of detergent compatible protease of a halophilic *Bacillus sp.* EMB9: differential role of metal ions in stability and activity. *Bioresource Technology*, 145, 357-361.
- Sundararajan, S., Kannan, C. N., & Chittibabu, S. (2011). Alkaline protease from *Bacillus cereus* VITSN04: Potential application as a dehairing agent. *Journal of Bioscience and Bioengineering*, 111(2), 128-133.
- Tan, S. H., Normi, Y. M., Leow, A. T. C., Salleh, A. B., Karjiban, R. A., Murad, A. M. A., Mahadi, N.M., & Rahman, M. B. A. (2014). A Sco protein among the hypothetical proteins of *Bacillus lehensis* G1: Its 3D macromolecular structure and association with Cytochrome C Oxidase. *BMC Structural Biology*, 14(1), 11.
- Wang, S. L., & Yeh, P. Y. (2006). Production of a surfactant-and solvent-stable alkaliphilic protease by bioconversion of shrimp shell wastes fermented by *Bacillus subtilis* TKU007. *Process Biochemistry*, 41(7), 1545-1552.
- Zhang, R. H., Xiao, L., Peng, Y., Wang, H. Y., Bai, F., & Zhang, Y. Z. (2005). Gene expression and characteristics of a novel fibrinolytic enzyme (subtilisin DFE) in *Escherichia coli*. *Letters in Applied Microbiology*, 41(2), 190-195.

## PRELIMINARY STUDY ON DIMENSIONAL STABILITY OF OIL PALM LUMBER TREATED WITH ROSIN-GUM RESIN

Hamid, A. B.,<sup>1</sup> Kamarudin, H.,<sup>2</sup> Hassan, A. R.<sup>3</sup> and Jalani, B. S.<sup>4</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1,3,4</sup>

Malaysian Palm Oil Board (MPOB), No. 6, Persiaran Institusi, Bandar Baru Bangi 43000 Kajang, Selangor<sup>2</sup>

hamidalkhair@gmail.com

### ABSTRACT

*This study focused on the dimensional stability of oil palm lumber treated with rosin-gum resin using a vacuum infusion method. Dry oil palm lumber was impregnated with the rosin-gum using a vacuum infusion system, and followed by a densification process. With the introduction of grooves and channels, the rosin flow was dispersed homogenously within the lumber macrostructure. The dimensional stability of densified rosin-treated lumber was determined by measuring its antiswelling efficiency (ASE). A positive ASE value indicated that the resin had penetrated into the cell wall and subsequently cross-linked, leading to bulking along interstitial spaces of parenchyma cells.*

**Keywords:** Oil palm lumber, rosin-gum resin, resin impregnation, dimensional stability.

### 1. INTRODUCTION

Oil palm lumber (OPL) is characterised as an anisotropic, heterogeneous and hygroscopic material. Without chemical pretreatment, the OPL is generally susceptible to biodegradation agents such as fungi and wood borers (Ho et al., 1985; William and Feist, 1999; Lim and Gan, 2005; Milling et al., 2005). The OPL, when exposed to cyclic humidity changes, is subjected to unequal expansion in the tangential and radial directions (Sulaiman et al., 2012). This creates internal stresses, which can result in distortion, such as checking, warping and twisting.

Certain chemicals used in traditional water repellents and preservative solutions, such as pentachlorophenol (PCP), copper-chrome-arsenate (CCA), borax and creosote are toxic to mammals (Weis and Weis, 2004; Patch and Maas, 2006; Rahman et al., 2009) and harmful to the environment (Cooper et al., 2001; Hale, 2003). Thus the use of these chemicals in many

regions of the world were banned, or being restricted in its uses (Drysdale, 2002; Townsend and Solo-Gabriele, 2006).

These developments have prompted the use of non-biocidal methods to minimise the environmental impact of treated finished products. One responds to this situation would be the use of rosin-gum resin as treatment chemical for OPL. The rosin, non-volatile solid form of resin, is produced by one of three routes, namely (a) oleoresin is obtained by tapping living pine trees, and then distilling the exudate to produce rosin and turpentine, (b) wood rosin is produced by solvent extraction of aged pine stumps, and (c) tall oil rosin, which is the useful by-product of the kraft pulping process. The production of rosin is more than 1 million tonnes per year (Liu & Urban, 2010; Ceisla, 2003). Over the years, the rosin is used in a wide range of applications such as in the manufacture of adhesives, paper sizing agents and printing inks (Smith et al., 2010; Yao & Zheng, 2000; Alexander & Shakesheff, 2006).

Rosin-gum resin consists primarily of abietic- and pimaric-type resin acids with a suitable hydrophobic character and affinity for lignocellulosics (Wiyono et al., 2006). Different chemical mechanisms between copper, rosin and wood constituents have also been investigated (Voulgaridis, 1993). For example, the copper-rosin soaps were extremely efficient repellents to both fungi (unsterile soil bed test) and termites (field test) (Pizzi, 1993). In another study, the decay resistance of wood seemed to improve due to the decreased in moisture absorbing tendency when subjected to a rosin sizing agent (Li et al., 2011).

The objective of this paper was focused on OPL treated with rosin gum resin that was impregnated using vacuum infusion method by measuring its ASE (Anti Swelling Efficiency) value and its water reduction as well.

## 2. MATERIALS AND METHOD

### 2.1 Preparation of Rosin-Treated Oil Palm Lumber

Thirty pieces of dry OPL with dimension of 100 mm wide by 100 mm thick by 400 mm long were obtained from Malaysian Palm Oil Board (MPOB). The OPL was individually treated

with a rosin-gum resin solution of 30% concentration (w/v) methyl ethyl ketone (MEK) using a vacuum infusion method. The rosin-treated OPL was then compressed to about 75% of its initial thickness using a cold press in order to remove excess rosin solution from the sample and to make it more compact. To complete cross-linking of the rosin-gum resin, the densified rosin-treated OPL sample was cured under pressure in an electric oven at  $60 \pm 5^\circ\text{C}$  for five days (Kamarudin et al., 2007).

## 2.2 Determination of Rosin-Gum Resin Penetration

Small cubic samples (in the form of a square cross-section of side 5 mm and length along the vascular bundles of 10 mm) were cut across the longitudinal direction with a razor-sharp knife. The samples were mounted on the stub using double-sided copper tape followed by sputter coating with a 38 nm layer of gold in a vacuum evaporator in order to give it electrical conductivity for scanning electron microscopy (SEM) observations.

In the SEM, the stub (with the specimen) was tilted and rotated in all directions in order to increase the area scanned by the beam, and therefore, creates perspective in the image. The penetration of resin in the lumber matrix was determined from the respective photomicrographs. A Hitachi S2700 SEM operated at 15kV was used for the study.

The penetration of rosin-gum in the OPL matrix was determined from the respective photomicrographs, while weight percent gain (WPG) due to rosin-gum load was calculated (Ozcifci et al., 2007) as follows:

$$WPG(\%) = \frac{W_{of} - W_{oi}}{W_{oi}} \times 100$$

where WPG is the weight percent gain on an oven-dry basis in percentage,  $W_{oi}$  is the oven-dry mass, in grams, of the OPL sample before rosin-gum treatment, and  $W_{of}$  is the oven-dry mass, in grams, of the OPL sample after subjected to rosin-gum treatment.

### 2.3 Dimensional Stability Test

Dimensional stabilization was quantified by comparing the volumetric swelling coefficients of both the densified rosin-treated OPL and OPL devoid of rosin treatment (control) samples. Measurements of anti-swelling efficiency (ASE) were carried out by subjecting to an exhaustive water-soaked treatment (Rowell & Ellis, 1978).

These samples were dried at 105 °C for 24 hr, cooled in a desiccator for 3 hr, and the three structural directions measured. After weighing, the samples were then individually vacuum-infused with distilled water until they were fully wet. The OPL samples were then soaked under water at 20 °C for 24 hr. After soaking, the samples were individually weighed and its dimensions were measured at the same point, which were taken as initial measurements. The ASE imparted by the treatment was computed (Islam et al., 2012) as follows:

$$ASE (\%) = \frac{S_{vc} - S_t}{S_{vc}} \times 100$$

where  $S_{vc}$  is the volumetric swelling coefficient of the match-OPL sample devoid of rosin-gum,  $S_t$  is the volumetric swelling coefficient of densified rosin-treated sample of OPL.

The volumetric swelling coefficient ( $S_{vc}$ ) was calculated by using the following equation:

$$S_{vc}(\%) = \frac{V_w - V_d}{V_d} \times 100$$

where  $V_w$  is the volume of OPL sample after soaking in  $m^3$  while  $V_d$  denotes the volume of respective oven-dry sample of OPL in  $m^3$  prior to soaking.

## 3. RESULTS AND DISCUSSION

### 3.1 Penetration of Rosin-Gum Resin Into the Lumber Matrix

The rosin-gum resin used was dispersed homogenously within the matrix of lumber sample, and thus, remove the rosin excess and dry the OPL sample (Figure 1). For the OPL samples, mean amount of rosin-gum absorbed and the time taken to infuse the rosin per unit volume of the lumber samples were  $0.56 \text{ g cm}^{-3}$  and  $1.4 \text{ s cm}^{-3}$ , respectively. Brouwer et al., (2007), has explained the vacuum infusion mechanism using different sample (bamboo) while applying the same method. The pressure required to vacuum the rosin was start from 0 MPa and suddenly increase to 1 MPa. The value of pressure continue to decrease which means the rosin gum resin have penetrated successfully.



**Figure 1:** A sample of densified rosin-treated palm lumber, y: the pressure value and x: time during pressure.

### 3.2 Effect of Rosin-Gum and Densification Treatments on Dimensional Stability

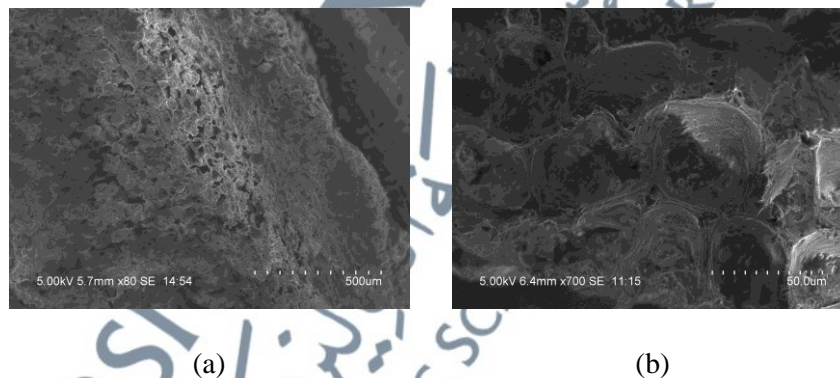
The anti-swelling efficiency (ASE) values show that the rosin-gum resin was able to impart a degree of stability to the densified rosin-treated OPL were sampled in middle/pith toward outer zones of OPL. The densified rosin-treated OPL yielded ASE values (Table 1), as follows: outer zone ranged from 3.3% to 23.03% and the middle/pith zone ranged from 0.89% to 9.27%.

**Table 1:** Variations of dimensional stability of densified rosin-treated oil palm lumber.

Zone	Number of Specimens	Antiswelling Efficiency (%)			
		Min	Mean*	Max	SE
Outer	9	3.3	7.85	23.03	2.39
Middle/pith	9	0.89	4.37	9.27	1.04

SE is standard error of the mean, values assigned the same letter do not statistically significant difference at 95% confidence level.

The positive ASE value indicated that the rosin-gum resin had been deposited into the cell wall and subsequently cross-linked, leading to bulking along interstitial spaces (Loh et al., 2010). These interstitial spaces were created due to the collapse of adjacent parenchymatous tissues during drying (Obataya et al., 2004). A typical SEM image of the vascular bundles of rosin-treated sample is shown in Figure 2. SEM was used in order to visualize the resin penetration (phenol formaldehyde and urea formaldehyde) into the pores of lumber (Abdullah et al., 2012). Rosin gum resin was successfully filled the pores of sample which shaped as a sponge.



**Figure 2:** Typical SEM photomicrographs showing the distribution of rosin-gum resin within the vascular bundles of densified resin-treated oil palm lumber a) magnification 80 and b) magnification 700.

The amount of water absorption was reduced from 73.9% to 69.2%, as shown in Table 2. This reduction shows that rosin gum resin could prevent the water penetration, reduce the swelling, and preserve the dimensional stability. It was noted that polymer grafting might have taken place with reactive group on OPL component within the cell wall while ungrafted bulk polymer

was formed in the OPL voids (Rowell et al., 1982). The PF resin used was deposited either in cell-lumen (Schneider, 1995), in cell-wall (Furano et al., 1992) or a combination of cell-lumen and cell-wall types (Schneider & Phillips, 1991).

**Table 2:** Variations of water reduction of densified rosin-treated oil palm lumber.

Zone	Number of Specimens	Amount (%)			
		Min	Mean*	Max	SE
Outer	9	55.97	69.2	83.02	3.49
Middle/pith	9	60.02	73.9	90.12	3.2

SE is standard error of the mean, values assigned the same letter do not statistically significant difference at 95% confidence level.

#### 4. CONCLUSION

The rosin-gum resin used was homogeneously dispersed within the OPL structure using a vacuum infusion method. It was noted that the rosin flow was directly proportional to the permeability of OPL macrostructure and the pressure differences between the inlet and outlet of the vacuum system, but inversely proportional to the viscosity of rosin-gum resin used. The positive ASE values proved that the rosin had penetrated and deposited in the OPL vascular bundles.

#### ACKNOWLEDGEMENT

A great many people have contributed, in their own ways, to the work embodied in this study. Gratitude must be expressed primarily to the management of MPOB for providing the oil palm lumber samples and practical assistance in dealing with the day-to day problems inherent in a study such as this one. Thanks of a rather different nature must be extended to the Faculty of Science and Technology of USIM for providing support to conduct the research.

## REFERENCES

- Abdullah, C. K., Jawaid, M., Khalil, H. A., Zaidon, A., & Hadiyane, A. (2012). Oil palm trunk polymer composite: Morphology, water absorption, and thickness swelling behaviours. *BioResources*, 7(3), 2948-2959.
- Alexander, C., & Shakesheff, K. M. (2006). Responsive polymers at the biology/materials science interface. *Advanced Materials*, 18(24), 3321-3328.
- Ceisla, W. M. 2003. *Products of Resin Processing*. Forest Health Management International, Fort Collins, Colorado, USA
- Cooper, P. A., Jeremic, D., & Taylor, J. L. (2001). Residual CCA levels in CCA-treated poles removed from service. *Forest Products Journal*, 51(10), 58-62.
- Drysdale, J. (2002). Global regulatory changes restrict the use of CCA. *New Zealand Journal of Forestry*, 47, 9-10.
- Furano, T., Uehara, T. and Jodai, S. 1992. The role of wall polymer in the decay durabilities of wood-polymer composites. *Makuzai Gakkaishi*. 38(3):285-293.
- Hale, M. D. C. 2003. *Novel method of wood preservation*. School of Agricultural and Forest Sciences, University of Wales: Bangor, UK.
- Ho, K. S., Choo, K. T. and Hong, L. T. 1985. Processing, seasoning and protection of oil palm lumber. *Proceedings of the National Symposium on Oil Palm By-products for Agro-based Industries*. OPTUC, Kuala Lumpur. p. 43-51.
- Islam, M. S., Hamdan, S., Rusop, M., Rahman, M. R., Ahmed, A. S., & Idrus, M. M. (2012). Dimensional stability and water repellent efficiency measurement of chemically modified tropical light hardwood. *BioResources*, 7(1), 1221-1231.
- Kamarudin, H., Anis, M., Choo, Y. M. and Mohd Basri, W. 2007. 3-N lumber from oil palm trunk. *MPOB Information Series MPOB TT No. 364*.
- Li, S., Thanh-Hien, N. T., Han, S., & Li, J. (2011). Application of rosin in wood preservation. *Chemistry and Industry of Forest Products*, 31(5), 117-121.
- Lim, S. C. and Gan, K. S. 2005. Characteristics and utilisation of oil palm stem. *Timber Technology Bulletin*. No. 35. Forest Research Institute, Kepong.
- Liu, F., & Urban, M. W. (2010). Recent advances and challenges in designing stimuli-responsive polymers. *Progress in Polymer Science*, 35(1), 3-23.
- Loh, Y. F., Paridah, M. T., Hoong, Y. B., Bakar, E. S., Hamdan, H., & Anis, M. (2010). Properties enhancement of oil palm plywood through veneer pretreatment with low molecular weight phenol-formaldehyde resin. *Journal of Adhesion Science and Technology*, 24(9), 1729-1738.

- Milling, A., Kehr, R., Wulf, A., & Smalla, K. (2005). Survival of bacteria on wood and plastic particles: Dependence on wood species and environmental conditions. *Holzforschung*, 59(1), 72-81.
- Obataya, E., Gril, J., & Thibaut, B. (2004). Shrinkage of cane (*Arundo donax*) I. Irregular shrinkage of green cane due to the collapse of parenchyma cells. *Journal of wood science*, 50(4), 295-300.
- Ozcifci, A., Toker, H., & Baysal, E. (2007). Fire properties of laminated veneer lumber treated with some fire retardants. *Wood Research*, 52(4), 37-46.
- Patch, S. C. and Maas, R. P. 2006. Arsenic release and exposure from lumber treated with chromated copper arsenate. Chapter 12. *Environmental Impacts of Treated Wood*. Taylor & Francis, Boca Raton.
- Pizzi, A. 1993. A new approach to nontoxic, wide-spectrum, ground-contact wood preservatives. II. Accelerated and long-term field tests. *Holzforschung*. 47(4):343-348.
- Rahman, M. M., Ng, J. C., & Naidu, R. (2009). Chronic exposure of arsenic via drinking water and its adverse health impacts on humans. *Environmental Geochemistry and Health*, 31(1), 189-200.
- Rowell, R. M., & Ellis, W. D. (2007). Determination of dimensional stabilization of wood using the water-soak method. *Wood and Fiber Science*, 10(2), 104-111.
- Rowell, R. M., Moisuk, R., & Meyer, J. A. (1982). Wood-polymer composites: cell wall grafting with alkylene oxides and lumen treatments with methyl methacrylate. *Wood Science*, 15, 90-96.
- Schneider, M. H. (1995). New cell wall and cell lumen wood polymer composites. *Wood Science and Technology*, 29(2), 121-127.
- Schneider, M. H., & Phillips, J. G. (1991). Elasticity of wood and wood polymer composites in tension compression and bending. *Wood Science and Technology*, 25(5), 361-364.
- Smith, A. E., Xu, X., & McCormick, C. L. (2010). Stimuli-responsive amphiphilic (co) polymers via RAFT polymerization. *Progress in Polymer Science*, 35(1), 45-93.
- Sulaiman, O., Fahmi Awalludin, M., Hashim, R., & Mondal, M. I. H. (2012). The effect of relative humidity on the physical and mechanical properties of oil palm trunk and rubberwood. *Cellulose Chemistry and Technology*, 46(5), 401-407.
- Townsend, T. and Solo-Gabriele, H. 2006. *Environmental impacts of treated wood*. CRC, Boca Raton, Florida, USA.

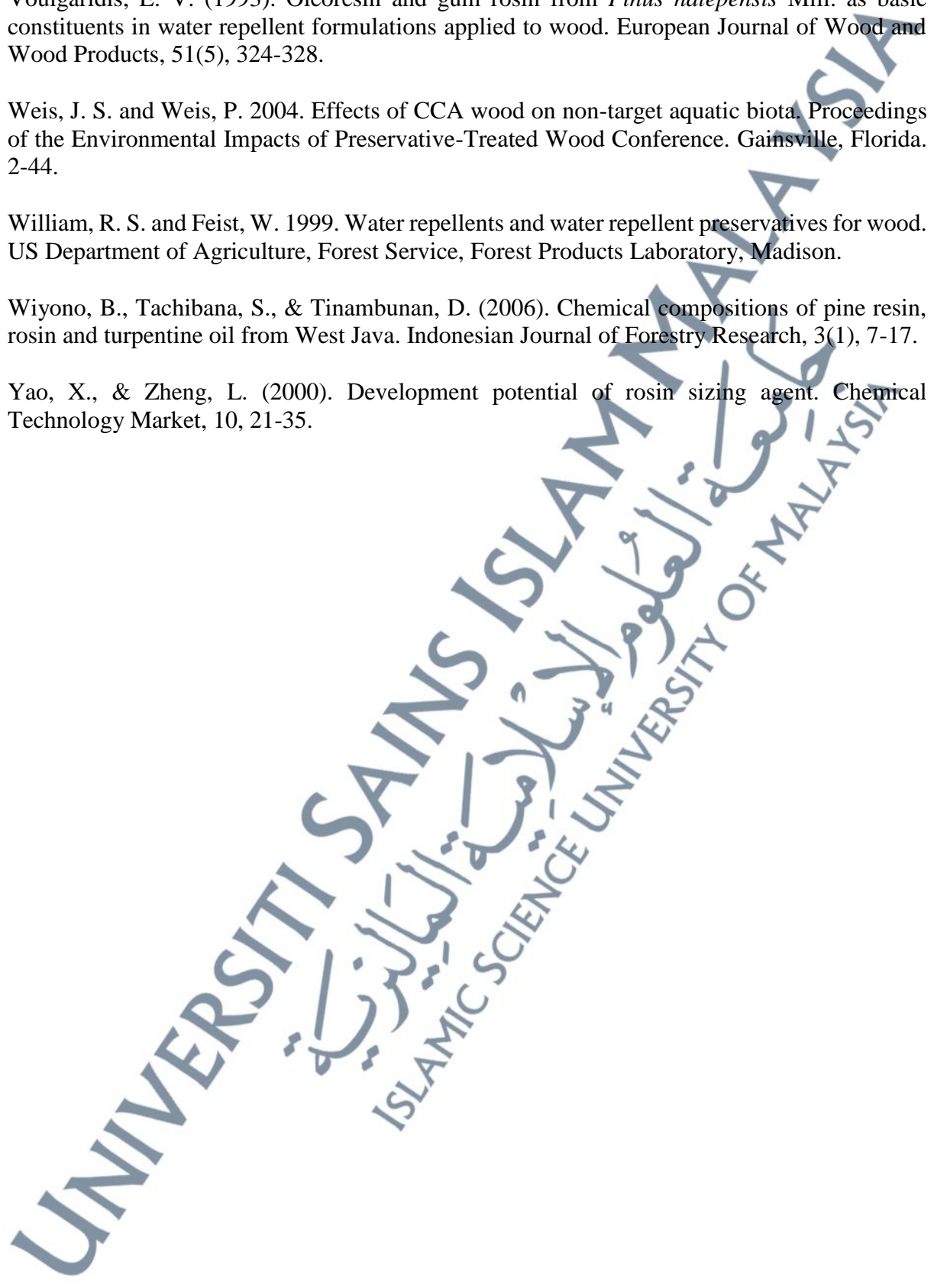
Voulgaridis, E. V. (1993). Oleoresin and gum rosin from *Pinus halepensis* Mill. as basic constituents in water repellent formulations applied to wood. *European Journal of Wood and Wood Products*, 51(5), 324-328.

Weis, J. S. and Weis, P. 2004. Effects of CCA wood on non-target aquatic biota. *Proceedings of the Environmental Impacts of Preservative-Treated Wood Conference*. Gainesville, Florida. 2-44.

William, R. S. and Feist, W. 1999. Water repellents and water repellent preservatives for wood. US Department of Agriculture, Forest Service, Forest Products Laboratory, Madison.

Wiyono, B., Tachibana, S., & Tinambunan, D. (2006). Chemical compositions of pine resin, rosin and turpentine oil from West Java. *Indonesian Journal of Forestry Research*, 3(1), 7-17.

Yao, X., & Zheng, L. (2000). Development potential of rosin sizing agent. *Chemical Technology Market*, 10, 21-35.



## ENHANCE CO<sub>2</sub> HYDRATE FORMATION INSIDE THE HPVA BY VIGOROUS STIRRING DURING SAMPLE PREPARATION

Mohd Hafiz, A. H.<sup>1</sup>, Snape, C. E.<sup>2</sup> and Stevens, L.<sup>3</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1</sup>

Cleaner Fossil Energy and Carbon Capture Technologies Research Group, Faculty of Engineering, The University of Nottingham<sup>2,3</sup>

mhafiz.a.h@usim.edu.my

### ABSTRACT

*A rise of 2 °C in the Earth's temperature is likely to occur when the concentration of carbon dioxide (CO<sub>2</sub>) in the atmosphere reaches approximately 450 ppm. Carbon capture and storage (CCS) is required to reduce CO<sub>2</sub> emissions. CO<sub>2</sub> hydrate (CO<sub>2</sub>:6H<sub>2</sub>O) formation has been investigated as a way to capture CO<sub>2</sub>. The formation of hydrate in this work was experimentally investigated in batch mode inside a high pressure volumetric analyser (HPVA). Standard silica gel with an average particle size of 200-500 μm, mean pore size of 5.14 nm, a pore volume of 0.64 cm<sup>3</sup>/g and a surface area of 499 m<sup>2</sup>/g was used as a porous medium. Hydrate formation was confirmed when the mole fraction of CO<sub>2</sub> dissolved in water exceeded the Henry's Law value as well as a two-stage pressure drop in the experimental pressure-time (P-t) curve. Various sample preparation methods (methods 1, 2, 3 and 4) were studied leading to the selection of method 4 (the use of vigorous stirring) as the best method which had the highest moisture content (14.8 wt.%) and the greatest water conversion to hydrate (40.5 mol%) at 275 K and 36 bar in a pure CO<sub>2</sub> gas system.*

**Keywords:** Greenhouse effect, CO<sub>2</sub> hydrate, silica gel, Henry's Law and P-t curve.

### 1. INTRODUCTION

Carbon dioxide (CO<sub>2</sub>) belongs to a category of gases known as 'greenhouse gases (GHG)' wherein human activities since the Industrial Revolution such as burning coal, natural gas and oil to power machines for manufacturing and transportation have now increased the concentration of CO<sub>2</sub> in the atmosphere, intensifying the natural warming caused by the greenhouse effect (Caitlyn, 2014). Biello (2013) reported that CO<sub>2</sub> levels in the atmosphere have reached 400 ppm for the first time in at least 800,000 years on May 7, 2013, around 40% greater than the value recorded before the Industrial Revolution: 280 ppm (Caitlyn, 2014). Biello (2013) also stated that global temperatures are estimated to rise by 2 °C if the

concentration of CO<sub>2</sub> in the atmosphere reaches 450 ppm. With the persistent use of fossil fuels, the capture of CO<sub>2</sub> needs to be considered (Kunze & Spliethoff, 2012). According to IEA (2008), carbon capture and storage (CCS) has been a favourable mitigation against climate change where it will be 70% more costly without CCS than with it by 2050. CCS is a 3-step process including CO<sub>2</sub> capture from power plants, industrial sources, and natural gas wells with high CO<sub>2</sub> content, transportation (usually via pipelines) to the storage site, and geological storage in deep saline formations, depleted oil/gas fields, unmineable coal seams, and enhanced oil or gas recovery (EOR or EGR) sites (IEA, 2006). There are three leading technologies for CO<sub>2</sub> capture namely post-combustion capture, pre-combustion capture in the integrated combined cycle (IGCC) plants and oxy-fuel combustion. D'Alessandro et al. (2010) highlighted several new materials for CO<sub>2</sub> capture methods such as physical or chemical absorbents, adsorption of solids, cryogenic systems, membranes, hydrate based gas separation (HBGS) and chemical looping combustion by metal oxides. HBGS is one of the most promising approaches in CO<sub>2</sub> capture field (Zheng et al., 2016; Babu et al., 2014; Linga et al., 2007, McCallum et al., 2005). This CO<sub>2</sub> capture method involves clathrate or gas hydrate crystallisation and can be applied to both post- (from flue gas) and pre-combustion (from fuel gas) capture respectively. However, the HBGS process is more suitable for pre-combustion CO<sub>2</sub> capture from a fuel gas mixture than a flue gas mixture because the partial pressure of the shifted fuel gas (40% CO<sub>2</sub> & 60% H<sub>2</sub>) is up to 1000 times higher than that of the flue gas (17% CO<sub>2</sub> & 83% N<sub>2</sub>) in post-combustion capture (IEA, 2011). Thus, HBGS has been chosen in this study because the CO<sub>2</sub> hydrate formation is a continuous operation, thus making it possible to treat a large amount of gaseous stream, recovering more than 99 mol% of CO<sub>2</sub> from the fuel gas (Babu et al., 2013).

The HBGS process relies on the ability of water to form non-stoichiometric crystalline compounds in the presence of CO<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub> and H<sub>2</sub>, as well as natural gas components at high pressures (10-70 bar) and low temperatures (near 273 K) (D'Alessandro et al., 2010). Carrol (2009) stated that the CO<sub>2</sub> hydrate formation conditions are in the range of 273.15-283 K at the pressure range of 12.7-45 bar wherein this is the case for a pure CO<sub>2</sub> system. A number of techniques have been investigated by researchers to capture CO<sub>2</sub> with gas hydrate formation by implementing the most traditional way (bulk water), solid adsorbent (porous medium) and promoters (oxygenated solvents, surfactants, semi-clathrate hydrate and hydrophobic former).

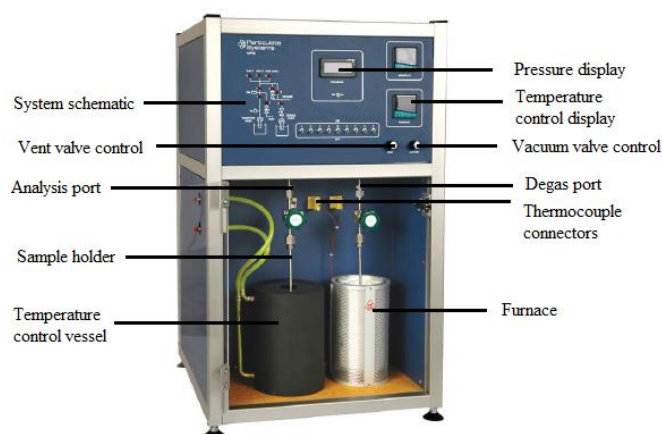
Also, numerous parameters have been investigated to improve CO<sub>2</sub> uptake, ease of hydrate formation and operational costs by employing solid adsorbent so that HBGS will become the most preferred method in the CCS industry. The stated parameters are promoter concentrations, driving forces, fixed bed heights and pore sizes of solid adsorbents. However, no research has been done in the determination of the best method to prepare solid adsorbent as a medium for CO<sub>2</sub> hydrate formation.

Consequently, several available options for sample preparation methods by using solid adsorbent in gas capture field were studied. Basically, each method has the same objective which is to completely fill the pores with water. Babu et al. (2013) used silica sand and silica gel as the solid beds in the reactor wherein they used a stirrer (3000 rpm) for 3 minutes to aid the dispersion of water into pore saturated silicas. This method was adopted from Adeyemo et al. (2010). Park et al. (2013) mixed an amount of water identical to the pore volume of silica gel inside the bottle consisting of dry silica gel and an ultrasonic wave was used to blend them. Wang et al. (2008) prepared numerous samples wherein the mixture of hydrophobic silica and water were stirred at three different speeds (16450, 17500 and 19000 rpm) for 90 seconds and they discovered that the formation of methane hydrate for the sample stirred at the highest speed gave the highest gas uptake. In other work, Dawson et al. (2014) mixed hydrophobic silica and alkaline solution and this mixture was stirred at 37000 rpm for 30 seconds to prepare a porous sample for CO<sub>2</sub> capture investigation at low pressure (2-3 bar). The sample preparation method employed by Wang et al. (2008) and Dawson et al. (2014) is beneficial because a water-in-air inverse foam consisting of water droplets is surrounded by a network of hydrophobic fumed silica which prevents coalescence of the droplets. In other words, it can increase the amount of water inside silica pores which is essential for hydrate formation. Thus, the focus of this study was to investigate the advantage of vigorous stirring during sample preparation on the formation of hydrate inside fixed bed reactor (FBR) or HPVA wherein silica gel was used as a solid adsorbent (porous medium) to omit the need of stirring process inside the reactor. Hence, two approaches were used to study the formation of hydrate inside the HPVA: analysis of P-t curves (Tang et al., 2013) and study of CO<sub>2</sub> dissolution in water suggested by Servio & Englezos (2001). Therefore, this study will help in providing the suitable sample preparation method for the formation of hydrate inside any reactor which employ adsorbent as a porous medium.

## 2. EXPERIMENTAL

### 2.1. Apparatus and Materials

The high pressure volumetric analyser (model HPVA-100 – Figure 1) consists of constant temperature bath or water bath (was used to control the operating temperature inside the temperature control vessel), vacuum pump (was needed to ensure the system reached full vacuum during the purging process before the commencement of each experiment) and degassing unit (was used to measure the equilibrium moisture content) was manufactured by Micromeritics. The oven used (Model AX30) was manufactured by Carbolite and has a maximum temperature of 250 °C and a minimum temperature of 40 °C. The weighing balance (Model: AEA – 220A) used was manufactured by AE Adam and it can measure a maximum mass up to 220 g and a minimum mass of 10 mg. The high-speed blender was used to vigorously mixed silica gel with water.



**Figure 1:** High pressure volumetric analyser (HPVA)

As for materials, standard silica gel (with a mean particle size of 200-500  $\mu\text{m}$ , mean pore size of 5.14 nm, pore volume of 0.64  $\text{cm}^3/\text{g}$  and surface area of 499  $\text{m}^2/\text{g}$ ) was purchased from Fisher Scientific.  $\text{CO}_2$  gas with a purity of 99.99% (maximum 103 bar), helium (He) gas with purity of 99.99% (34.4 bar) for venting or cleaning purposes and nitrogen ( $\text{N}_2$ ) gas/compressed air (5.2 to 5.5 bar) to control the pneumatic valves of the HPVA were supplied by BOC (a member of the Linde Group). Antifreeze was purchased from ASDA. All the materials were used without further purification.

## 2.2. Various Sample Preparation Methods

Four methods were used for the preparation of wet silica gel, namely methods 1 (silica was left to naturally adsorbed moisture), 2 (the lowest rates of stirring of silica), 3 (silica was submerged in excess water) and 4 (the highest rates of stirring of silica). For all methods, silica gel was initially dried inside the oven for one night at 200°C before the commencement of each experiment. Finally, each final equilibrium mass of wet silica gel was deducted with this dry silica gel mass to obtain the final moisture content respectively.

### Method 1

- i. Dry silica gel (2.5 g) was placed inside weighing boat and left at atmospheric condition.
- ii. The silica gel naturally adsorbed water until the final mass reached equilibrium.

### Method 2

- i. Dry silica gel (2.5 g) was placed inside beaker and water was added in excess (19 times the mass of dry silica gel) so that the total mass of the mixture was 50 g.
- ii. The silica gel and water mixture were slowly stirred by using a magnetic stirrer for one night.
- iii. Then, the beaker was left at atmospheric condition until the final mass of the mixture reached equilibrium.

### Method 3

- i. Dry silica gel (0.5 g) was placed inside weighing boat and water was added in excess (19 times the mass of dry silica gel) so that the total mass of the mixture was 50 g.
- ii. Then, it was left at atmospheric condition until the final mass reached equilibrium.

### Method 4

- i. Dry silica gel (0.5 g) was placed inside blender and water was added in excess (19 times the mass of dry silica gel) so that the total mass of the mixture was 50 g.
- ii. The silica gel and water mixture were vigorously stirred at the speed of 37,000 rpm (Dawson et al., 2014) by using a high-speed blender for 90 seconds.

- iii. Then, the mixture was left at atmospheric condition until the final mass reached equilibrium.

### 2.3. *Equilibrium Moisture Content Measurement*

The sample cell was charged with wet silica gel and was placed inside furnace and the cell's valve was initially closed. Then, the furnace was heated to 200 °C by temperature controller and at the same time the vacuum pump was switched on to vent out any gas and moisture presence in the system. After the operating temperature was achieved, then the cell's valve was slowly opened until maximum opening. The experiment was left running for one night. Finally, the amount of equilibrium moisture content was calculated.

### 2.4. *CO<sub>2</sub> Hydrate Formation*

#### 2.4.1. *Principles of the HPVA*

The HPVA-100 (Figure 1) was designed to obtain high-pressure adsorption isotherms by employing the static volumetric method where it consists of introducing (dosing) a known amount of gas (adsorptive) into the chamber containing the sample to be analysed. It is capable of achieving pressures up to 100 bar. When the sample reaches equilibrium with the adsorbate gas, the final equilibrium pressure is recorded and this process is repeated at given pressure intervals until the maximum pre-selected pressure is reached. These data are then used to calculate the quantity of gas adsorbed by the sample and resulting equilibrium points (volume adsorbed and equilibrium pressure) are then plotted to provide an isotherm. Excellent reproducibility and accuracy are obtained by using separate transducers for dosing the sample and for monitoring the pressure in the sample chamber.

#### 2.4.2. *Experimental Procedure*

##### *Before the commencement of experiment*

The system was manually purged with He gas to clean the line from any impurities. This step was repeated three times. Next, the operating conditions such as experiment time, analysis gas port, operating pressure and temperature were pre-defined. After that, the sample cell was charged with wet silica gel and was placed inside water bath. The mixture of 70 vol% water +

30 vol% antifreeze was used to avoid the formation of ice inside water bath and to make sure that the water mixture was consistently circulated throughout the process. The cell's valve was initially closed.

#### *During experiment*

The cell was pressurised to required pressure through a supply vessel containing CO<sub>2</sub> gas with a purity of 99.99% and at the same time the desired operating temperature was established through constant temperature bath. After the operating conditions were achieved, the cell's valve was fully opened. Subsequently, the experiment was left running for 1200 minutes.

#### *After the completion of experiment*

The pressure was reduced to atmospheric pressure at the same operating temperature for hydrate decomposition. Then, the system was automatically vented with He gas several times to clean the line for the next experiment. Subsequently, the cell's valve was fully closed and the sample cell was removed from the HPVA. Finally, the pressure-time (P-t) curve obtained after completion of the experiment was analysed.

### 3. CALCULATIONS

#### 3.1. Calculation of The Amount of Equilibrium Moisture Content (wt%)

The amount of equilibrium moisture content was presented in weight percent of wet silica gel. Equation 1 was used to calculate weight percent of moisture content available inside wet silica gel pores.

$$\text{equilibrium moisture content (wt\%)} = \frac{\text{moisture content (g)}}{\text{wet silica gel (g)}} \times 100\% \quad (\text{Eq. 1})$$

#### 3.2. Calculation of the Amount of CO<sub>2</sub> Consumed (mol)

The number of moles of gas consumed for hydrate formation is the difference between the number of moles of gas at the start of experiment (t = 0) and the number of moles of gas at time t in the sample cell (S). Equation 2 was used to calculate the number of moles of CO<sub>2</sub> gas consumed,  $\Delta n_1$ .

$$(\Delta n_{\downarrow})_t = V_S \left( \frac{P}{zRT} \right)_0 - V_S \left( \frac{P}{zRT} \right)_t \quad (\text{Eq. 2})$$

where the compressibility factor,  $z$ , is obtained from REFPROP\*,  $V_S^{**}$  is the volume of the sample cell,  $R$  is the gas constant,  $P$  and  $T$  are the pressure and temperature of the sample cell.

\*NIST standard reference database – National Institute of Standards and Technology

\*\* $V_S$  obtained in the data sheet was in ccSTP. The conversion of ccSTP to  $\text{cm}^3$  was performed by using combined gas law.

### 3.3. Calculation of the Amount of $\text{CO}_2$ Involved In Hydrate Formation

The amount of  $\text{CO}_2$  consumed for each experiment was further analysed to investigate the formation of hydrate by using Henry's Law. If the total mole fraction of  $\text{CO}_2$ ,  $x_{D,t}$  dissolved in water exceeded the value calculated from Henry's Law, it shows the formation of hydrate in the system.

#### 3.3.1. Henry's Law

Equation 3 was used to calculate the theoretical value of  $x_{D,t}$  by using Henry's Law.

$$x_{D,t} (\text{mole fraction}) = \frac{(\text{Partial pressure of } \text{CO}_2)_t}{(\text{KH})_t} \quad (\text{Eq. 3})$$

where,  $\text{KH}$  is the value of Henry's constant at operating pressure and temperature respectively (Carroll et al., 1991). Equation 4 was used to calculate the value of  $x_{D,e}$  at time  $t$  during the experiment.

$$x_{D,e} \text{ at time } t (\text{mole fraction}) = \frac{(\text{mol of } \text{CO}_2)_t}{(\text{mol of } \text{CO}_2 + \text{mol of } \text{H}_2\text{O})_t} \quad (\text{Eq. 4})$$

#### 3.3.2. Total number of moles of $\text{CO}_2$ consumed to form hydrate

Equation 5 was used to calculate the number of moles of  $\text{CO}_2$  consumed to form hydrate,  $\Delta n_{H,\downarrow}$  (Sun et al., 2014).

$$(\Delta n_{H,\downarrow})_t = (n_{\downarrow})_0 - (n_{\downarrow})_t - n_{D,t} \quad (\text{Eq. 5})$$

where,  $(n_{\downarrow})_0$  is  $\text{CO}_2$  consumed at time  $t = 0$ ,  $(n_{\downarrow})_t$  is the  $\text{CO}_2$  consumed at time  $t$  and  $n_{D,t}$  is the  $\text{CO}_2$  dissolved in water phase. Equation 6 was used to calculate the number of moles of  $\text{CO}_2$  dissolved in water phase,  $n_{D,t}$  (Servio & Englezos, 2001).

$$n_{D,t} = \frac{\left[ \left( \frac{x_{D,t} \times n_{water}}{1 - x_{D,t}} \right) \times \left( \frac{1}{x_{D,e}} - 1 \right) - n_{water} \right]}{\left[ 1 - \frac{1}{x_{D,e}} \right]} \quad (Eq. 6)$$

### 3.3.3. Conversion of water to hydrate (mol%)

Equation 7 was used to calculate the conversion of water to CO<sub>2</sub> hydrate.

Conversion of water to hydrate (mol%)

$$= \frac{\Delta n_{H,\downarrow} \times \text{hydration number}}{n_{H_2O}} \times 100\% \quad (Eq. 7)$$

where  $\Delta n_{H,\downarrow}$  is the number of moles of gas consumed for hydrate formation at the end of the experiment determined from the gas uptake and  $n_{H_2O}$  is the total number of moles of water in the system. The hydration number is the number of water molecules per guest molecule. The hydration number used for pure CO<sub>2</sub> gas system was 5.75 (Babu et al., 2015).

### 3.3.4. CO<sub>2</sub> uptake (mmol of CO<sub>2</sub>/g of H<sub>2</sub>O)

Equation 8 was used to calculate the amount of CO<sub>2</sub> uptake with respect to dry silica gel weight.

$$CO_2 \text{ uptake (mmol of } CO_2 \text{/g of } H_2O) = \frac{\Delta n_{H,\downarrow}}{\text{weight of } H_2O} \times 1000 \quad (Eq. 8)$$

### 3.3.5. Rate of hydrate formation (mmol CO<sub>2</sub> per g H<sub>2</sub>O per min)

Equation 9 was used to calculate the rate of hydrate formation.

$$\left( \frac{\Delta n_{H,\downarrow}}{\Delta t} \right) = \frac{(\Delta n_{H,\downarrow})_{t+\Delta t} - (\Delta n_{H,\downarrow})_t}{\Delta t}, \Delta t = 30 \text{ min} \quad (Eq. 9)$$

The rate of hydrate formation was reported every 30 minutes because the data was very large; the sampling time for data acquisition of pressure and temperature were taken every 5 s by the HPVA.

## 4. RESULTS AND DISCUSSION

Four sample preparation methods were investigated in this work namely methods 1, 2, 3 and 4. The full description of each method was previously described in *section 2.2*. Initially, the results of equilibrium moisture content from each method are presented in which the

confidence interval (CI) calculated was based on a 90% CI level and is presented in the results and discussion throughout this work when necessary. Then, the formation of hydrate in the HPVA was examined by studying the P-t curves obtained together with the study of CO<sub>2</sub> dissolution in water according to Henry's Law (Diamond & Akinfiev, 2003; Servio & Englezos, 2001; Carrol et al., 1991) and then followed by analysis of CO<sub>2</sub> uptake and rate of hydrate formation.

#### 4.1. Vigorous Stirring Increased the Amount of Equilibrium Moisture Content inside Silica Gel Pores with Minimum Error Bars

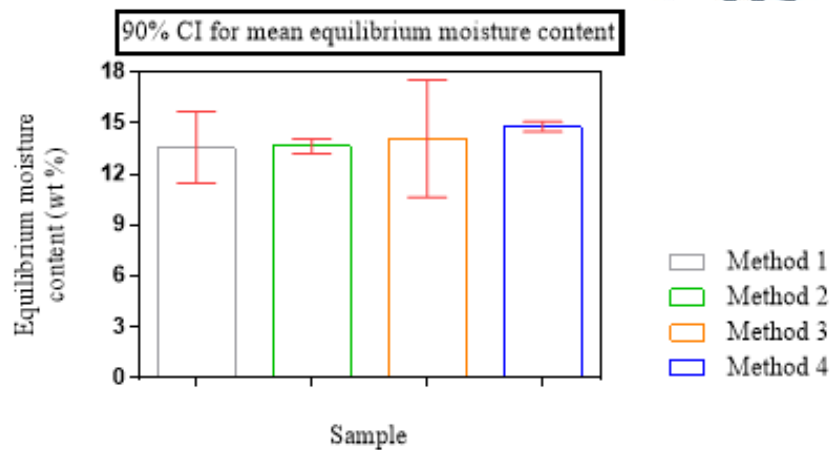
The amount of water available inside silica pores is vital for the formation of CO<sub>2</sub> hydrate in batch FBR. Thus, the calculated amount of water available for each prepared sample is first presented in Table 1.

**Table 1:** The equilibrium moisture content in wet silica gel for different types of sample preparation methods

Sample	Exp.	SiG (wet) (g)	SiG (dry) (g)	Equil. moisture content (wt%)	Mean equil. moisture content (wt%)	SD	90% C1 (±)
Method 1	1	0.06	0.05	15.05	13.56	2.11	2.45
	2	0.08	0.07	12.07			
Method 2	1	0.06	0.05	13.33	13.64	0.44	0.51
	2	0.09	0.07	13.95			
Method 3	1	0.05	0.04	16.52	13.83	2.47	2.87
	2	0.11	0.10	11.65			
Method 4	1	0.14	0.12	15.00	14.79	0.29	0.34
	2	0.10	0.08	14.58			

Previously, Babu et al. (2013) and Adeyemo et al. (2010) used a magnetic stirrer and Park et al. (2013) used ultrasonic wave during sample preparation with the same intention - to aid dispersion of water inside silica pores. However, the sample prepared by method 4 with vigorous stirring using a high-speed blender in this work had the highest equilibrium moisture content with the lowest CI which was  $14.79 \pm 0.34$  wt%. As for methods 1-3 in this work, the amount of equilibrium moisture content was almost identical which was around 13.68 wt% but the CI for method 2 (gentle stirring by magnetic stirrer) was the lowest among those three with

the value of  $\pm 0.51$  wt%. Both methods 1 and 3 without stirring demonstrated quite high CI values,  $\pm 2.45$  and  $\pm 2.87$  wt% respectively. The relatively low CI obtained for methods 2 and 4 is explained by the need for stirring during sample preparation to ensure the water is well distributed inside silica gel pores. These also indicated that samples prepared from both methods had high reproducibility as compared to methods 1 and 3. Figure 2 summarises the results shown in Table 1.

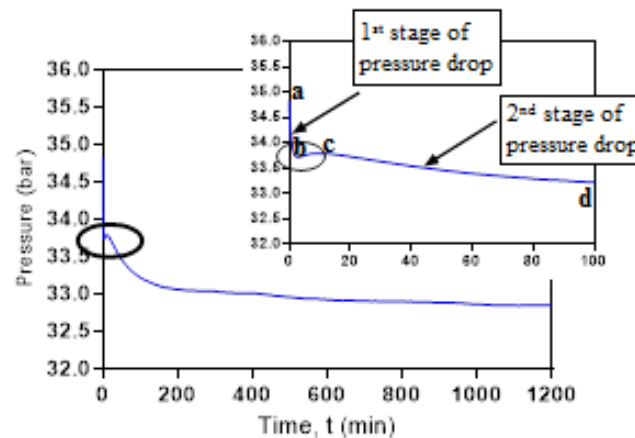


**Figure 2:** The equilibrium moisture content in wet silica gel for methods 1 – 4.

#### 4.2. Justification on CO<sub>2</sub> Hydrate Formation

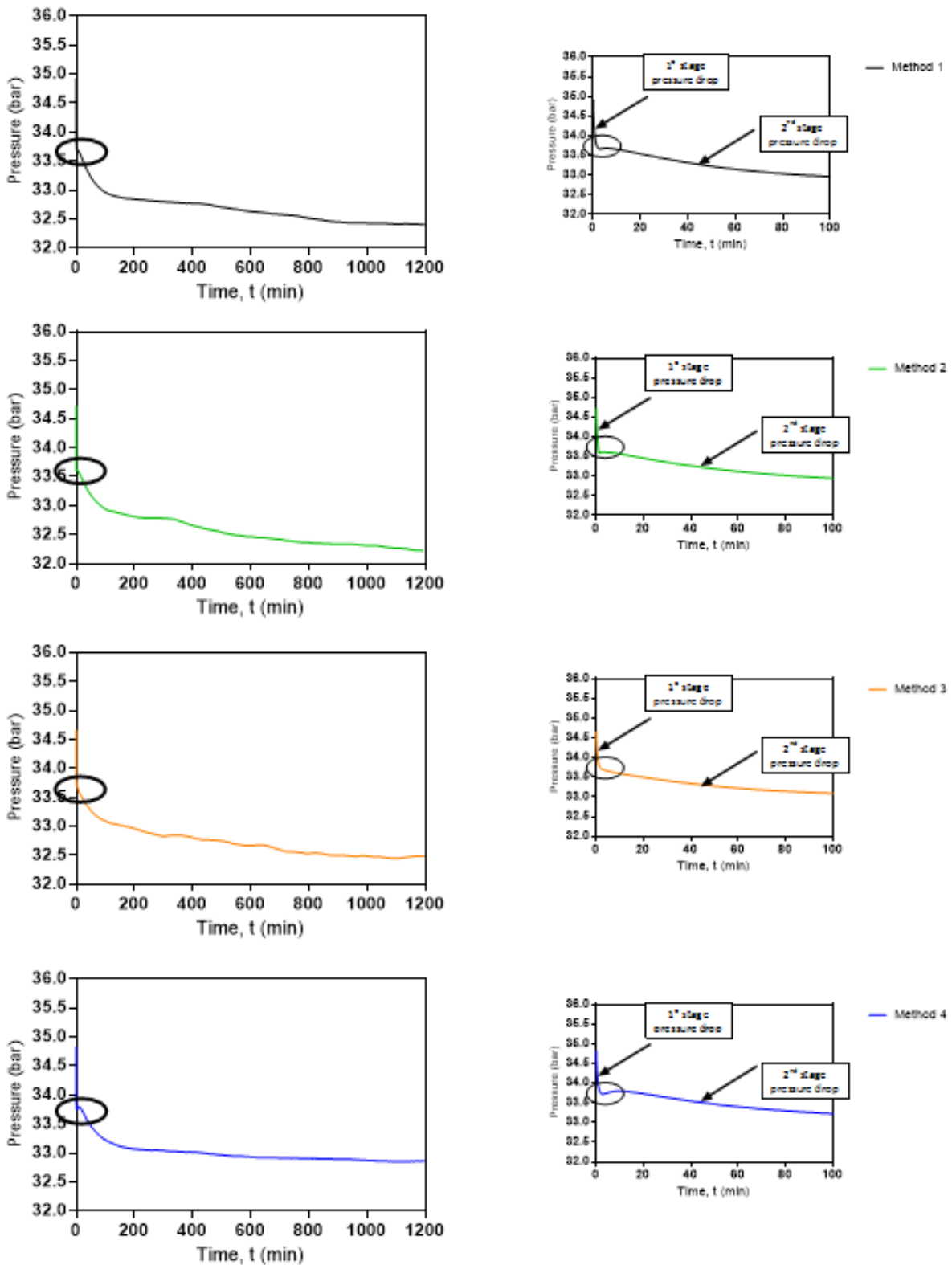
##### 4.2.1. Two-stage Pressure Drop of P-t Curve

Three sets of experiment were performed for each sample preparation method (methods 1-4) for assessing hydrate; only P-t curve for experiment number 1 for method 4 (Figure 3) is fully explained here. The analyses on other experiments were not presented because almost the same trends were observed for each method (Figure 4). All prepared samples with approximately 0.5 g of wet silica were used to investigate hydrate formation in the HPVA at 275 K and 35 bar by using pure CO<sub>2</sub> gas (99.99% purity). All experiments were conducted for 1200 minutes to obtain maximum CO<sub>2</sub> uptake.



**Figure 3:** P-t curves that show 2-stage pressure drop in 1200 minutes and the first 100 minutes at the experimental conditions of 35 bar and 275 K (method 4).

As seen in Figure 3, the two-stage pressure drop trend described by Tang et al. (2013) was observed during the hydrate formation experiment where the total pressure drop achieved after 1200 minutes was around 2 bar. The complete dissolution of CO<sub>2</sub> in water inside silica gel pores was observed after the pressure dropped approximately from 35 to 33.8 bar as indicated by point **a-c** (the first stage of pressure drop). Initially, point **a-b** indicates that the dissolution of CO<sub>2</sub> in water happened around 5 minutes in which Sloan & Koh (2008) stated that upon dissolution of gas in water, labile clusters form immediately. Concurrently, labile clusters started to agglomerate by sharing faces, thus increasing disorder which explained no pressure drop from point **b-c**. This process continued until the size of the cluster agglomerate reached a critical value at point **c** wherein Sloan & Koh (2008) said this was the point where primary nucleation happens. Also, Tang et al. (2013) described the time from point **a-c** as an induction time for hydrate formation. Moreover, the fast induction time observed in this work which was around 10 minutes by employing FBR agreed with the one reported in the literature (Zheng et al., 2016; Babu et al., 2014 & 2013; Linga et al., 2007). Then, the second stage of pressure drop was observed immediately after point **c** and this significant pressure drop is known as the hydrate growth stage until no more drop in pressure was observed in batch FBR as indicated by point **d**. Finally, several stages of pressure drop were observed after 100 minutes (point **d**) before it became a plateau started approximately at 1000 minutes.



**Figure 4:** P-t curves that show 2-stage pressure drop in 1200 minutes and the first 100 minutes respectively for methods 1 – 4

#### 4.2.2. CO<sub>2</sub> Solubility in Water Exceeded the Henry's Law Value

Since the experiments were conducted in batch mode or isochoric condition, the total number of moles of CO<sub>2</sub> consumed can easily be calculated by using the Ideal Gas Law. Then, the mole fraction of CO<sub>2</sub> dissolved in water was calculated to be 0.0438 where the value of Henry's constant at 275 K was obtained from Carrol et al. (1991). This value was not considered for the justification of hydrate formation because the value of Henry's constant in their work was calculated at atmospheric pressure which was not the same with the operating pressures employed in this work. However, this value was presented in the figure for the comparison purpose. Thus, the equilibrium mole fraction of CO<sub>2</sub> in water at various operating temperatures and pressures plotted by Servio & Englezos (2001) was used as a guideline to determine the formation of hydrate in the system. The formation of hydrate was confirmed when the total experimental mole fraction was higher than the equilibrium mole fraction at the experimental conditions (275 K and 35 bar). Based on the data given by Servio & Englezos (2001), the equilibrium mole fraction was found to be 0.0165 at these experimental conditions.



**Figure 5:** Mole fraction of CO<sub>2</sub> in water inside silica gel pores for methods 1 – 4 at experimental conditions of 275 K and 35 bar in 1200 minutes; (a) total mole fraction of CO<sub>2</sub>, (b) mole fraction of CO<sub>2</sub> during hydrate growth

Figure 5 (a) illustrates the total mole fraction of CO<sub>2</sub> consumed in water for all samples throughout the experiments while Figure 5 (b) illustrates the mole fraction of CO<sub>2</sub> consumed during hydrate growth. The red-dashed line in Figure 5 (a) shows the total CO<sub>2</sub> dissolved in

water at the experimental conditions in which further CO<sub>2</sub> consumed after that is known as the growth of hydrate. The highest total mole fraction of CO<sub>2</sub> involved in hydrate formation was observed for the sample prepared using method 4 (the highest rates of stirring of silica) (0.060) followed by methods 2 (the slowest rates of stirring of silica) (0.059), 1 (silica was left to naturally adsorbed moisture) (0.057) and 3 (silica was submerged in excess water) (0.050) as shown in Figure 5 (b). Since the formation of hydrate was justified, next the study on final water to hydrate conversion, CO<sub>2</sub> uptake and rate of hydrate formation for all samples are presented in the next section.

#### 4.3. Vigorous Stirring Enhanced the CO<sub>2</sub> Hydrate Formation inside Silica Gel Pores with Minimum Error Bars

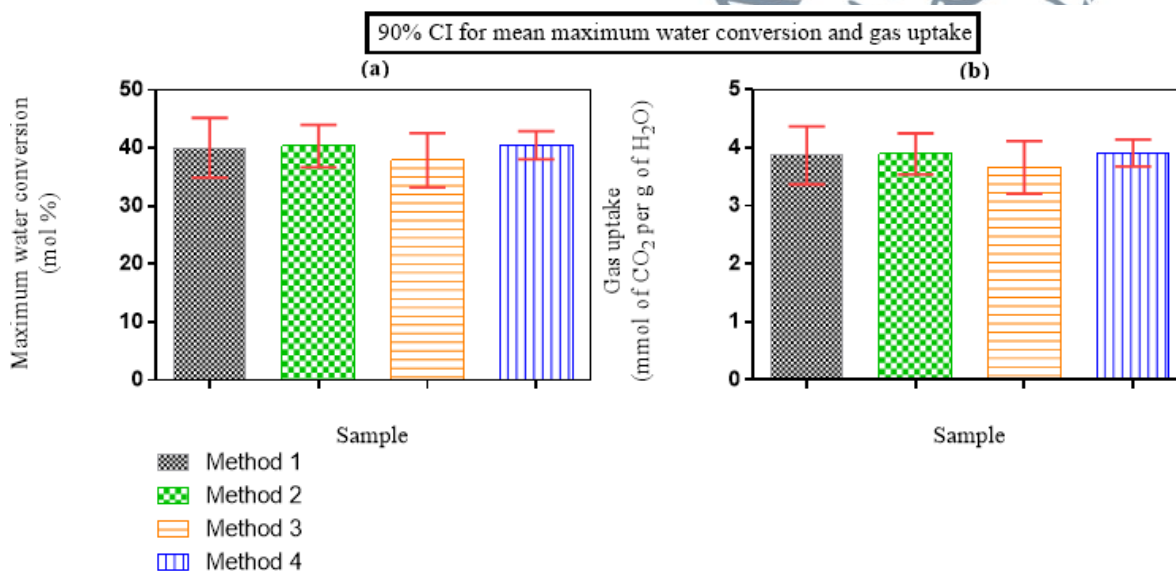
##### 4.3.1. Water Conversion to Hydrate and CO<sub>2</sub> Uptake

Table 2 shows the summary of results obtained where the experiments were repeated three times for each sample to analyse the reproducibility of the results.

**Table 2:** Summary of results for hydrate formation experiments performed at 275 K and 35 bar in 1200 minutes (samples prepared by Methods 1 – 4).

Sample	Exp.	Water conversion to hydrate (mol%)	Mean water conversion to hydrate (mol%) (90% CI)	SD	CO <sub>2</sub> uptake (mmol of CO <sub>2</sub> /g of H <sub>2</sub> O)	Mean CO <sub>2</sub> uptake (mmol of CO <sub>2</sub> /g of H <sub>2</sub> O) (90% CI)	SD	CO <sub>2</sub> formed in hydrate (mmol)	Mean CO <sub>2</sub> formed in hydrate (mmol)
Method 1	1	37.8	40.0 ± 4.84	5.1	3.7	3.9 ± 0.47	0.5	0.24	0.26
	2	36.4			3.5			0.23	
	3	45.9			4.4			0.30	
Method 2	1	36.8	40.3 ± 3.42	3.6	3.6	3.9 ± 0.38	0.4	0.24	0.27
	2	40.1			3.9			0.27	
	3	44.0			4.3			0.29	
Method 3	1	34.8	37.9 ± 4.46	4.7	3.4	3.7 ± 0.47	0.5	0.24	0.26
	2	43.3			4.2			0.29	
	3	35.6			3.4			0.24	
Method 4	1	40.6	40.5 ± 2.28	2.4	3.9	3.9 ± 0.19	0.2	0.29	0.29
	2	38.0			3.7			0.28	
	3	42.8			4.1			0.31	

Studying total water conversion to hydrate and gas uptake further showed the significance of vigorous stirring. The sample prepared by the highest rates of stirring (method 4) had the highest water conversion to hydrate with a value of  $40.5 \pm 2.28$  mol% followed by method 2 ( $40.3 \pm 3.42$  mol%), method 1 ( $40.0 \pm 4.84$  mol%) and method 3 ( $37.9 \pm 4.46$  mol%). It was observed that the sample with the highest equilibrium moisture content (method 4; Table 1) yielded the highest water conversion to hydrate. Consequently, the amount of gas uptake was directly related to the amount of water conversion to hydrate. Hence, the gas uptake obtained for methods 1, 2 and 4 were 3.9 mmol of CO<sub>2</sub>/g of H<sub>2</sub>O, 0.2 mmol of CO<sub>2</sub>/g of H<sub>2</sub>O higher than method 3. However, method 4 consumed the highest amount of CO<sub>2</sub> molecules with the value of 0.29 mmol followed by methods 2, 1 and 3 with values of 0.27, 0.26 and 0.26 mmol respectively. Figure 6 illustrates the water conversion to hydrate and CO<sub>2</sub> uptake for each sample.



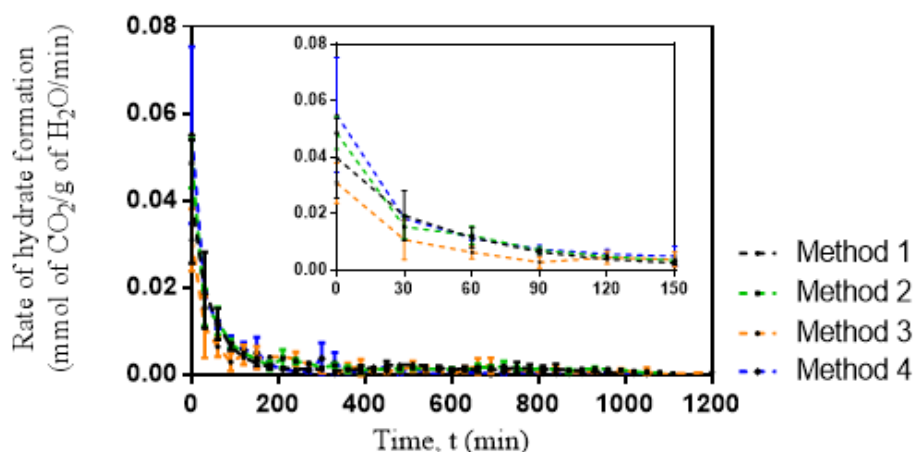
**Figure 6:** Comparison of (a) water conversion to hydrate and (b) gas uptake for methods 1 - 4 at the operating conditions of 275 K and 35 bar in 1200 minutes

The CI for samples with stirring process was lower than methods 1 and 3. High regeneration and repeatability (Table 2) were similar to past studies with consistent results using silica sand and PU foam (Babu et al., 2014; Babu et al., 2014(b); Mekala et al., 2014). Hence, the results observed explain the advantages of the vigorous stirring process during sample preparation as the amount of water that gets occluded inside silica gel pores (Babu et al., 2015) is overcome and beneficial for the utilisation of water in hydrate formation. Also, the results observed in this work agreed with the one obtained by Wang et al. (2008) wherein the formation of methane

hydrate for the sample (mixture of silica and water) stirred at the highest speed attained the highest gas uptake. Additionally, Rajnauth et al. (2012) mentioned that gas hydrate has a high gas to solid ratio where 1 ft<sup>3</sup> of hydrate contains 150-180 ft<sup>3</sup> of gas per ft<sup>3</sup> of water. Since the sample prepared by method 4 had the highest moisture content, this is the main explanation for the highest CO<sub>2</sub> uptake attained. The vigorous stirring of the silica mixture with water employed by several researchers (Dawson et al., 2014; Wang et al., 2008) and this work is useful because water droplets are surrounded by a network of silica which prevents coalescence of the droplets (Wang et al., 2008) thus increasing the amount of water inside silica pores. Wang et al. (2008) mentioned that the higher the mixing speed, the lower the average particle size which increased the gas-water interfacial surface area for hydrate formation due to the presence of dispersed water phase. Moreover, Yuan & Lee (2013) stated that a small contact angle was observed when the liquid spread on the surface, while a large contact angle was observed when the liquid beads (compact liquid droplet) exist on the surface. This contact angle is closely related to the spreadability or wettability (Zisman, 1964) in which a small contact angle (< 90°) indicates that wetting of the surface is favourable thus increasing the interaction between CO<sub>2</sub> gas-water with porous medium and as a result, enhancing hydrate formation. Yuan & Lee (2013) added that small droplets and bubbles are spherical, which gives a minimum surface area for a fixed volume and increases the surface tension of water. Thus, the use of vigorous stirring which dispersed water molecules is expected to reduce the surface tension of water inside silica pores.

#### 4.3.2. Rate of Hydrate Formation

Figure 7 illustrates the rate of hydrate formation in 1200 min and inset for the first 150 minutes. At 1200 min period, the rate of hydrate formation could not be seen clearly, thus the comparison was done for the first 150 min. The sample prepared by the highest rates of stirring (method 4) demonstrated the fastest kinetics in which the initial rate of hydrate formation was more than 0.05 mmol of CO<sub>2</sub>/g of H<sub>2</sub>O/min, followed by methods 2, 1 and 3. The slowest initial rate was demonstrated by method 3, almost 40% slower than method 4. The kinetics for all methods were almost the same after 120 minutes of the experiments. Thus, the initial rate of hydrate formation for the samples prepared with stirring (methods 2 and 4) was better than the methods without stirring (methods 1 and 3). Overall, the silica contacted with water from method 4 showed the best results.



**Figure 7:** Rate of hydrate formation for 1200 minutes and inset for the first 150 minutes (methods 1 - 4).

## 5. CONCLUSION

The solubility of CO<sub>2</sub> in water using Henry's Law and the experimental pressure–time (P-t) curve were analysed to determine the formation of hydrate. Hydrate formation was confirmed when the mole fraction of CO<sub>2</sub> dissolved in water exceeded the Henry's Law value as well as a two-stage pressure drop in the experimental P-t curve. Method 4 sample preparation with the use of vigorous stirring obtained the highest equilibrium moisture content ( $14.79 \pm 0.34$  wt%) which led to the greatest water conversion to hydrate ( $40.5 \pm 2.28$  mol%) in 1200 minutes. The high mixing speed (37,000 rpm) which resulted in the presence of dispersed water phase increased the gas-water interfacial surface area for hydrate formation. The repeatability of equilibrium moisture content and gas uptake attained for the sample prepared by the highest rates of stirring (method 4) was the greatest as compared to other methods (1,2 and 3) with the CIs of  $\pm 0.34$  wt% and  $\pm 0.19$  mmol of CO<sub>2</sub> per g of H<sub>2</sub>O respectively because the amount of water occluded inside silica gel pores was not an issue or in other words vigorous stirring increased the spreadability.

## REFERENCES

Adeyemo, A., Kumar, R., Linga, P., Ripmeester, J., Englezos, P. (2010). Capture of CO<sub>2</sub> from flue or fuel gas mixtures by clathrate crystallization in a silica gel column. *International Journal of Greenhouse Gas Control*, 4, 478-485.

Babu, P., Linga, P., Kumar, R., Englezos, P., (2015). A review of the hydrate based gas separation (HBGS) process for carbon dioxide pre-combustion capture. *Energy*, 185, 261-279.

Babu, P., Ho, C.Y., Kumar, R., Linga, P. (2014). Enhanced kinetics for the clathrate process in a fixed bed reactor in the presence of liquid promoters for pre-combustion carbon dioxide capture. *Energy*, 70, 664-673.

Babu, P., Kumar, R., Linga, P. (2014b). Unusual behaviour of propane as a co-guest during hydrate formation in silica sand: Potential application to seawater desalination and carbon dioxide capture. *Chemical Engineering Science*, 117, 342-351.

Babu, P., Kumar, R., Linga, P. (2013), Pre-combustion capture of CO<sub>2</sub> in a fixed bed reactor using the Clathrate hydrate process. *Energy*, 50, 364-373.

Biello, D. (2013). 400 ppm: Carbon Dioxide in the Atmosphere Reaches Prehistoric Levels. *Scientific American*, <http://blogs.scientificamerican.com/observations/2013/05/09>

Caitlyn, K. (2014). Climate change: Atmospheric Carbon Dioxide. National Centers for Environmental Information, <https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide>

Carrol, J., Slupsky, J., Mather, E. (1991). The solubility of CO<sub>2</sub> in water at low pressure. *Journal of Physic and Chemistry*, 20(6), 1201-1209.

D'Alessandro, D.M., Smit, B., Long, J.R. (2010). Carbon dioxide capture: Prospects for new materials. *Angewandte Chemie International Edition*, 49, 6058-6082.

Dawson, R., Stevens, L.A., Williams, O.S.A., Wang, W., Carter, B.O., Sutton, S., Drage, T.C., Blanc, F., Adams, D.J., Cooper, A.I. (2014). Dry bases: carbon dioxide capture using alkaline dry water. *Energy & Environmental Science*, 7, 1786-1791.

Diamond, L.W., Akinfiev, N.N. (2003). Solubility of CO<sub>2</sub> in water from -1.5 to 100 °C and from 0.1 to 100 MPa: evaluation of literature data and thermodynamic modelling. *Fluid Phase Equilibria*, 208, 265-290.

IEA Clean Coal Centre (2011). Pre-combustion capture of CO<sub>2</sub> in IGCC plants. No 11/14, United Kingdom: London.

IEA (2008). Energy technology perspectives. OECD/IEA Publications 9, France: Paris.

IEA (2006). Energy technology essentials. OECD/IEA 2006, ETE01.

Kunze, C., Spliethoff, H. (2012). Assessment of oxy-fuel, pre- and post-combustion-based carbon capture for future IGCC plants. *Applied Energy*, 94, 109-116.

Linga, P., Kumar, R., Englezos, P. (2007). Gas Hydrate formation from hydrogen/carbon dioxide and nitrogen/carbon dioxide gas mixtures. *Chemical Engineering Science*, 62, 4268-4276.

- McCallum, S.D., Riestenberg, D.E., Zatsepina, O.Y., Phelps, T.J. (2005). Effect of pressure vessel size on the formation of gas hydrates. *Journal of Petroleum Science & Engineering*, 56, 54-64.
- Mekala, P., Busch, M., Mech, D., Patel, R.S. (2014). Effect of silica sand size on the formation kinetics of CO<sub>2</sub> hydrate in porous media in the presence of pure water and seawater relevant for CO<sub>2</sub> sequestration. *Journal of Petroleum Science & Engineering*, 122, 1-9.
- Park, S., Lee, S., Lee, Y., Lee, Y., Seo, Y. (2013). Hydrate-based pre-combustion capture of carbon dioxide in the presence of a thermodynamic promoter and porous silica gels. *International Journal of Greenhouse Gas Control*, 14, 193-199.
- Rajnauth, J., Barrufet, M., Falcone, G., (2012). Hydrate formation: Considering the effects of pressure, temperature, composition and water. *Energy Science & Technology*, 4, 60-67.
- Servio, P., Englezos, P. (2001). Effect of temperature and pressure on the solubility of carbon dioxide in water in the presence of gas hydrate. *Fluid Phase Equilibria*, 190, 127-134.
- Sloan, E.D., Koh, C.A. (Third Edition) (2008). *Clathrate hydrates of natural gases*. Taylor & Francis Group USA: Boca Raton.
- Sun, D., Sun, Englezos, P. (2014). Storage of CO<sub>2</sub> in a partially water saturated porous medium at gas hydrate formation conditions. *International Journal of Greenhouse Gas Control*, 25, 1-8.
- Tang, J., Zeng, D., Wang, C., Chen, Y., He, L., Cai, N. (2013). Study on the influence of SDS and THF on hydrate-based gas separation performance. *Chemical Engineering Research and Design*, 91, 1777-1782.
- Wang, W., Bray, C.L., Adams, D.J., Cooper, A.I., (2008). Methane storage in dry water gas hydrates. *JACS Communications*, 130, 11608-11609.
- Yuan, Y., Lee, T.R. (2013). Contact angle and wetting properties. *Surface Science Technique*, 3-34.
- Zheng, J., Lee, Y.K., Babu, P., Zhang, P., Linga, P. (2016). Impact of fixed bed reactor orientation, liquid saturation, bed volume and temperature on the clathrate hydrate process for pre-combustion capture. *Journal of Natural Gas Science and Engineering*, 35, 1499-1510.
- Zisman, W.A. (1964). Relation of the equilibrium contact angle to liquid and solid constitution. *Advances in Chemistry*, 43, 1-5.

## DFT CALCULATIONS OF 3C- AND 2H-SiC STRUCTURAL AND ELECTRONIC PROPERTIES AND THEIR RESPONSE TO A HYDROSTATIC PRESSURE

Emhemed, A. M. M.<sup>1</sup> and Rosli, A. N.<sup>2</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1,2</sup>

ali55ly@yahoo.com

### ABSTRACT

Density functional theory (DFT) calculations have been used to investigate the structural and electronic properties of 3C-SiC and 2H-SiC and their response to hydrostatic pressure. The results obtained are in well agreement with the experimentally measured data and other theoretically calculated results. The bond lengths, angles, and interbilayer distances are calculated of the crystals. The calculations show that 3C-SiC has an identical tetrahedral bond lengths and angle with a high cubic symmetry. The off-plane bonds of 2H-SiC are shorter than the in-plane bonds whereas the off-plane angle was larger than the in-plane angle, and this implies that Si (C) atom was displaced slightly toward C (Si) atom in next bilayer. The electronic properties of the polytypes are investigated from the energy band structures. Partial density of states (PDOS) and total density of states (DOS) are calculated and analyzed with respect to the band gap structure. The electronic energy gap structures show that 2H-SiC has a wider band gap than 3C-SiC with clear differences in the conduction and valence bands characteristics. DOS at the conduction band edge of 3C-SiC (0.15e/eV) increases gradually whereas that of 2H-SiC (0.17e/eV) increases sharply. A brief discussion of the hydrostatic pressure effect on the structural and electronic properties of 3C and 2H silicon carbide polytypes is presented. The measurements show an increase in the band gap and DOS of SiC polytypes with increasing the pressure. The volume reduction ( $V_p/V_0$ ) of 3C-SiC is lower than the volume reduction of 2H-SiC, and the  $c/a$  ratio is almost constant of both crystals. The enthalpy increases with increasing the pressure. 3C appears to be more stable than 2H where the calculated enthalpy of 2H is higher than that of 3C.

**Keywords:** SiC polytypes, lattice structures, Electronic structures, pressure effect

### 1. INTRODUCTION

Silicon carbide (SiC) is a covalently bonded semiconductor where each silicon atom is bonded to four neighboring carbon atoms, or vice versa, to form a tetrahedron which is the main block of SiC. Each a close-packed structure of SiC crystal consists of Si-C double-atomic layer called Si-C bilayer which falls in one of three types of sites in arranging the Si-C bilayers named A, B, and C. These bilayers are stacking along c axis in different sequences forming a specific

polytype of SiC, and this character is called the polytypism. Silicon carbide appears in more than 200 polytypes, and the most known polytypes are cubic silicon carbide (3C-SiC) and some hexagonal polytypes such as 2H-, 4H-, and 6H-SiC. 3C-SiC and 2H-SiC are the only polytypes that have a pure cubic and hexagonal configurations respectively. Because of its superior properties, such as wide band gap, high-temperature resistance, chemical inertness, and breakdown field, silicon carbide has become one of the most candidate materials for future semiconductor electronic and optics devices. However, the properties and applicability of SiC polytypes are influenced strongly by the stacking sequence of the bilayers. These structures and some other properties were studied experimentally and theoretically by several researchers (Liu et al., 2010; Luo et al., 2012; Huang et al., 2015). Some of these studies were extended to investigate the effect of temperature and pressure on SiC properties (Zhuravlev et al., 2013; Varshney et al., 2015; Carlson et al., 2015). However, because of the limitation of the sample availability and the cost of theoretical simulations, most of the studies have been on 3C-SiC, 6H-SiC and very recently on 4H-SiC. Although the structure of 2H-SiC is less complex than 4H- and 6H-SiC, still there is a lack of the pressure effect investigations on its properties.

This paper focuses on investigating the structural and electronic properties of 3C-SiC and 2H-SiC polytypes, and a brief discussion of the pressure effect on these properties is presented. Furthermore, in order to evaluate our work, our results are compared with the experimentally measured data and other theoretically calculated results when available.

## 2. THE COMPUTATIONAL PROCEDURE

Based on density functional theory, the present first-principles calculations were performed using the Cambridge Serial Total Energy Package (CASTEP) code within the plane-wave pseudopotential approach, and the exchange-correlation energy was approximated using Perdew-Burke-Ernzerhof generalized gradient approximation (PBE-GGA) functional. These calculations were performed for the contributions from valence electrons ( $2S^2$  and  $2P^2$  for C, and  $3S^2$  and  $3P^2$  for Si) in order to achieve more accuracy and reduce the computational complexity. During the calculations, the plane-wave cut off energy was 240 eV. Special K-points sampling integration over the Brillouin zone (BZ) were employed with  $5 \times 5 \times 2$  mesh of hexagonal SiC polytypes, and  $5 \times 5 \times 5$  mesh of cubic SiC. This set of parameters uses the

tolerance in the self-consistent field calculation of  $2 \times 10^{-6}$  eV/atom, the change in total energy of  $2 \times 10^{-5}$  eV/atom, the maximum force of 0.05 eV/Å, the maximum stress of 0.1 GPa, and the maximum displacement of 0.002 Å. With these parameters, we calculated the electronic energy gap structure and the density of states of the SiC polytypes crystals. The geometric conformations of SiC crystals were optimized using the lattice constants and fractional coordinates of the internal atomic positions that listed in Table 1. The cubic structure of 3C-SiC crystal was built with space group F43m while the hexagonal structure of 2H-SiC crystal was built with space group P6<sub>3</sub>mc.

**Table 1:** SiC polytypes lattice constants and atomic fractional coordinates

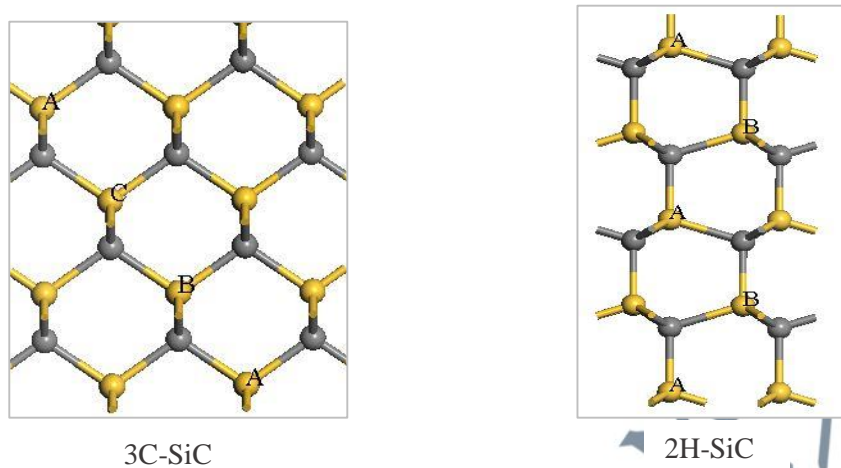
SiC Polytype	a (Å)	c (Å)	Fractional coordinates of atoms
3C-SiC	4.3596	4.3596	Si 0 0 0
			C $\frac{3}{4}$ $\frac{3}{4}$ $\frac{3}{4}$
2H-SiC	3.073	5.048	Si $\frac{1}{3}$ $\frac{2}{3}$ 0
			C $\frac{1}{3}$ $\frac{2}{3}$ $\frac{3}{8}$

### 3. RESULTS AND DISCUSSION

#### 3.1 The Crystal Structure of 3C- and 2H-SiC

Silicon carbide, as mentioned above, consists of a tetrahedron formed by bonding one Si atom to four C atoms, and Si-C bilayers are created when these tetrahedrons are stacked along C-axis. The different stacking sequence of the Si-C bilayers is the reason of polytypism character in SiC. Each bilayer in SiC can be situated in one of the three possible positions (A, B, or C) with respect to the lattice while maintaining the tetrahedral bonding scheme. The Si-C bonds in bilayer planes, depending on the stacking sequence, are either rotated by 60° with respect to their nearest neighbors to create a hexagonal nature, or appeared as mirror images and have a cubic nature. In our crystals, 3C-SiC has a cubic structure with stacking sequence ABCABC which is the only cubic SiC polytype, while 2H-SiC has a hexagonal structure with stacking sequence ABAB which is the simplest and the only polytype with pure hexagonal structure among SiC polytypes.

The optimized structures of the two polytypes are shown in Figure 1 where the crystals of 3C-SiC and 2H-SiC are composed of 621 atoms and 525 atoms respectively.



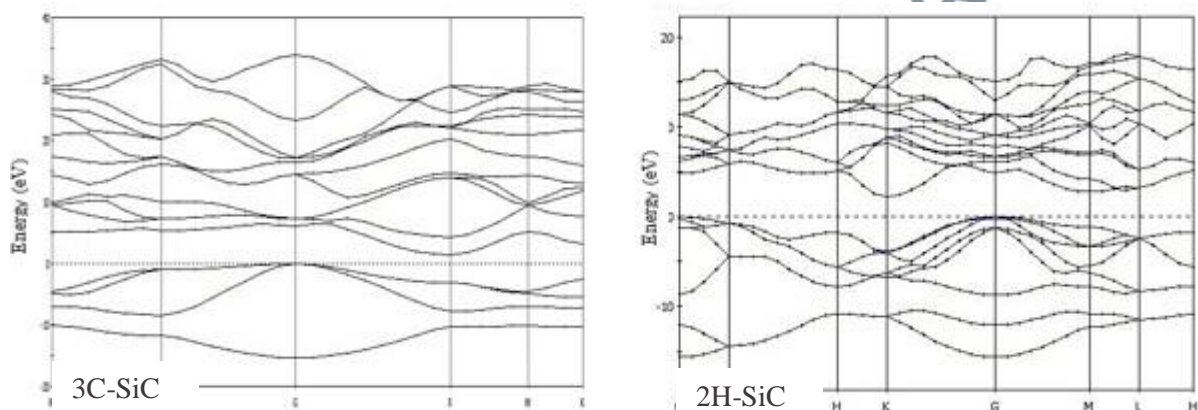
**Figure 1:** Crystal structures of SiC polytypes.

Figure 1 shows that 3C-SiC is a crystal with an identical orientation of Si-C tetrahedral where the orientation of the Si-C tetrahedral in the top layer coincides with that of the bottom layer. The length of Si-C bonds is 1.89 Å. The maximum deviation from symmetry is  $2 \times 10^{-15}$  ANG. The stacking sequence of 3C-SiC atomic layers has a local cubic environment (*k*-type) with respect to the immediate neighbors. The C-Si-C angle is  $109.471^\circ$  which agrees with the experimental measured identical tetrahedron angle (Capitani et al. 2007). The distance between A-B and B-C bilayers (Si1-Si2, Si2-Si3) is 3.09 Å, and the distance between A-C bilayers is 5.352 Å. On the other hand, the calculations of the length bonds of 2H-SiC crystal with 100% hexagonality showed that the in-plane bonds (Si1-C1, Si2-C2) are slightly longer with an average (1.893 Å) than the off-plane bonds (Si1-C2, Si2-C1) with an average (1.88307 Å). The distance between A-B bilayers (Si1-Si2) is 3.085 Å while the distance between A-A or B-B bilayers is 5.048 Å. The in-plane angle ( $C1-Si1-C1=109.364^\circ$ ) is smaller than the off-plane angle ( $C1-Si1-C2=109.578^\circ$ ), and the identical tetrahedron angle value falls in the middle between those two measured angles which affirms the crystal symmetry. The calculated values of 2H bonds and angles indicate to a slight displacement of Si (C) atom toward C (Si) atom that belongs to next bilayer. These calculated lattice parameters agree well with the results obtained by Nowak (2001). The slight differences in the bond lengths and angles between 3C- and 2H-SiC crystals can affect their properties strongly, for instance, the short bonds enhance

the hardness and increase the width of the band gap. However, the hexagonality also plays an important role in defining the crystal properties (Luo et al. 2012; Huang et al. 2015).

### 3.2 The Electronic Properties of SiC (3C and 2H) Polytypes

The electronic energy band structures of the SiC crystals in the energy range (-15 to 15) eV along the high-symmetry directions in the first Brillion zone (BZ) are depicted in Figure 2.

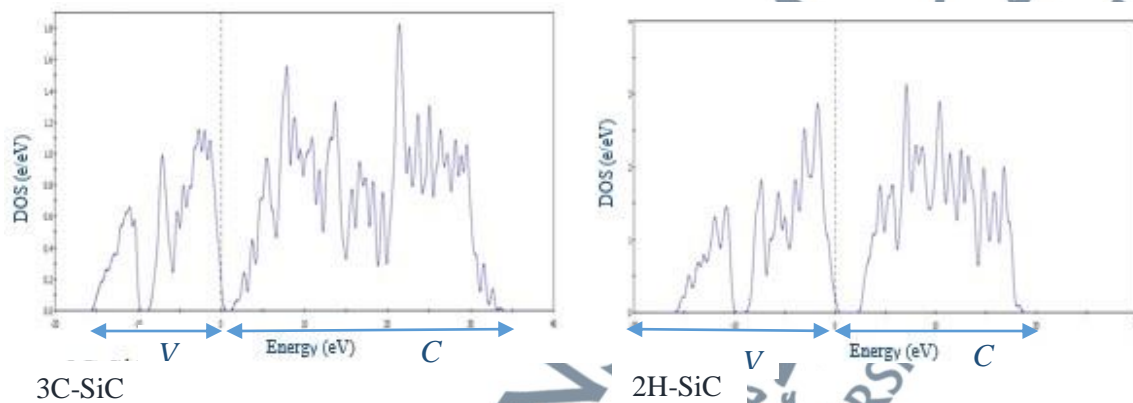


**Figure 2:** Band-gap structure of SiC polytypes.

The width of the band gap is estimated from the differences between the highest occupied orbital in the valence band and the lowest unoccupied orbital in the conduction band. All energies are with reference to the top of valence band for 3C and 2H polytypes where the top of occupied states in the valence band is situated at G point with Zero value of the energy due to the automatically shifting process, and this facilitates the evaluation of the band gap.

It is well known that DFT based on local density approximation (LDA) or generalized gradient approximation (GGA) underestimates the band gap calculations by (40-60) % because of the Kohn-Sham theory that systematically underestimates photoemission gaps. Therefore, the band gaps of 3C-SiC and 2H-SiC crystals are corrected by adding the quasiparticle correction (1.13 eV) of the calculated values, so that the band gaps were found to be 2.4 eV and 3.37 eV of 3C-SiC and 2H-SiC respectively. These results agree well with the experimental data and others calculations (Nowak, 2001; Laref & Laref, 2008), and better than those of Huang et al. (2015) calculations. It can be seen from the band gap diagram of 3C-SiC that the first minima of the conduction band is located at X point, and the second minima is at L point with about 3.14 eV

difference. The direct band gap is 6.1 eV which agrees with Huang et al. (2015) results. Whereas the first minima in the conduction band of 2H-SiC is located at K point while the second minima is located at M point with a difference of 0.76 eV which is very small when compared with that of 3C-SiC, and the direct gap of 2H-SiC (4.95 eV) is narrower than 3C-SiC direct gap. The conduction band minima in cubic SiC at X point is folded onto  $2/3$  LM in 2H hexagonal BZ. This point has a lower symmetry, and the bonding combinations of C 2s orbital and a Si 3p orbital can interact with closer lying states. On the other hand, the minima at K point gives rise to the lowest empty band. The electronic energy gap structure can be viewed and discussed with DOS diagrams. Figure 3 shows the total density of states (DOS) of 3C-SiC and 2H-SiC polytypes.



**Figure 3:** Density of states of SiC polytypes.

It can be seen that the width of 3C-SiC and 2H-SiC valence (V) bands are almost same with a lower width of 2H-SiC. In the 3C-SiC structure, there is a subband gap appears between (-8.75) eV and (-9.75) eV where the lower band is dominated by S orbitals while the upper band is P-orbital dominated. The higher peak at (-7) eV comes from the onset of the second valence band at X and L points where the states at the beginning are dominated by S electrons and rapidly changed to P electrons dominated. The subband in 2H-SiC valence band is located between (-9) eV and (-10) eV where the lower band is dominated by S electrons while the upper is dominated by P electrons. The large peak located at the onset of the upper band appears at (-7.5) eV which also rapidly dominated by P-state electrons.

The calculations showed that the valence bandwidth of 3C-SiC is slightly higher than that of 2H-SiC which agrees well with Laref & Laref (2008) values. However there is a clear variance

between these polytypes in the conduction band ( $C$ ) width and structure, and this means that the reason for the band gap variations between the polytypes comes mainly from the conduction band differences where the conduction band of cubic polytype is wider than that of the hexagonal polytype. Another interesting point is that the minima location of the conduction band moved from  $X$  point in cubic SiC, which folded between  $M$  and  $L$  in a hexagonal structure, to  $K$  point in hexagonal SiC which is located in the center of the  $BZ$  edge parallel to  $c$ -axis. The DOS at the edge of the conduction band of 3C-SiC is 0.15 e/eV and increases gradually, while the DOS at 2H-SiC conduction band edge is 0.17 e/eV and increases sharply, and this value consists very well with the calculated value by Nuruzzaman (2015). These variations are generated mainly from the diversity in the stacking sequence between the two polytypes.

### 3.3 *The Pressure Effect on 3C-SiC and 2H-SiC Properties*

While investigating the pressure effect on SiC polytypes crystal structures, the cut-off energy was changed to 310 eV, and the  $K$ -point sets became  $5 \times 5 \times 4$  mesh for the hexagonal SiC and  $4 \times 4 \times 4$  mesh for the cubic SiC, so that the ultrafine quality of the functional was used to enhance the calculations accuracy. This work presents results of DFT measurements of the band gap, DOS, and volume changes of 3C-SiC and 2H-SiC crystals under pressures up to 70 GPa at zero temperature.

The results showed a decrease in the band gap width of 3C-SiC crystal. Also, the trend between the density and pressure matched what was expected where the density increased with pressure increase. The highest density was obtained at pressure 70 GPa, and the lowest density was obtained at pressure 10 GPa. These findings agree well with Carlson (2015) and Guedda (2016). The volume of the crystal decreased when the pressure increased which consists of Zhuravlev (2013) experimental findings, and  $(c/a)$  ratio was almost constant where the hydrostatic pressure unchanged the original cubic symmetry of the unit cell. The volume collapse ( $V_p/V_0$ ) was determined and found to be 16% which is comparable to experiments data and others calculations (Lu et al., 2008; Varshney et al., 2015). The band gap width of 2H-SiC crystal became wider, and the density became higher with increasing the pressure. The measurements showed that the crystal volume of 2H-SiC decreased with increasing the

pressure where the calculated volume collapse was about 20%. The equilibrium ratio of  $c/a$  was 1.64 which is very close to the identical value. These results agreed very well with those of Karch et al. (1996).

Since the hydrostatic pressure kept the cubic symmetry unchanged of 3C-SiC. The strong covalent Si-C bonds lie in [111] direction. Therefore the effect of the compression along  $c$  direction was weak, and the reduction volume of 3C-SiC (16%) was found to be lower than that of 2H-SiC which coincides with Guedda results (Guedda et al., 2016).

The crystal stability at a given temperature and pressure is governed by what is called Gibbs free energy or the enthalpy (Karch et al., 1996; Varshney et al., 2015). In our calculations, we measured the enthalpy of 3C and 2H under hydrostatic pressure effect and zero temperature. The calculations of the SiC crystals showed an increase in the enthalpy with increasing the pressure which agrees with the fact that the most thermodynamically stable phase at a given pressure and temperature is the one with the lowest enthalpy.

#### 4. CONCLUSION

The structural and electronic properties of silicon carbide (3C-SiC and 2H-SiC) polytypes were investigated employing density functional theory (DFT) calculations with CASTEP code. The calculations showed that the different stacking sequences of the SiC tetrahedron gave a rise of diversity in the structure parameters which caused a variation in the configuration nature of the bilayers between those polytypes. The structure of 3C-SiC showed identical tetrahedral bond lengths and angle with a high cubic symmetry. The off-plane bonds of 2H-SiC were shorter than the in-plane bonds whereas the off-plane angle was larger than the in-plane angle, and this implies that Si (C) atom was displaced slightly toward C (Si) atom in next bilayer. The slight differences in the bond lengths and angles between the crystals of SiC polytypes can affect their properties strongly, for instance, the short bonds enhance the hardness and increase the width of the band gap. However, the hexagonality also plays an important role in defining the crystal properties. The position of the conduction band minima in the band gap structure is different for the polytypes. The electronic energy gap structures showed that 2H-SiC has a wider band gap than 3C-SiC with clear differences in the conduction and valence bands

characteristics. The electronic properties of SiC polytypes, such as the transition process from the valence band to conduction band, are controlled by the band energy structure. The measurements of the pressure effect on the SiC crystals showed an increase in the band gap and DOS of the crystals, and the volume reduction ( $V_p/V_0$ ) of 3C-SiC was lower than the volume reduction of 2H-SiC with almost a constant  $c/a$  ratio of both crystals. 3C appeared to be more stable than 2H where the calculated enthalpy of 2H was higher than that of 3C.

## REFERENCES

- Capitani, G. C., Di Pierro, S., & Tempesta, G. (2007). The 6H-SiC structure model: Further refinement from SCXRD data from a terrestrial moissanite. *American Mineralogist*, 92(2-3), 403-407.
- Carlson, T. A., Welch, C. R., Kriven, W. M., & Marsh, C. P. (2015). Effects of time, temperature, and pressure on the microstructure of spark plasma sintered silicon carbide. *Journal of Ceramic Processing Research*, 16(3), 303-307.
- Guedda, H. Z., Ouahrani, T., Morales-García, A., Franco, R., Salvadó, M. A., Pertierra, P., & Recio, J. M. (2016). Computer simulations of 3C-SiC under hydrostatic and non-hydrostatic stresses. *Physical Chemistry Chemical Physics*, 18(11), 8132-8139.
- Huang, Z., Lü, T. Y., Wang, H. Q., & Zheng, J. C. (2015). Thermoelectric properties of the 3C, 2H, 4H, and 6H polytypes of the wide-band-gap semiconductors SiC, GaN, and ZnO. *AIP Advances*, 5(9), 097204.
- Karch, K., Bechstedt, F., Pavone, P., & Strauch, D. (1996). Pressure-dependent properties of SiC polytypes. *Physical Review B*, 53(20), 13400.
- Laref, A., & Laref, S. (2008). Electronic and optical properties of SiC polytypes using a transferable semi-empirical tight-binding model. *Physica Status Solidi (B)*, 245(1), 89-100.
- Liu, L., Yiu, Y. M., Sham, T. K., Zhang, L., & Zhang, Y. (2010). Electronic structures and optical properties of 6H-and 3C-SiC microstructures and nanostructures from X-ray absorption fine structures, X-ray excited optical luminescence, and theoretical studies. *The Journal of Physical Chemistry C*, 114(15), 6966-6975.
- Lu, Y. P., He, D. W., Zhu, J., & Yang, X. D. (2008). First-principles study of pressure-induced phase transition in silicon carbide. *Physica B: Condensed Matter*, 403(19), 3543-3546.
- Luo, X., Goel, S., & Reuben, R. L. (2012). A quantitative assessment of nanometric machinability of major polytypes of single crystal silicon carbide. *Journal of the European Ceramic Society*, 32(12), 3423-3434.

Nowak, S. (2001, August). Crystal lattice dynamics of various silicon-carbide polytypes. In International Conference on Solid State Crystals 2000 (pp. 181-186). International Society for Optics and Photonics.

Nuruzzaman, Md., Ariful Islam, M., Ashraful Alam, M., Hadi Sah, M. A., & Tanveer Karim, A. M. M. (2015). Structural, elastic and electronic properties of 2H- and 4H-SiC. International Journal of Engineering Research and Applications, 5(5), 48-52.

Varshney, D., Shriya, S., Varshney, M., Singh, N., & Khenata, R. (2015). Elastic and thermodynamical properties of cubic (3C) silicon carbide under high pressure and high temperature. Journal of Theoretical and Applied Physics, 9(3), 221-249.

Zhuravlev, K. K., Goncharov, A. F., Tkachev, S. N., Dera, P., & Prakapenka, V. B. (2013). Vibrational, elastic, and structural properties of cubic silicon carbide under pressure up to 75 GPa: Implication for a primary pressure scale. *Journal of Applied Physics*, 113(11), 113503.

UNIVERSITI SAINS ISLAM  
جامعة العلوم الإسلامية  
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

## A ROBUST RIDGE REGRESSION FOR MULTICOLLINEARITY PROBLEM IN THE PRESENCE OF OUTLIERS IN THE DATA

Ferdaos, N. A.<sup>1</sup> and Mohamad Shariff, N. S.<sup>2</sup>

Faculty of Science & Technology (FST), Universiti Sains Islam Malaysia, Nilai  
Negeri Sembilan, Malaysia<sup>1,2</sup>

nuraqilahferdaos@yahoo.com<sup>1</sup>

### ABSTRACT

*Multicollinearity and outlying points are common problems in multiple regression analysis, which produce inconsistent and unreliable estimates of the model parameter and subsequently the inference of parameter. The well-known method to combat multicollinearity problem is the ridge regression method. However, this procedure is affected by the presence of outliers where it can cause fatter tails in the error distribution than the normal distribution. Therefore, it is important to combine methods of estimation that is more robust to deal with these problems simultaneously. In this paper, the performance of the proposed method is discussed and comparative investigation is made between existing estimators namely, Ordinary Least Squares (OLS), ridge regression (R), and robust ridge regression based on GM-estimates. The initial finding of this study is expected to produce a reliable parameter estimates in the presence of both anomalies in the data.*

**Keywords:** multicollinearity, outlying points, ridge regression, robust method

### 1. INTRODUCTION

In regression analysis, the Ordinary Least Squares (OLS) estimation is widely used to investigate the linear relationship between the variables of interest because it has the most favourable properties and easy to compute. It is also the Best Linear Unbiased Estimator (BLUE) when it follows two assumptions; i) independency of explanatory variables and ii) normality of error distribution. However, when multicollinearity and outliers are present in the data, these assumptions are invalid due to the increased in standard errors of the estimated coefficients in regression analysis. Multicollinearity is defined as a phenomenon where the explanatory variables are highly dependent with each other, while outliers is defined as an observation that is very far away from the entire observed data. When OLS estimator suffers from multicollinearity problem, the situation will be worsen in the presence of outliers in the data. The most suitable method to overcome multicollinearity is ridge regression method;

however in the presence of outliers in the data, it will cause fatter tails in the error distribution rather than in the normal distribution. This will produce inconsistent and unreliable parameter estimates that results in incorrect inference of the parameter and model.

There are some studies to illustrate the application of robust ridge regression in the literatures. For examples, Bagheri & Midi (2009) use Gujarati's (2003) child mortality data set to evaluate the performance the proposed robust methods which are GM-DRGP-L1, GM-DRGP-LTS, M-DRGP, MM-DRGP and DRGP-MM. The final results produce a reasonable interpretation that the proposed robust estimation process are able to solve multicollinearity problem caused by high leverage points. Another study done by Pati et al. (2015) evaluated the performance of proposed method in the presence of multicollinearity and outliers by using data set of Hald (1952). This data contains 4 variables and 13 observations that have 3 outliers. The study shows that the VIF of each variable are extremely high indicating a present of high level of multicollinearity in the data. In this study, they are able to conclude that the use of robust method is suitable to provide a more stable parameter estimates in the presence of multicollinearity and outliers in the data.

The main issue that arises in regression analysis is multicollinearity especially when modelling the relationship with a large number of explanatory variables. The existing ridge estimator is robust towards the multicollinearity problem; however it is believed that this estimator has a limitation where it is being influenced by the presence of outliers. To overcome the problem involved in modelling multicollinearity in the presence of outliers, a robust version of the ridge estimator is introduced in this study. This method is adopted from Bagheri & Midi (2009) technique with the selected proposed robust estimates. The performance of the proposed procedure is examined via Monte Carlo simulation study and compared with the existing methods by using the coefficient bias.

## 2. METHOD

### 2.1 Ridge Regression Model

Let the linear regression model be;

$$Y = Xb + e \quad (1)$$

where  $Y$  is a vector response (dependent) values,  $X$  represents  $n \times p$  of explanatory (independent) variables with rank  $p$ ,  $b$  is the vector of  $p$  coefficient for explanatory variables and  $e$  are random errors with the assumptions of zero mean,  $E(e) = 0$  and a constant variance,  $Var(e) = \sigma^2 I_n$ . Under OLS assumptions, the estimates of  $b$  in equation (1) is computed as  $\hat{b} = (X^T X)^{-1} X^T Y$ .

However, when multicollinearity exists, this OLS estimates is unreliable due to large variation in  $b$ . It is said to be inconsistent. As such, Hoerl & Kennard (1970a) introduced a new method of estimation and it is called ridge estimator. This method is adapted from the OLS estimator by adding  $kI_n$  in  $(X^T X)^{-1}$ ;

$$\hat{b}_R = (X^T X + kI_n)^{-1} X^T Y \quad (2)$$

The estimated value  $k$  in (2) will reduce the dependency in explanatory variables  $X$  (Mansson et al., 2010). In the presence of outliers, the response variable  $Y$  will subsequently affect  $\hat{b}_R$  and will cause unreliable parameter estimates. Therefore, a robust version of ridge estimator is introduced to overcome this problem.

## 2.2 Ridge Regression Estimators

Over the years, there are many studies have been proposed to estimate the value of  $k$ , for example; see Hoerl & Kennard (1970), McDonald & Galarnau (1975), Kibria (2003), Khalaf & Shukur (2005), Alkhamisi et al. (2006), Alkhamisi & Shukur (2008), Muniz & Kibria (2009), and recently Mansson et al. (2010). Different methods will yield different estimates in (2). Thus, based on the performance in simulation study of Mansson et al. (2010), the technique from Kibria (2003) and Muniz & Kibria (2009) are selected to estimate the value of  $k$  in this study.

Kibria (2003) proposed the following  $k$ ;

$$k_2 = \frac{\hat{\sigma}^2}{\prod_{i=1}^p \hat{\beta}_i^2} \quad (5)$$

$$k_3 = \text{median}\{m_i^2\} \quad (6)$$

Based on the previous estimator, Muniz & Kibria (2009) proposed the following;

$$k_4 = \left( \prod_{i=1}^p \frac{1}{m_i} \right)^{\frac{1}{p}} \quad (7)$$

$$\text{where } m_i = \sqrt{\frac{\hat{\sigma}^2}{\hat{\beta}_i^2}} \quad (8)$$

### 2.3 Robust Ridge Regression

Generalized M-estimator (GM) provides a good parameter estimate in (1) by filtering outliers that are present in both X and Y- direction by using the weight function. The GM estimator is commonly computed as  $\hat{b}_{Rob} = (X^T W X)^{-1} X^T W y$ . Due to the problem involved in the presence of outliers, the technique of Bagheri & Midi (2009) is adopted in this study, where the robust ridge regression is obtained by introducing  $k^*$  in equation (2);

$$\hat{b}_{RRobR} = (X^T X + k^* I_n)^{-1} X^T X \hat{b}_{Rob} \quad (9)$$

$k^*$  is computed as  $k^* = \frac{p \cdot \hat{\sigma}_{Rob}}{\hat{b}_{Rob}^T \hat{b}_{Rob}}$ , where  $p$  and  $\hat{\sigma}_{Rob}$  are the number of explanatory variables

and robust scale respectively.  $\hat{\sigma}_{Rob}$  is the Median Absolute Deviation (MAD) and computed as

$$\hat{\sigma}_{Rob} = 1.4825 \text{median}|\hat{e} - \text{median}(\hat{e})|. \hat{e} \text{ are the estimated errors and obtained via } \hat{e} = Y - X \hat{b}_{Rob}$$

. To study more about the performance of  $k^*$  in the presence of both multicollinearity and

outliers in the data, the  $k^*$  in (9) is being replaced by the different estimators ( $k^* = k_2, k_3, k_4$ )

chosen for this study.

### 2.4 Monte Carlo Simulation Study

Let the following model be;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad \text{where } i = 1, 2, \dots, n \quad (10)$$

and  $\varepsilon_i \sim iidN(0,1)$ . The parameter values are set as;  $b_0 = b_1 = b_2 = 1$  and  $x_{ij} = (1 - \rho^2)z_{ij} + \rho z_{ij}$ , where  $\rho$  is the correlation between  $x_{ij}$  and  $z_{ij} \sim iidN(0,1)$ . The correlations chosen in this study are fixed at 0.00, 0.35 and 0.95 and represent as; no multicollinearity, moderate and high level of multicollinearity. These degree of correlation are chosen to describe the variability of dependency among independent variables in studying the performance of all methods. These values are set to be closed to Pati et al. (2015). The sample size chosen are  $n = 50, 100$  with 1000 number of simulations.

When the outliers are present at time  $t = \tau$ , the residual takes the following form;

$$\varepsilon_i = \begin{cases} \varepsilon_i & \text{for } i \neq \tau \\ \varepsilon_i + m_i & \text{for } i = \tau \end{cases} \quad \text{for } i = 1, 2, \dots, n; \text{ with } m_i \sim U[0,10] \text{ and 1\% and 5\% contaminations}$$

are chosen randomly to illustrate the presence of outliers for  $n = 50, 100$ , respectively.

In this study, the performances of these procedures are measured by using the coefficient bias; the different between fitted and true values and it is computed as follow;

$$bias = E(\hat{b} - b) \quad (11)$$

The estimated coefficient is larger than the true parameter value when the bias is positive, while the estimated coefficient is smaller when the bias is negative.

### 3. RESULTS AND DISCUSSION

Table 1 shows the simulation results of biasness for parameter estimates ( $\beta_1$  and  $\beta_2$ ) for both  $n = 50$  and 100 in the presence of multicollinearity in the data. The performance of OLS, ridge regression with  $k = k_2, k_3, k_4$ , GM estimator, GM with  $k = k^*, k_2, k_3, k_4$  are shown in the table below. For  $n = 50$  with 1% outliers, with no multicollinearity ( $\rho = 0$ ) in the data, the estimators that provides the smallest coefficient bias are; OLS, ridge regression with  $k_4$ , GM estimator, robust ridge regression with  $k^*$  and GM estimator with  $k_4$  result in low coefficient bias yielding the favourable estimate among others. The results shown for moderate and high multicollinearity are GM estimator and GM estimator with  $k_4$  which show a good result,

respectively. As the sample size increases to  $n=100$  with 5% contamination in the data, the GM estimator with  $k_4$  has the smallest coefficient bias, therefore it is said as the best estimator in this study.

#### 4. CONCLUSION

This study proposes as an alternative approach to deal with multicollinearity in a contaminated data by using a robust version of ridge regression and the study found that the proposed test is robust towards the effect of multicollinearity and outliers in the data. It is believed that this estimator produces unbiased estimates in the presence of both multicollinearity and outliers and yet confirmed with the simulation results. The result have shown that the proposed test (GM estimator with  $k = k_4$ ) outperformed other estimator when multicollinearity and outliers are present simultaneously in the data

**Table 1:** Bias result for contaminated data.

$n$	$\rho$		0	0.35	0.95
	Estimator	Coefficient	Bias		
50	OLS	$\beta_1$	-0.0122	0.0038	-0.0118
		$\beta_2$	0.0133	0.0278	0.0433
	R	$\beta_1$	0.9863	0.9844	0.9885
		$\beta_2$	0.9867	0.9847	0.9885
	R3	$\beta_1$	0.9732	0.9610	0.9777
		$\beta_2$	0.9744	0.9713	0.9786
	R4	$\beta_1$	-0.0101	0.0241	0.0145
		$\beta_2$	0.0087	0.0595	0.0400
	GM	$\beta_1$	-0.0204	-0.0032	0.0061
		$\beta_2$	0.0359	0.0243	0.0112
	ROB	$\beta_1$	0.0029	0.0112	0.0195
		$\beta_2$	0.0340	0.0248	0.0296

100	ROB2	$\beta_1$	0.9866	0.0092	0.0412
		$\beta_2$	0.9868	0.0517	0.0383
	ROB3	$\beta_1$	0.9746	-0.0058	0.0194
		$\beta_2$	0.9764	0.0367	0.0172
	ROB4	$\beta_1$	-0.0068	-0.0230	-0.0064
		$\beta_2$	0.0060	0.0350	0.0297
	OLS	$\beta_1$	-0.0003	-0.0194	-0.0068
		$\beta_2$	0.0158	0.0470	0.0065
	R	$\beta_1$	0.9942	0.9981	0.9985
		$\beta_2$	0.9942	0.9982	0.9985
	R3	$\beta_1$	0.9885	0.9963	0.9971
		$\beta_2$	0.9886	0.9965	0.9971
	R4	$\beta_1$	0.0122	0.0028	0.0086
		$\beta_2$	0.0131	0.0285	0.0113
	GM	$\beta_1$	0.0174	-0.0057	0.0069
		$\beta_2$	0.0095	0.0196	0.0258
	ROB	$\beta_1$	0.0058	0.0075	0.0170
		$\beta_2$	0.0114	0.0107	0.0462
	ROB2	$\beta_1$	0.9942	0.0166	0.0108
		$\beta_2$	0.9942	0.0200	0.0355
ROB3	$\beta_1$	0.9888	0.0090	0.0156	
	$\beta_2$	0.9891	0.0122	0.0369	
ROB4	$\beta_1$	0.0034	-0.0183	0.0091	
	$\beta_2$	0.0333	0.0316	-0.007	

R is ridge regression with  $k = k_2$ , R3 is ridge regression with  $k = k_3$ , R4 is ridge regression with  $k = k_4$ . ROB is GM estimator with  $k = k^*$ , ROB2 is GM estimator with  $k = k_2$ , ROB3 is GM estimator with  $k = k_3$  and ROB4 is GM estimator with  $k = k_4$ .

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the grant support (RAGS-FST-50214-55) from the Universiti Sains Islam Malaysia (USIM), Malaysia

## REFERENCES

- Alkhamisi, M., Khalaf, G., & Shukur, G. (2006). Some modifications for choosing ridge parameters. *Communications in Statistics—Theory and Methods*, 35(11), 2005-2020.
- Alkhamisi, M. A., & Shukur, G. (2008). Developing ridge parameters for SUR model. *Communications in Statistics—Theory and Methods*, 37(4), 544-564.
- Bagheri, A., & Midi, H. (2009). Robust estimations as a remedy for multicollinearity caused by multiple high leverage points. *Journal of Mathematics and Statistics*, 5(4), 311-321.
- Gujarati, D. N. (2003). *Basic Econometrics*. 4th. New York: McGraw-Hill.
- Hald, A. (1952). *Statistical theory with engineering applications*. In *Statistical theory with engineering applications*. John Wiley & Sons.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1), 69-82.
- Hoerl, A. E., & Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Khalaf, G., & Shukur, G. (2005). Choosing ridge parameter for regression problems. *Communications in Statistics-Theory and Methods*, 34, 1177-1182.
- Kibria, B. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 32(2), 419-435.
- Mansson, K., Shukur, G., & Kibria, B. G. (2010). A simulation study of some ridge regression estimators under different distributional assumptions. *Communications in Statistics-Simulation and Computation*, 39(8), 1639-1670.
- Mansson, K., Shukur, G., & Kibria, B. G. (2010). On some ridge regression estimators: A Monte Carlo simulation study under different error variances. *Journal of Statistics*, 17(1).
- McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350), 407-416.
- Muniz, G., & Kibria, B. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics—Simulation and Computation*, 38(3), 621-630.
- Pati, K. D., Adnan, R., & Rasheed, B. A. (2015). Using ridge least median squares to estimate the parameter by solving multicollinearity and outliers problems. *Modern Applied Science*, 9(2), 191.

## MODIFIED HPM FOR LINEAR FRACTIONAL INTEGRO-DIFFERENTIAL EQUATIONS OF FREDHOLM-VOLTERRA TYPE

Hamson, K. M.,<sup>1</sup> Eshkuvatov, Z. K.<sup>2</sup> and Taib, B. M.<sup>3</sup>

Faculty of Science and Technology (FST), University Sains Islam Malaysia  
Nilai, Negeri Sembilan<sup>1,2,3</sup>

Institute for Mathematical Research, Universiti Putra Malaysia (UPM)<sup>2</sup>

khadijah\_hamson@yahoo.com<sup>1</sup>

### ABSTRACT

*In this study, we consider linear fractional integro-differential equations of order  $1 < \alpha \leq 2$  of Fredholm-Volterra type with mixed boundary conditions. Modified homotopy perturbation method (HPM) is applied to solve the fractional integro-differential equations of order  $\alpha$ . Suitable choices of unknown parameters together with boundary conditions lead the higher accurate approximate solution. Finally, two numerical examples are provided to show the validity and the efficiency of the method presented. Comparisons with other method are also made and results reveal that the proposed method is very efficient and accurate.*

**Keywords:** HPM, Fractional operator, Fredholm-Volterra integral equation, Discretization.

### 1. INTRODUCTION

The concept of differentiation and integration to non integer orders is not new. In fact, the idea of fractional calculus has been around since the 19<sup>th</sup> century. In the recent years, fractional integro-differential equations has taken the interest of researchers for their wide application in various fields. Fractional integro-differential equation appears in many modelling problems such as economics (Baillie, 1996), control theory (Bohannan, 2008) and so on.

Fractional integro-differential equations (Fractional IDEs) are mostly difficult to solve analytically, thus they are usually solved by approximation. There are numerous techniques that have been used to solve Fractional IDEs approximately. Some of them includes Adomian decomposition method (Momani & Qaralleh, 2006), variational iteration method (Sayevand, 2015), homotopy perturbation method (Elbeleze et al., 2016) and discrete Galerkin method (Mokhtary, 2016).

The modified technique was proposed by Ghorbani & Saberi-Nadjafi (2008) for nonlinear integral equations of the forms

$$y(x) = g(x) + \int_a^x k(x,t)[y(t)]^r dt \quad a \leq x, t \leq b,$$

$$y(x) = g(x) + \int_a^b k(x,t)[y(t)]^r dt \quad a \leq x \leq b,$$

They have shown through examples that the method is reliable and give better results than the standard homotopy perturbation method.

In this note, we consider Fredholm-Volterra integro-differential equation of fractional order  $1 < \alpha \leq 2$ ,

$$D_*^\alpha y(x) - g(x)y(x) = f(x) + \lambda_1 \int_a^b K_1(x,t)y(t)dt + \lambda_2 \int_a^x K_2(x,t)y(t)dt \quad (1)$$

under mixed boundary conditions

$$\begin{cases} a_1 y(a) + a_2 y'(a) = c_1 \\ b_1 y(b) + b_2 y'(b) = c_2 \end{cases} \quad (2)$$

where  $D_*^\alpha$  is the fractional derivative in the Caputo sense.

Equation (1)-(2) is solved by modified homotopy perturbation method (HPM) and compared with various methods, in Elbeleze et al. (2016) and Meng et al. (2015).

The paper is organized as follows: a brief review of the theory of fractional calculus is given in Section 2. Application of modified HPM is given in Section 3 to solve the fractional integro-differential equation (1)-(2). Section 4 presents examples to show the efficiency of the proposed method and finally followed by results and discussion.

## 2. PRELIMINARIES

**Definition 2.1** A real function  $f(x), x > 0$  is said to be in space  $C_\mu, \mu \in R$  if there exists a real number  $p > \mu$ , such that  $f(x) = x^p f_1(x)$  where  $f_1(x) \in C(0, \infty)$ , and it is said to be in space  $C_\mu^n$  if  $f^n \in R_\mu, n \in N$ .

**Definition 2.2** The Riemann-Liouville fractional integral operator of order  $\alpha \geq 0$  of a function  $f \in C_\mu, \mu \geq -1$ , is defined as

$$\begin{cases} J^\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_a^x (x-t)^{\alpha-1} f(t) dt, \alpha > 0, t > 0 \\ J^0 f(x) = f(x) \end{cases} \quad (3)$$

For  $f \in C_\mu, \mu \geq -1, \alpha, \beta \geq 0$  and  $\gamma \geq -1$ , some properties of the operator  $J^\alpha$ , which are needed here, are as follows:

$$(i) J^\alpha J^\beta f(x) = J^{\alpha+\beta} f(x) \quad (ii) J^\alpha J^\beta f(x) = J^\beta J^\alpha f(x) \quad (iii) J^\alpha x^\gamma = \frac{\Gamma(\gamma+1)x^{\alpha+\gamma}}{\Gamma(\alpha+\gamma+1)}$$

**Definition 2.3.** The fractional derivative of  $f(x)$  in the Caputo sense is defined as

$$D_*^\alpha f(x) = \frac{1}{\Gamma(m-\alpha)} \int_a^x (x-t)^{m-\alpha-1} f^{(m)}(t) dt, \quad (4)$$

for  $m-1 < \alpha \leq m, m \in N, t > 0, f \in C_{-1}^m$ .

**Lemma 2.1** (Kilbas et al., 2006) If the fractional derivative  $D^\alpha f(x), m-1 < \alpha \leq m$ , of a function  $f(x)$  is integrable, then the following properties hold:

- 1)  $D^\alpha J^\alpha f(x) = f(x)$ ,
- 2)  $J^\alpha [D^\alpha f(x)] = f(x) - \sum_{j=0}^{m-1} f^{(j)}(0) \left( \frac{(x-a)^j}{j!} \right)$

In Ahmad & Sivasundara (2008), a Cauchy problem involving a fractional integro-differential equation with nonlocal condition is given by

$$\begin{cases} {}^c D^\alpha x(t) = f(t, x(t)) + \int_0^t k(t, \eta, x(\eta)) d\eta, t \in [0, T], T > 0, 0 < \alpha \leq 1, \\ x(0) = x_0 - g(x) \end{cases} \quad (5)$$

where  ${}^c D^\alpha x(t)$  denotes the Caputo fractional derivative,  $f : [0, T] \times X \rightarrow X, k : [0, T] \times X \rightarrow X$  are jointly continuous,  $g : C \rightarrow X$  is continuous. Here,  $(X, \|\cdot\|)$  is a Banach space and  $C = C([0, T], X)$  denotes the Banach space of all continuous functions from  $[0, T] \rightarrow X$

endowed with a topology of uniform convergence with the norm denoted by  $\|\cdot\|_C$ . The following theorems discusses the existence of (5).

**Theorem 2.1** (Meng et al., 2015). Assume that

- A.  $\|f(t, x_1) - f(t, x_2)\| \leq L_1 \|x_1 - x_2\|, \forall t \in [0, T], x_1, x_2 \in X$ ;
- B.  $\|k(t, s, x_1) - k(t, s, x_2)\| \leq L_2 \|x_1 - x_2\|, \forall t, s \in [0, T], x_1, x_2 \in X$ ;
- C.  $\|g(x_1) - g(x_2)\| \leq b \|x_1 - x_2\|, \forall x_1, x_2 \in C$ .

Then the Cauchy problem (5) has a unique solution provided

$$b < \frac{1}{2}, \quad L_1 \leq \frac{\Gamma(\alpha + 1)}{4T^\alpha}, \quad L_2 \leq \frac{\Gamma(\alpha + 2)}{4\alpha T^{\alpha+1}}.$$

**Theorem 2.2** (Meng et al., 2015). Assume that

- A.  $\|f(t, x)\| \leq \mu(t), \mu \in L_1([0, T], R^+), \forall (t, x) \in [0, T] \times X$ ,
- B.  $\|k(t, s, x)\| \leq \sigma(t), \sigma \in L_1([0, T], R^+), \forall (t, s, x) \in [0, T] \times [0, T] \times X$ ,
- C.  $\|g(x) - g(y)\| \leq b \|x - y\|, \forall x, y \in C, b < 1$ .

Then the Cauchy problem (5) has at least one solution on  $[0, T]$ .

### 3. MODIFIED HOMOTOPY PERTURBATION METHOD

Let us consider a linear Fredholm-Volterra integro-differential equations of fractional order of the form

$$D_*^\alpha y(x) - g(x)y(x) = f(x) + \lambda_1 \int_a^b K_1(x, t)y(t)dt + \lambda_2 \int_a^x K_2(x, t)y(t)dt \quad (6)$$

with mixed boundary conditions

$$\begin{cases} a_1 y(a) + a_2 y'(a) = c_1 \\ b_1 y(b) + b_2 y'(b) = c_2 \end{cases} \quad (7)$$

where  $1 < \alpha \leq 2$ . Equation (6) is a generalization of linear fractional differential equations, Fredholm integral equations, and Volterra integral equations.  $D_*^\alpha y(x)$  represent the Caputo fractional derivative of order  $\alpha$ .

By imposing operator  $J^\alpha$  in (3) on both sides of (6) and making use of Lemma 2.1, we get,

$$y(x) - y(a) - y'(a)(x - a) = J^\alpha \left[ g(x)y(x) + f(x) + \lambda_1 \int_a^b K_1(x,t)y(t)dt + \lambda_2 \int_a^x K_2(x,t)y(t)dt \right] \quad (8)$$

To illustrate the basic concepts of modified HPM equation (8) can be re-written as,

$$Ly = f + Ny \quad (9)$$

where

$$\begin{aligned} Ly &= y, \quad f = J^\alpha [f(x)], \\ Ny &= y(a) + y'(a)(x - a) \\ &+ J^\alpha \left[ g(x)y(x) + \lambda_1 \int_a^b K_1(x,t)y(t)dt + \lambda_2 \int_a^x K_2(x,t)y(t)dt \right] \end{aligned} \quad (10)$$

Following Ghorbani and Saberi-Nadjafi (2008), homotopy function constructed as follows:

$$H(v, \beta, p) = (1 - p)[L(v) - L(y_0)] + p[L(v) - f - N(v)], \quad (11)$$

with

$$y_0 = \sum_{j=0}^N \beta_j g_j(x). \quad (12)$$

where  $g_j(x), j = 0, \dots, n$  are the selective functions and  $\beta_j, j = 0, \dots, n$  are unknown coefficients to be determined later.

In equation (11),  $p \in [0, 1]$  is an imbedding parameter, and hence, it is obvious that

$$\begin{aligned} H(v, \beta, 0) &= L(v) - L(y_0), \\ H(v, \beta, 1) &= L(v) - f - N(v). \end{aligned} \quad (13)$$

and as  $p$  increases from 0 to 1,  $H(v, \beta, p)$  varies continuously from  $L(v) - L(y_0)$  to  $L(v) - f - N(v)$ . In topology, such continuous variation is called deformation, and  $L(v) - L(y_0)$  and  $L(v) - f - N(v)$  are called homotopic.

Equating equation (11) to be zero, we have

$$L(v) = \sum_{j=0}^N \beta_j g_j(x) + p(f + N(v) - \sum_{j=0}^N \beta_j g_j(x)) \quad (14)$$

In view of basic assumption of homotopy perturbation method, solution of equation (8) can be expressed as a power series in  $p$ ,

$$v = \sum_{k=0}^{\infty} p^k v_k \quad (15)$$

when  $p \rightarrow 1$ , then equation (15) corresponding to equation (14) becomes the approximate solution of equation (8) i.e.,

$$v(x) = v_0(x) + v_1(x) + v_2(x) + \dots \quad (16)$$

The series (15) is convergent for most cases, and also the rate of convergence depends on  $N(v)$ .

The proof can be found in He (1999).

Substituting (15) into (14) yields:

$$L(\sum_{k=0}^{\infty} p^k v_k) = \sum_{j=0}^N \beta_j g_j(x) + p \left( f + N(\sum_{k=0}^{\infty} p^k v_k) - \sum_{j=0}^N \beta_j g_j(x) \right) \quad (17)$$

and equating the terms with having identical power of  $p$ , the following series are obtained:

$$\left\{ \begin{array}{l} p^0 : v_0 = \sum_{j=0}^N \beta_j g_j(x), \\ p^1 : v_1 = A + B(x-a) + J^\alpha [g(x)y(x) + f(x)] \\ \quad + J^\alpha \left[ \lambda_1 \int_a^b K_1(x,t)v_0(t)dt + \lambda_2 \int_a^x K_2(x,t)v_0(t)dt \right] - v_0, \\ p^2 : v_2 = J^\alpha \left[ g(x)v_1(x) + \lambda_1 \int_a^b K_1(x,t)v_1(t)dt + \lambda_2 \int_a^x K_2(x,t)v_1(t)dt \right], \\ \quad \vdots \\ p^k : v_k = J^\alpha \left[ g(x)v_{k-1}(x) + \lambda_1 \int_a^b K_1(x,t)v_{k-1}(t)dt + \lambda_2 \int_a^x K_2(x,t)v_{k-1}(t)dt \right] \\ \quad \text{for } k \geq 2. \end{array} \right. \quad (18)$$

where  $A = y(a)$  and  $B = y'(a)$ .

In MHPM,  $v_1$  is forced to be equal to zero to find the values of  $\beta_j, j = 0, 1, 2, \dots$ . If  $v_1 = 0$

then,  $v_2 = v_3 = \dots = 0$  and the solution will be obtained as

$$v(x) = v_0(x) \quad (19)$$

The unknowns  $A$  and  $B$  are found by imposing boundary conditions (7) on (16).

**Remark:** In fractional equations where non integer differentiation involve, forcing  $v_2 = 0$  does not give unique solution and hence some terms of  $v_2 = 0$  are not equal to zero. Fortunately these nonzero terms impact to the next solution is very little, therefore the approximate solution of the problem is searched as two steps i.e.

$$v(x) = v_0(x) + v_1(x) \quad (20)$$

#### 4. NUMERICAL EXAMPLES

In this section, we have applied MHPM to linear Fredholm-Volterra integro-differential equations of fractional order with known exact solution.

**Remark:** In the examples, the following notations are used

$$\begin{aligned}
 E_1(x) &= \max_{0 \leq x \leq 1} |(\text{exact solution}) - v(x)|_{\alpha=1.25} \\
 E_2(x) &= \max_{0 \leq x \leq 1} |(\text{exact solution}) - v(x)|_{\alpha=1.5} \\
 E_3(x) &= \max_{0 \leq x \leq 1} |(\text{exact solution}) - v(x)|_{\alpha=1.75} \\
 E(x) &= \max_{0 \leq x \leq 1} |(\text{exact solution}) - v(x)|_{\alpha=2}
 \end{aligned} \tag{21}$$

**Example 1:** Consider the following fractional linear integro-differential equation:

$$D_*^{1.7} u(x) = \int_0^x (x-y)u(y)dy + \int_0^1 (x+y)u(y)dy + g(x), \quad 1 \leq \alpha \leq 2, \tag{22}$$

with these conditions  $u'(0) = u(0) = 0$  and,

$$g(x) = \frac{\Gamma(3)}{\Gamma(1.3)} x^{0.3} + \frac{\Gamma(4)}{\Gamma(2.3)} x^{1.3} - \frac{x^4}{12} - \frac{x^5}{20} - \frac{7x}{12} - \frac{9}{20} \tag{23}$$

The exact solution of this problem is  $u(x) = x^2 + x^3$ , Equation (21) was solved using Legendre wavelets method in (Meng et al., 2015).

Applying operator  $J^\alpha$  to both sides of (21) and according to modified HPM, we construct the following homotopy:

$$H(v, \beta, p) = (1-p) \left[ L(v) - \sum_{j=0}^N \beta_j g_j(x) \right] + p \left[ L(v) - f - N(v) \right] = 0, \tag{24}$$

where  $f = J^\alpha(g(x)) = J^\alpha \left( \frac{\Gamma(3)}{\Gamma(1.3)} x^{0.3} + \frac{\Gamma(4)}{\Gamma(2.3)} x^{1.3} - \frac{x^4}{12} - \frac{x^5}{20} - \frac{7x}{12} - \frac{9}{20} \right)$  and

$$N(v) = J^\alpha \left[ \int_0^x (x-t)v(t)dt + \int_0^1 (x+t)v(t)dt \right].$$

Substituting (15) into (23) and equating the terms with same power of  $p$ , we get the series

$$\left\{ \begin{array}{l}
 p^0 : v_0 = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3, \\
 p^1 : v_1 = J^\alpha \left( \frac{\Gamma(3)}{\Gamma(1.3)} x^{0.3} + \frac{\Gamma(4)}{\Gamma(2.3)} x^{1.3} - \frac{x^4}{12} - \frac{x^5}{20} - \frac{7x}{12} - \frac{9}{20} \right) \\
 \quad + J^\alpha \left[ \int_0^x (x-t)(\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) dt \right] \\
 \quad + J^\alpha \left[ \int_0^1 (x+t)(\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) dt \right] - (\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3), \quad (25) \\
 p^2 : v_2 = -J^\alpha \left[ \int_0^x (x-t)v_1(t) dt + \int_0^1 (x+t)v_1(t) dt \right], \\
 \quad \vdots \\
 p^k : v_k = -J^\alpha \left[ \int_0^x (x-t)v_{k-1}(t) dt + \int_0^1 (x+t)v_{k-1}(t) dt \right] \text{ for } k \geq 3.
 \end{array} \right.$$

Applying  $J^\alpha$ , we will obtain,

$$\left\{ \begin{array}{l}
 p^0 : v_0 = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3, \\
 p^1 : v_1 = x^2 + x^3 - \frac{\Gamma(5)}{12\Gamma(6.7)} x^{5.7} - \frac{\Gamma(6)}{20\Gamma(7.7)} x^{6.7} - \frac{7\Gamma(2)}{12\Gamma(3.7)} x^{2.7} - \frac{9}{20\Gamma(2.7)} x^{1.7} \\
 \quad + 0.002166197641 x^{6.7} b_3 + 0.004837841398 x^{5.7} b_2 \\
 \quad + 0.06480288553 x^{3.7} b_0 + 0.2397706764 x^{2.7} b_0 + 0.1198853383 x^{2.7} b_1 \\
 \quad + 0.07992355882 x^{2.7} b_2 + 0.05994266912 x^{2.7} b_3 + 0.3236904133 x^{1.7} b_0 \\
 \quad + 0.2157936089 x^{1.7} b_1 + 0.1618452066 x^{1.7} b_2 + 0.1294761653 x^{1.7} b_3 \\
 \quad + 0.01378784799 x^{5.7} b_1 - (\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3), \quad (26)
 \end{array} \right.$$

We had chosen  $N=3$  and  $g_j(x) = x^j, j=0,1,2,3$ . The unknowns  $\beta_j, j=0,1,2,3$  is found out by equating  $v_1 = 0$ .

$$\begin{aligned}
 v_1 = & x^2 + x^3 - 0.004837841400 x^{5.7} - 0.002166197642 x^{6.7} - 0.1398662280 x^{2.7} \\
 & - 0.2913213720 x^{1.7} + 0.002166197641 x^{6.7} b_3 + 0.004837841398 x^{5.7} b_2 \\
 & + 0.01378784799 x^{5.7} b_1 + 0.06480288553 x^{3.7} b_0 + 0.2397706764 x^{2.7} b_0 \\
 & + 0.1198853383 x^{2.7} b_1 + 0.07992355882 x^{2.7} b_2 + 0.05994266912 x^{2.7} b_3 \\
 & + 0.3236904133 x^{1.7} b_0 + 0.2157936089 x^{1.7} b_1 + 0.1618452066 x^{1.7} b_2 \\
 & + 0.1294761653 x^{1.7} b_3 - (\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3) = 0 \quad (27)
 \end{aligned}$$

From (26), we obtain system of algebraic equation:

$$\left\{ \begin{array}{l} -\beta_0 = 0, \\ -\beta_1 = 0, \\ -\beta_2 + 1 = 0, \\ -\beta_3 + 1 = 0, \\ -0.004837841400 + 0.004837841398b_2 = 0, \\ -0.002166197642 + 0.002166197641b_3 = 0, \\ 0.01378784799b_1 = 0, \\ 0.06480288553b_0 = 0, \\ -0.1398662280 + 0.2397706764b_0 + 0.1198853383b_1 + 0.07992355882b_2 \\ + 0.05994266912b_3 = 0 \\ -0.2913213720x^{1.7} + 0.3236904133x^{1.7}b_0 + 0.2157936089x^{1.7}b_1 \\ + 0.1618452066x^{1.7}b_2 + 0.1294761653x^{1.7}b_3 = 0 \end{array} \right. \quad (28)$$

We have chosen that  $\beta_0 = 0, \beta_1 = 0, \beta_2 = 1, \beta_3 = 1$ . Since it was found that  $v_1$  is not equal but very close to 0, thus, the solution becomes

$$\begin{aligned} v &= v_0 + v_1 \\ &= \sum_{j=0}^3 \beta_j g_j(x) + v_1 \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + v_1 \\ &= x^2 + x^3 + x^2 + x^3 - 0.004837841400x^{5.7} - 0.002166197642x^{6.7} - 0.1398662280x^{2.7} \\ &\quad - 0.2913213720x^{1.7} + 0.002166197641x^{6.7}b_3 + 0.004837841398x^{5.7}b_2 \\ &\quad + 0.01378784799x^{5.7}b_1 + 0.06480288553x^{3.7}b_0 + 0.2397706764x^{2.7}b_0 \\ &\quad + 0.1198853383x^{2.7}b_1 + 0.07992355882x^{2.7}b_2 + 0.05994266912x^{2.7}b_3 \\ &\quad + 0.3236904133x^{1.7}b_0 + 0.2157936089x^{1.7}b_1 + 0.1618452066x^{1.7}b_2 \\ &\quad + 0.1294761653x^{1.7}b_3 - (\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3) \end{aligned}$$

**Table 1:** Comparison with the exact solution and Lagendre wavelet method (LWM)

x	Exact	LWM	MHPM
0	0	0	0
0.1	0.017578125	0.017566	0.0175781249
0.2	0.078125000	0.078115	0.0781249000
0.3	0.193359375	0.193351	0.1933593749

0.4	0.375000000	0.374988	0.374900000
0.5	0.634765625	0.634693	0.6347656249
0.6	0.984375000	0.984162	0.9843749000
0.7	1.435546875	1.434649	1.4355468749

**Example 2:** Consider the following linear Fredholm integro-differential equation:

$$D_*^\alpha u(x) = 2e^x - 1 - \int_0^x u(t)dt, \quad 0 \leq x \leq 1, \quad 1 \leq \alpha \leq 2, \quad (29)$$

with initial boundary conditions

$$u(0) = 1, u'(1) = e \quad (30)$$

The exact is  $u(x) = e^x$ , when  $\alpha = 2$ . We do comparisons approximate solution which was obtained by modified HPM with HPM and VIM in (Elbeleze et. al., 2016).

Applying operator  $J^\alpha$  to both sides of (28) and according to modified HPM, we construct the following homotopy:

$$H(v, \beta, p) = (1-p)[L(v) - \sum_{j=0}^N \beta_j g_j(x)] + p[L(v) - f - N(v)] = 0, \quad (31)$$

where  $f = 1 + Ax$  and  $N(v) = J^\alpha \left[ 2e^x - 1 - \int_0^x v(t)dt \right]$ .

Substituting (15) into (30) and equating the terms with same power of  $p$ , we get

$$\left\{ \begin{array}{l} p^0 : v_0 = \beta_0 + \beta_1 x + \beta_2 x^2, \\ p^1 : v_1 = 1 - \beta_0 + (A - \beta_1)x - \beta_2 x^2 + \frac{x^\alpha}{\Gamma(\alpha+1)} + \frac{(2-\beta_0)}{\Gamma(\alpha+2)} x^{1+\alpha} + \frac{(2-\beta_1)}{\Gamma(\alpha+3)} x^{2+\alpha} \\ \quad + \frac{(2-2\beta_2)}{\Gamma(\alpha+4)} x^{3+\alpha} - (\beta_0 + \beta_1 x + \beta_2 x^2), \\ p^2 : v_2 = \frac{(\beta_0-1)}{\Gamma(\alpha+2)} x^{1+\alpha} + \frac{(\beta_1-A)}{\Gamma(\alpha+3)} x^{2+\alpha} + \frac{2\beta_2}{\Gamma(\alpha+4)} x^{3+\alpha} - \frac{x^{2\alpha+1}}{\Gamma(2\alpha+2)} \\ \quad + \frac{(\beta_0-2)}{\Gamma(2\alpha+3)} x^{2+2\alpha} + \frac{(\beta_1-2)}{\Gamma(2\alpha+4)} x^{3+2\alpha} + \frac{(2\beta_2-2)}{\Gamma(2\alpha+5)} x^{4+2\alpha}, \\ \quad \vdots \\ p^k : v_k = -J^\alpha \left[ \int_0^x v_{k-1}(t)dt \right] \text{ for } k \geq 3. \end{array} \right. \quad (32)$$

We have chosen  $N=2$ . The unknowns  $\beta_j, j = 0,1,2$ , is found out by equating  $v_2 = 0$ .

$$v_2 = \frac{(\beta_0 - 1)}{\Gamma(\alpha + 2)} x^{1+\alpha} + \frac{(\beta_1 - A)}{\Gamma(\alpha + 3)} x^{2+\alpha} + \frac{2\beta_2}{\Gamma(\alpha + 4)} x^{3+\alpha} - \frac{x^{2\alpha+1}}{\Gamma(2\alpha + 2)} + \frac{(\beta_0 - 2)}{\Gamma(2\alpha + 3)} x^{2+2\alpha} + \frac{(\beta_1 - 2)}{\Gamma(2\alpha + 4)} x^{3+2\alpha} + \frac{(2\beta_2 - 2)}{\Gamma(2\alpha + 5)} x^{4+2\alpha} \quad (33)$$

From (32), we obtain system algebraic below:

$$\begin{cases} \beta_0 - 1 = 0, \\ \beta_1 - A = 0, \\ 2\beta_2 = 0 \end{cases} \quad (34)$$

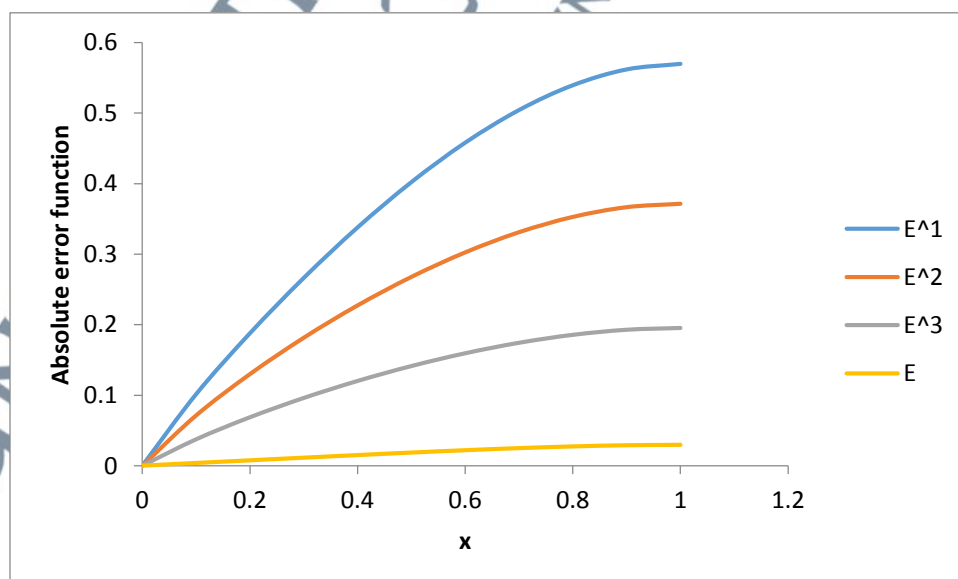
We had chosen so that  $\beta_0 = 1, \beta_1 = A, \beta_2 = 0$ . Thus, the solution becomes

$$v = v_0 + v_1 = 1 + Ax + \frac{x^\alpha}{\Gamma(\alpha + 1)} + \frac{x^{1+\alpha}}{\Gamma(\alpha + 2)} + \frac{(2 - A)}{\Gamma(\alpha + 3)} x^{2+\alpha} + \frac{2x^{\alpha+3}}{\Gamma(\alpha + 4)} + \dots \quad (35)$$

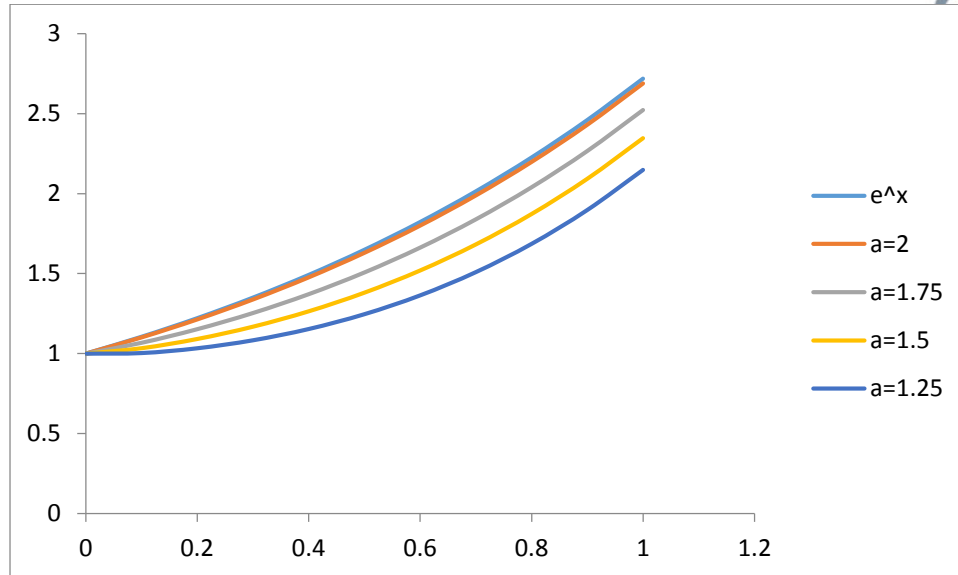
Unknown  $A$  is found out by imposing boundary conditions (29). Different values of  $\alpha$  will give different values of  $A$  as shown in Table 2.

**Table 2:** Value of  $A$  for different values of  $\alpha$  using (29)

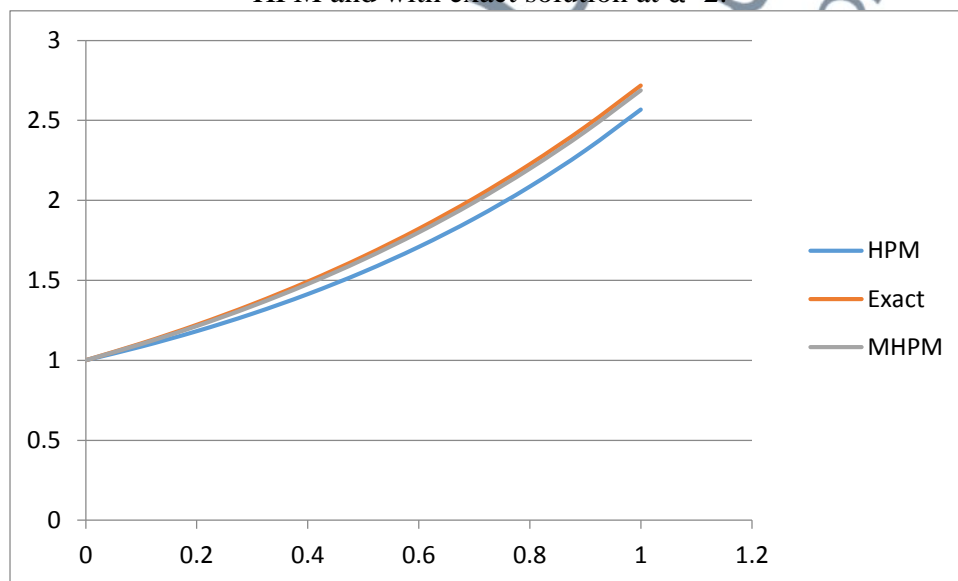
	$\alpha=1.25$	$\alpha=1.5$	$\alpha=1.75$	$\alpha=2$
$A$	-0.482996524	0.091409342	0.562986332	0.961938195



**Figure 1:** Absolute error of functions  $E_1(x), E_2(x), E_3(x)$  and  $E(x)$  obtained by 2-term modified HPM.



**Figure 2:** Comparison of approximate solutions with various  $\alpha$  obtained by 2-term modified HPM and with exact solution at  $\alpha=2$ .



**Figure 3:** Comparison of approximate solutions obtained by 2-term modified HPM and HPM with exact solution at  $\alpha=2$ .

**Table 3:** Comparison with the exact solution and HPM

x	Exact	HPM	MHPM
0	1.0000	1.0000	1.0000
0.1	1.1052	1.0853	1.1014
0.2	1.2214	1.1818	1.2138
0.3	1.3499	1.2907	1.3385
0.4	1.4918	1.4136	1.4767
0.5	1.6487	1.5524	1.6300
0.6	1.8221	1.7091	1.8001
0.7	2.0138	1.8861	1.9887
0.8	2.2255	2.0862	2.1981
0.9	2.4596	2.3125	2.4305
1.0	2.7183	2.5683	2.6885

**Remark** We could have chosen to iterate using the actual kernel  $e^x$  except that the integrals will result in a long chain, which gets longer from one iteration to the next. Instead, we chose to approximate this kernel by a limited number of terms. (Explanation on choosing limited term for exponential function).

## 5. RESULTS AND DISCUSSION

In this study, we introduced the modified homotopy perturbation method for fractional integro-differential equation with boundary conditions (1)-(2). For Example 1, we have got exact solution at  $\alpha = 1.7$  and comparison is done with Legendre Wavelet method as shown in Table 1. For Example 2, Table 2 and Figure 1 show that the absolute error decreases as the value of  $\alpha$  increases from 1.25 to 1.75. Figure 2 shows the comparison of approximate solutions for different value of  $\alpha = [1.25, 1.5, 1.75, 2]$  with the exact solution. Example 2 was also

compared with homotopy perturbation method at  $\alpha = 2$  as shown in Figure 3 and Table 3. These examples show that the proposed method works well and dominate HPM. It is noted that the approximate values can be more accurate with more number of terms chosen.

## ACKNOWLEDGEMENT

This work was supported by Universiti Sains Islam Malaysia (USIM) under Research Grant of USIM PPP/GP/FST/30/14915). Authors are grateful for sponsor and financial support of the Research Management Center (RMC) of USIM.

## REFERENCES

- Ahmad, B., & Sivasundaram, S. (2008). Some existence results for fractional integro-differential equations with nonlinear conditions. *Communications in Applied Analysis*, 12(2), 107.
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1), 5-59.
- Bohannan, G. W. (2008). Analog fractional order controller in temperature and motor control applications. *Journal of Vibration and Control*, 14(9-10), 1487-1498.
- Elbeleze, A. A., Kılıçman, A., & Taib, B. M. (2016). Approximate solution of integro-differential equation of fractional (arbitrary) order. *Journal of King Saud University-Science*, 28(1), 61-68.
- Ghorbani, A., & Saberi-Nadjafi, J. (2008). Exact solutions for nonlinear integral equations by a modified homotopy perturbation method. *Computers & Mathematics with Applications*, 56(4), 1032-1039.
- He, J. H. (1999). Homotopy perturbation technique. *Computer methods in applied mechanics and engineering*, 178(3), 257-262.
- Kilbas, A. A. A., Srivastava, H. M., & Trujillo, J. J. (2006). *Theory and applications of fractional differential equations* (Vol. 204). Elsevier Science Limited.
- Meng, Z., Wang, L., Li, H., & Zhang, W. (2015). Legendre wavelets method for solving fractional integro-differential equations. *International Journal of Computer Mathematics*, 92(6), 1275-1291.
- Mokhtary, P. (2016). Discrete Galerkin method for fractional integro-differential equations. *Acta Mathematica Scientia*, 36(2), 560-578.

Momani, S., & Qaralleh, R. (2006). An efficient method for solving systems of fractional integro-differential equations. *Computers & Mathematics with Applications*, 52(3), 459-470.

Sayevand, K. (2015). Analytical treatment of Volterra integro-differential equations of fractional order. *Applied Mathematical Modelling*, 39(15), 4330-4336.

UNIVERSITI SAINS ISLAM MALAYSIA  
جامعة العلوم الإسلامية  
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

## BACHOK--HASHAM POLYNOMIALS APPROACH FOR EIGENVALUES PROBLEM OF LOGARITHMIC SINGULAR INTEGRAL EQUATIONS

Alhawamda, H.,<sup>1</sup> Taib, B. M.,<sup>2</sup> Eshkuvatov, Z. K.<sup>3</sup> and Ibrahim, R. I.<sup>4</sup>

Faculty of Science & Technology (FST), University Sains Islam Malaysia  
Nilai, Negeri Sembilan<sup>1, 2, 3, 4</sup>

hisham\_math77@yahoo.com

### ABSTRACT

*This study aims to propose semi-analytical approximate solutions of linear homogeneous Logarithmic Singular Integral Equations of second kind (LogSIEs). Eigenvalues and corresponding Eigen functions problems are considered. It is found that certain class of LogSIE can be solved analytically. The proposed method is based on the Bachok-Hasham's polynomials ( $H_n^k(x)$ ) via Galerkin Method and implementation of this method is more efficient and more accurate for solving such class of integral equations. On the other hand, the benefit of this method is low cost of computing operations. The applied method transforms the Linear homogeneous LogSIEs of second kind into linear algebraic system that can be solved easily.*

**Keywords:** *Logarithmic Singular Integral Equations; Bachok-Hasham polynomials; Galerkin method; Chebyshev polynomials.*

### 1. INTRODUCTION

Recently, most of the problems are mathematically described by integral equations. Modern life science is the development of mathematical models that appropriately indicate the problem, which is given a complication of the real physical situation. The analytical solution of these integral equations is facing great difficulties and limitations. Mathematical models are utilized to represent the problem properly and to use numerical methods in order to achieve approximate solutions. Lately, the development of computer software has made the use of numerical methods for solving integral equations important in scientific investigations. In addition to the numerical methods that has been spreading commonly, which are fundamental to our understanding of scientific phenomena, in which there were too much difficulty in the past (Corduneanu, 1991; Polyanin & Manzhirov, 2008; Hochstadt, 2011).

Orthogonal functions methods are developed for the numerical solution of singular integral equations with Cauchy kernel by considering an appropriate weight function. Additionally, orthogonal functions methods have been proposed to solve linear integral equations (Hsiao & Chen, 1979). Special attentions have been given to applications of block-pulse functions (Nosrati et al., 2015), Legendre polynomials (Jafari et al., 2011), Laguerre Gaussians (Aboiyar et al., 2010), and Chebyshev polynomials (Paryab & Rostami, 2008). On the other hand, the advancement and innovation of numerical technology in modeling engineering problems have been experiencing constant technological developments, as well as enlarging its scope for its efficiency and versatility.

Many studies have conducted the intention of dealing effectively with the Logarithmic Singular Integral Equations. These surveys began in the mid-seventies and after many years, with the advancement of technology, the effective computational approach for the evaluation of singular integrals undergoes changes and modernizations (Sizikov & Sidorov, 2016; Berthold et al., 1992). The classic formulation of the Logarithmic Singular Integral Equations (LogSIEs) has been used in several areas, such as mechanics, elastodinamicos problems, plates, soil mechanics and potential problems. Singular integral equations are connected with such boundary values problems in a natural way. In this case it comes to hyper singular formulation, as this is where the boundary integral equations have improper cores, (Eshkuvatov et al., 2009; Nadir & Antidze, 2004).

However, the theory of singular integral equations with their integrals in the sense of its principal value was arose out after the development of the classical theory of integral equations by Fredholm in 1903. Logarithmic Singular Integral Equations and there properties were investigated by many authors (Okecha, 2007), furthermore, the theory of singular integral equations did not encounter the attention of the mathematicians for a certain time.

LogSIEs as a mathematical tools in several science fields such as applied mathematical, medicine, mechanical, economics, and applications in theory of elasticity, plasticity, aerodynamics, filtration theory, hydromechanics engineering, fluid dynamics, and electrical engineering which have been extensively studied by several authors such as Mennouni (2013)

and Chakrabarti & Martha (2012). More information can be found, in one-dimensional equations, on the classics books (Sen & Powers, 2001).

The aim of this work is to present semi-analytical and numerical solutions for special case of Linear homogeneous LogSIEs of second kind as defined

$$f(t) = \lambda \int_{-1}^1 k(x,t) f(x) dx, \quad x, t \in [-1,1] \quad (1)$$

Into two kernel cases:

- $k(x,t) = \log|x^k - t^k|$ ,  $k$  is positive odd integer
- $k(x,t) = \frac{x^{k-1}}{\sqrt{1-x^{2k}}} \log|x^k - t^k|$ ,  $k$  is positive odd integer.

The propose method is reducing the given Linear homogeneous LogSIEs to a set of algebraic equations by expanding the unknown function  $f(t)$  using Bachok–Hasham polynomials ( $H_n^k(x)$ ) with unknown coefficients. Galerkin method is utilized to evaluate the unknown coefficients because the orthogonality of Bachok–Hasham polynomials  $H_n^k(x)$  with weight

$$w_k(x) = \frac{x^{k-1}}{\sqrt{1-x^{2k}}}, \text{ where } k \text{ is positive odd integer.}$$

## 2. MATHEMATICAL PRELIMINARIES

In this section introduces briefly in order to give some basic definitions of integral equations theory. For more details see any introductory text on real analysis (Zarowski, 2004).

### 2.1 Linear Space and Function Spaces

A linear space also is called a vector space which is a set  $X$  over with two operations defined on its elements, addition and scalar multiplication under  $X$ , there is a need for a generalization idea for the size of the individual elements between two vectors of an arbitrary vector space by

supplying the vector space  $X$  with a real valued function, which is called a norm, moreover, a vector space provides with a norm is called a normed linear space,  $(X, \|\cdot\|)$

Another important concept is an inner product to the space. An inner product on a real vector space  $X$  is a real-valued function defined on  $X \times X$  and denoted with the following properties:

**Definition 1:** An inner product  $\langle \cdot, \cdot \rangle$  is a linear function of elements  $f, g, h, \dots$  of a vector space that satisfies the axioms:

- 1-  $\langle f, f \rangle = 0 \Leftrightarrow f = 0$
- 2-  $\langle f, g \rangle = \langle g, f \rangle$
- 3-  $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$
- 4-  $\langle \alpha f, g \rangle = \alpha \langle f, g \rangle$  for any scalar

Using the notation  $L_2[-1, 1]$  to denote the space of square integrable functions of a single variable on the interval  $[-1, 1]$ , the inner product corresponding to its norm is

$$\langle f, g \rangle = \left[ \int_{-1}^1 f(x)g(x)dx \right]^{1/2} \quad (3)$$

Moreover, we say that two functions  $f$  and  $g$  are orthogonal if

$$\langle f, g \rangle = 0 \quad (4)$$

## 2.2 Orthogonal Polynomials and Weight Functions

Orthogonal Polynomials play an important role in numerical analysis, because functions are belong to general classes which can be expanded in sense of orthogonal functions. Additionally, the orthogonality property of these polynomials have many other general properties; a large literature in which the contributions of researchers and mathematicians like Bernstein, Chebyshev, etc see (Boyd, 2001).

**Definition 2:** Two functions  $f(x)$  and  $g(x)$  in  $L_2[a, b]$  are said to be orthogonal on the interval  $[a, b]$  with respect to a given continuous and non-negative weight function  $w(x)$  if

$$\int_a^b w(x)f(x)g(x)dx = 0 \quad (5)$$

for convenience, we use the inner product notation:

$$\langle f, g \rangle_w = \int_a^b w(x)f(x)g(x)dx = 0$$

where  $f$  and  $g$  are functions of  $x$  on  $[a, b]$ , then the orthogonality condition Eq. (5) is equivalent to

$$\langle f, g \rangle_w = 0 \quad (6)$$

### 3. BACHOK-HASHAM POLYNOMIALS OF ORDER K

Orthogonal polynomials are widely used in applications to various fields in mathematics, mathematical physics, engineering and computer science. One of the most common set of this kind of polynomials is the set of first kind of Chebyshev polynomials,  $T_0(x), T_1(x), T_2(x), \dots$  which are orthogonal with respect to the weight function  $w(x) = \frac{1}{\sqrt{1-x^2}}$   $x \in [-1, 1]$ , (Mason & Handscomb, 2002).

Similarly, the Bachok–Hasham polynomials  $H_n^k(x)$  are an extension of the Chebyshev polynomials. Beside, their polynomials are orthogonal with weight  $w_k(x) = \frac{x^{k-1}}{\sqrt{1-x^{2k}}}$ , where  $k$  is positive odd integer.

**Definition 3: Bachok-Hasham Polynomials of order k:** First kind of the Bachok–Hasham polynomials  $H_n^k(x)$  of order  $k$ , where  $k$  is positive odd integer number, is a polynomial in  $x$  of degree  $(kn)$ , defined by the relation (for fixed  $k$ ):

$$H_n^k(x) = \cos(n\theta), \quad x^k = \cos(\theta), \quad \forall k = 1, 3, \dots, N; n = 0, 1, 2, \dots \quad (7)$$

The range of the variable  $x$  is the interval  $[-1, 1]$  and the range of the corresponding variable  $\theta$  is  $[0, \pi]$ . These ranges are traversed in opposite directions i.e.  $x = -1$ , corresponds to  $\theta = \pi$  and  $x = 1$ , corresponds to  $\theta = 0$ .

On the other hand the above definition is well defined since  $x \in [-1, 1]$  and  $|\cos(\theta)| \leq 1$  implies  $|x^k| \leq 1$ .

Here, we illustrate the first few Bachok-Hasham polynomials of order  $k$ ,

$$\begin{aligned}
 H_0^k(x) &= 1 \\
 H_1^k(x) &= x^k \\
 H_2^k(x) &= 2x^{2k} - 1 \\
 H_3^k(x) &= 4x^{3k} - 3x^k \\
 H_4^k(x) &= 8x^{4k} - 8x^{2k} + 1 \\
 H_5^k(x) &= x^k(16x^{4k} - 20x^{2k} + 5) \\
 H_6^k(x) &= 32x^{6k} - 48x^{4k} + 18x^{2k} - 1
 \end{aligned} \tag{8}$$

It's easy to show that when  $k = 1$ ,

$$H_n^1(x) = T_n(x), \quad n = 0, 1, 2, \dots; \tag{9}$$

where  $T_n(x)$  is the Chebyshev polynomials, indeed a first kind of Bachok-Hasham polynomial of order  $k$ ,  $H_n^k(x)$  is a function of Chebyshev polynomials in  $x^k$ .

### 3.1 Bachok-Hasham Polynomials of First Kind for Order $K$ and Weight Functions

#### Theorem(1):

The first kind Bachok-Hasham polynomials  $H_n^k(x)$  satisfy:

$$\langle H_n^k, H_m^k \rangle = \int_{-1}^1 \frac{x^{k-1} H_n^k(x) H_m^k(x)}{\sqrt{1-x^{2k}}} dx = \begin{cases} 0, & m \neq n \\ \frac{\pi}{2k}, & m = n \\ \frac{\pi}{k}, & m = n = 0 \end{cases}; \tag{10}$$

$n, m = 0, 1, 2, 3, \dots$ ;  $k$  is positive odd integer.

#### Proof:

Consider  $H_n^k(x) = \cos(n \cos^{-1}(x^k))$ , and  $H_m^k(x) = \cos(m \cos^{-1}(x^k))$ ,  $n, m, 0, 1, 2, 3, \dots$  we have,

$$\langle H_n^k, H_m^k \rangle = \int_{-1}^1 \frac{x^{k-1} H_n^k(x) H_m^k(x)}{\sqrt{1-x^{2k}}} dx = \int_{-1}^1 \frac{x^{k-1} \cos(n \cos^{-1}(x^k)) \cos(m \cos^{-1}(x^k))}{\sqrt{1-x^{2k}}} dx \tag{11}$$

Let  $u = x^k \Rightarrow du = kx^{k-1}dx$

$$\frac{1}{k} \int_{-1}^1 \frac{\cos(n \cos^{-1}(u)) \cos(m \cos^{-1}(u))}{\sqrt{1-u^2}} du \quad (12)$$

Setting  $u = \cos(\theta) \Rightarrow du = -\sin(\theta)d\theta$ , substitute in Eq. (11):

$$\begin{aligned} &= \frac{1}{k} \int_0^\pi \frac{\cos(n \cos^{-1}(\cos(\theta))) \cos(m \cos^{-1}(\cos(\theta)))}{\sqrt{1-\sin^2(\theta)}} d\theta \\ &= \frac{1}{k} \int_0^\pi \cos(n\theta) \cos(m\theta) d\theta \end{aligned}$$

Using relation  $\cos(n\theta) \cos(m\theta) = \frac{1}{2} [\cos(n+m)\theta + \cos(n-m)\theta]$

Now, for  $n \neq m$

$$\begin{aligned} \frac{1}{k} \int_0^\pi \cos(n\theta) \cos(m\theta) d\theta &= \frac{1}{2k} \int_0^\pi [\cos(n+m)\theta + \cos(n-m)\theta] d\theta \\ &= \frac{1}{2k} \left[ \frac{\sin(n+m)\theta}{n+m} + \frac{\sin(n-m)\theta}{n-m} \right]_0^\pi \\ &= 0 \end{aligned}$$

The norm of  $H_n^k(x)$  is given by

$$\begin{aligned} \|H_n^k\|^2 &= \langle H_n^k, H_n^k \rangle = \frac{1}{k} \int_0^\pi \frac{\cos(n\theta) \cos(n\theta)}{\sqrt{1-\sin^2(\theta)}} d\theta \\ &= \frac{1}{k} \int_0^\pi \cos(n\theta) \cos(n\theta) d\theta \\ &= \frac{1}{2k} \int_0^\pi (1 + \cos(2n\theta)) d\theta \quad (13) \\ &= \frac{1}{2k} \left[ \theta + \frac{\sin(2n\theta)}{2n} \right]_0^\pi \\ &= \frac{\pi}{2k} \end{aligned}$$

#### 4. ANALYTIC AND APPROXIMATION SOLUTIONS

This section investigates Linear homogeneous LogSIEs of second kind, our technique in dealing with logarithmic kernels is essentially to reduce the integral equations problems to one

involving matrices. The key to this is the observation to write the unknown function as finite series of the Bachok–Hasham polynomial  $H_n^k(x)$ .

The integral equation:

$$\lambda \phi(t) = \int_{-1}^1 \frac{x^{k-1}}{\sqrt{1-x^{2k}}} \phi(x) \log|x^k - t^k| dx \quad (14)$$

has an Eigen solutions  $\phi(t) = H_n^k(t)$ , and eigenvalues  $\lambda = \lambda_n = -\frac{\pi}{nk}$  for  $n = 1, 2, 3, \dots$

,where  $H_n^k(x)$  is first kind of Bachok-Hasham polynomials,  $k$  is positive odd integer.

- The integral equation:

$$\lambda f(t) = \int_{-1}^1 f(x) \log|x^k - t^k| dx, \quad k \text{ is positive odd integer } t, x \in [-1, 1] \quad (15)$$

has a trivial solution,  $f(t) = 0$ .

Now, we construct an approximation solution for nonhomogeneous case of LogSIEs:

$$g(t) = \int_{-1}^1 f(x) \log|x^k - t^k| dx, \quad t, x \in [-1, 1] \quad k \text{ is positive odd integer} \quad (16)$$

we assume that the unknown function  $f(t)$  can be written as:

$$f(t) = \frac{t^{k-1}}{\sqrt{1-t^{2k}}} \phi(t) \quad (17)$$

where  $f(t)$  is a well-defined function of  $t$ , approximated by a finite sum of Bachok-Hasham polynomials of first kind  $H_n^k(t)$  as

$$\phi(t) \cong a_0 H_0^k(t) + \sum_{i=1}^n a_i H_i^k(t) \quad (18)$$

where  $a_i, i = 0, 1, 2, \dots, n$  are the unknown coefficients.

Now, Substitute Eq.(17) and Eq.(18) into Eq.(16), we have

$$g(t) = \int_{-1}^1 a_0 H_0^k(t) \frac{t^{k-1}}{\sqrt{1-t^{2k}}} + \sum_{i=1}^n a_i \frac{t^{k-1}}{\sqrt{1-t^{2k}}} H_i^k(t) \log|x^k - t^k| dx [0, 1] \quad (19)$$

The first kind of Bachok–Hasham polynomials  $H_n^k(x)$  are orthogonal polynomials with weight  $w(t) = \frac{t^{k-1}}{\sqrt{1-t^{2k}}}$ , moreover, apply theorem (1), and multiply both side by

$H_j^k(x), j = 0, 1, 2, \dots, n$ , Eq.(19) can be formulated a system of linear equations :

for  $i = 0$

$$a_0 \frac{\pi}{k} \log 2 \beta_0^k + \pi a_1 \frac{\alpha_{0,1}}{1 \times k} + \pi a_2 \frac{\alpha_{0,2}}{2 \times k} + \pi a_3 \frac{\alpha_{0,3}}{3 \times k} + \dots + \pi a_n \frac{\alpha_{0,n}}{n \times k} = C_0^k$$

for  $i = 1$

$$a_0 \frac{\pi}{k} \log 2 \beta_1^k + \pi a_1 \frac{\alpha_{1,1}}{1 \times k} + \pi a_2 \frac{\alpha_{1,2}}{2 \times k} + \pi a_3 \frac{\alpha_{1,3}}{3 \times k} + \dots + \pi a_n \frac{\alpha_{1,n}}{n \times k} = C_1^k$$

for  $i = 2$

$$a_0 \frac{\pi}{k} \log 2 \beta_2^k + \pi a_1 \frac{\alpha_{2,1}}{1 \times k} + \pi a_2 \frac{\alpha_{2,2}}{2 \times k} + \pi a_3 \frac{\alpha_{2,3}}{3 \times k} + \dots + \pi a_n \frac{\alpha_{2,n}}{n \times k} = C_2^k \quad (20)$$

$$\begin{array}{ccccccc} \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \end{array}$$

for  $i = n$

$$a_0 \frac{\pi}{k} \log 2 \beta_n^k + \pi a_1 \frac{\alpha_{n,1}}{1 \times k} + \pi a_2 \frac{\alpha_{n,2}}{2 \times k} + \pi a_3 \frac{\alpha_{n,3}}{3 \times k} + \dots + \pi a_n \frac{\alpha_{n,n}}{n \times k} = C_n^k$$

Where

$$\begin{aligned} \beta_i^k &= \int_{-1}^1 H_i^k(t) dt \\ \alpha_{i,j}^k &= \int_{-1}^1 H_i^k(t) H_j^k(t) dt, \quad i, j = 0, 1, 2, \dots, n \\ C_J^K &= \int_{-1}^1 g(t) H_j^k(t) dt \end{aligned} \quad (21)$$

Eq . (20) forge a system of linear algebraic equations which can be represented as a matrix form:

$$\begin{bmatrix} \frac{\pi}{k} \log 2 \beta_0^k & \pi \frac{\alpha_{0,1}}{1 \times k} & \pi \frac{\alpha_{0,2}}{2 \times k} & \pi \frac{\alpha_{0,3}}{3 \times k} & \dots & \pi \frac{\alpha_{0,n}}{n \times k} \\ \frac{\pi}{k} \log 2 \beta_1^k & \pi \frac{\alpha_{1,1}}{1 \times k} & \pi \frac{\alpha_{1,2}}{2 \times k} & \pi \frac{\alpha_{1,3}}{3 \times k} & \dots & \pi \frac{\alpha_{1,n}}{n \times k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\pi}{k} \log 2 \beta_3^k & \pi \frac{\alpha_{i,1}}{i \times k} & \pi \frac{\alpha_{i,2}}{i \times k} & \pi \frac{\alpha_{i,3}}{i \times k} & \dots & \pi \frac{\alpha_{i,n}}{i \times k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\pi}{k} \log 2 \beta_n^k & \pi \frac{\alpha_{n,1}}{n \times k} & \pi \frac{\alpha_{n,2}}{n \times k} & \pi \frac{\alpha_{n,3}}{n \times k} & \dots & \pi \frac{\alpha_{n,n}}{n \times k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} C_0^k \\ C_1^k \\ \vdots \\ C_i^k \\ \vdots \\ C_n^k \end{bmatrix} \quad (22)$$

Where  $\beta_i^k, \alpha_{i,j}^k, C_j^k$  are described in Eq. (21).

The above matrix has size  $(n + 1) \times (n + 1)$ , we use a software MATLAB program, to determine Coefficients  $a_i, i = 0, 1, 2, 3, \dots, n$ .

## 5. NUMERICAL EXAMPLES

In this section, we consider two examples to study accuracy and performance of the proposed method. It is performed with  $n = 4$ , we restrict to the simplest cases

$k = 3, 5$ . The right-hand side for the Galerkin method is always computed as explained in section 3.

**Example 1** Consider a Nonhomogeneous linear LogSIEs of first kind:

$$t^5 = -\frac{5}{\pi} \int_{-1}^1 f(x) \log |x^5 - t^5| dx, \quad x, t \in [-1, 1] \quad (23)$$

The exact solution of Eq. (23) is  $f(t) = \frac{t^9}{\sqrt{1-t^{10}}}$ .

From the relations (17)-(20) and (22) we obtain the system of linear equations, by solving this system for the unknown coefficients  $a_i, i = 0, 1, 2, 3, 4$ , we have:

$$\begin{bmatrix} -0.8710344 & 0 & \frac{9}{25}\pi & 0 & -\frac{151}{2310}\pi \\ 0 & \frac{2}{55}\pi & 0 & -\frac{191}{1155}\pi & 0 \\ 0.712664 & 0 & -\frac{191}{1155}\pi & 0 & \frac{1713}{23870}\pi \\ 0 & \frac{38}{1155}\pi & 0 & -\frac{914}{35805}\pi & 0 \\ -0.5693774 & 0 & \frac{1713}{11935}\pi & 0 & -\frac{231041}{2936010}\pi \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{2}{55}\pi \\ 0 \\ \frac{38}{1155}\pi \\ 0 \end{bmatrix}$$

The error of approximate solution of Equation (23) at  $n = 4$  is given by Table 1.

Table 1

$x_i$	Absolute Error
0.9	5.55111512312578 E-17
0.7	1.38777878078145 E-17
0.5	8.67361737988404 E-19
0.3	6.26804380968182 E-19
0.1	3.99218043133405 E-21
-0.1	5.96562657773244 E-21
-0.3	1.82959116606929 E-19
-0.5	6.93889390390723 E-18
-0.7	6.93889390390723 E-18
-0.9	5.55111512312578 E-17

**Example 2** Consider a Nonhomogeneous linear LogSIEs of second kind:

$$f(t) = \frac{3t^5 + \pi t^3 \sqrt{1-t^6}}{3\sqrt{1-t^6}} + \int_{-1}^1 f(x) \log|x^3 - t^3| dx, \quad x, t \in [-1, 1] \quad (25)$$

The exact solution of Eq. (25) is  $f(t) = \frac{t^5}{\sqrt{1-t^6}}$ , the corresponding of unknown coefficients

$a_i, i = 0, 1, 2, 3, 4$  can be reduced to a system of linear algebraic

$$\begin{bmatrix} 2.498216 & 0 & -0.747998 & 0 & 0.247414 \\ 0 & 0.822798 & 0 & -0.084389 & 0 \\ -1.036945 & 0 & 1.294612 & 0 & -0.3013191 \\ 0 & -0.253168 & 0 & 0.720238 & 0 \\ 0.685979 & 0 & -0.602638 & 0 & 0.885557 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.822798 \\ 0 \\ -0.253169 \\ 0 \end{bmatrix}$$

Table 2

$x_i$	Absolute Error
0.9	1.11022302462516 E-16
0.7	1.11022302462516 E-16
0.5	2.08166817117217 E-17
0.3	5.20417042793042 E-18
0.1	1.10453096321961 E-18
-0.1	1.08420217248550 E-19
-0.3	6.07153216591883 E-18
-0.5	6.93889390390723 E-18
-0.7	5.55111512312578 E-17
-0.9	2.22044604925031 E-16

The presented results in both Table 1 and 2 show an excellent convergence of the values by using Bachok–Hasham polynomials. The computations were carried out by manipulation MATLAB program and 20 significant digits as well.

## 6. CONCLUSION

Analytical and approximation solutions are provided for a class of Linear homogeneous LogSIEs with two kernel cases. Unknown functions approximated by Bachok–Hasham polynomials ( $H_n^k(x)$ ). The benefits of this method are low cost of setting up the equations and more accuracy and efficiency.

## REFERENCES

- Aboiyar, T., Kuhe, D. A., & Soomiyol, M. C. (2010). On Laguerre-Gaussians and Fredholm integral equations. *African Journal of Mathematics and Computer Science Research*, 3(10), 220-224.
- Berthold, D., Hoppe, W., & Silbermann, B. (1992). The numerical solution of the generalized airfoil equation. *Journal of Integral Equations and Applications*, 4(2), 309-336.
- Boyd, J. P. (2001). *Chebyshev and Fourier spectral methods*. Courier Corporation.
- Chakrabarti, A., & Martha, S. C. (2012). Methods of solution of singular integral equations. *Mathematical Sciences*, 6(1), 1-29.
- Corduneanu, C. (1991). *Integral equations and applications* (Vol. 148). Cambridge: Cambridge University Press.

- Eshkuvatov, Z. K., Long, N. N., & Abdulkawi, M. (2009). Approximate solution of singular integral equations of the first kind with Cauchy kernel. *Applied Mathematics Letters*, 22(5), 651-657.
- Hochstadt, H. (2011). *Integral equations* (Vol. 91). John Wiley and Sons.
- Hsiao, C. H., & Chen, C. F. (1979). Solving integral equations via Walsh functions. *Computers and Electrical Engineering*, 6(4), 279-292.
- Jafari, H., Yousefi, S. A., Firoozjaee, M. A., Momani, S., & Khalique, C. M. (2011). Application of Legendre wavelets for solving fractional differential equations. *Computers & Mathematics with Applications*, 62(3), 1038-1045.
- Mason, J. C., & Handscomb, D. C. (2002). *Chebyshev polynomials*. CRC Press.
- Mennouni, A. (2013). A regularization procedure for solving some singular integral equations of the second kind. *International Journal of Difference Equations*, 8(1), 71-76.
- Nadir, M., & Antidze, J. (2004). On the numerical solution of singular integral equations using Sanikidzes approximation. *Comperative Methods in Science and Technology*, 10, 83-89.
- Nosrati, S. M., Marasi, H. R., & Ghahramani, F. (2015). Block-pulse functions approach to numerical solution of Abels integral equation. *Cogent Mathematics*, 2(1), 1047111.
- Okecha, G. E. (2007). Solution of Cauchy-Type singular integral equations of the First Kind with Zeros of Jacobi Polynomials as interpolation nodes. *International Journal of Mathematics and Mathematical Sciences*, 2007.
- Paryab, K., & Rostami, M. (2008). Expansion method for linear Fredholm integral equations of second kind by Chebyshev, Legendre and Shannon wavelets and the comparison of their numerical results. *Mathematical Sciences Quarterly Journal*, 2(4), 335-346.
- Polyanin, A. D., & Manzhirov, A. V. (2008). *Handbook of integral equations*. CRC press.
- Sen, M., & Powers, J. M. (2001). *Lecture Notes on Mathematical Methods*. University of Notre Dame, USA.
- Sizikov, V., & Sidorov, D. (2016). Generalized quadrature for solving singular integral equations of Abel type in application to infrared tomography. *Applied Numerical Mathematics*, 106, 69-78.
- Zarowski, C. J. (2004). *An introduction to numerical analysis for electrical and computer engineers*. John Wiley and Sons.

## A REVIEW OF SECURITY METHODS FOR MOBILE CLOUD COMPUTING

Elfurgani, Z. B. M.<sup>1</sup> and Abdullah, I.<sup>2</sup>

Faculty of Science and Technology, Universiti Sains Islam Malaysia, Malaysia<sup>1,2</sup>

eng.zhra@gmail.com<sup>1</sup>, isbah@usim.edu.my<sup>2</sup>

### ABSTRACT

*Nowadays, information systems is used widely by organizations in various industries such as health, educations, and financial in order to provide accurate services at the real time. The distributed information systems have many drawbacks such as information redundancy, delay the services processing time due to needed time to retrieve information from other departments, and confliction between workers responsibilities. Cloud computing technology was developed to reduce the information systems expenses and increase the effectiveness of provided services using central services and infrastructures. Cloud computing technology provide many advantages over distributed information system such as availability of services anytime and anywhere, reduce the costs that required to install and deploy many information systems, speed up the services time through, and avoid the confliction between working tasks inside working environment through central organizational structure. Mobile cloud computing is the ability of benefits from cloud computing services through using mobile devices of users. The main aim of mobile cloud computing is to portable the services that provided by organizations between users hands anytime and anywhere. The data security of mobile cloud computing stills an issue. The security problem of mobile cloud computing is related to three main issues which are; characteristics of mobile devices, cloud computing security, and connection between mobile devices and cloud services. The main objective of this study is to review and identify the most applicable security methods and processes to increase the security performance of mobile cloud computing. The significant results of this study show that the security methods of mobile cloud computing can be classified as three levels: 1) security of mobile device like antivirus. 2) Security of connection between mobile and cloud such as data encryption. 3) Security of cloud such as access using username and password.*

**Keywords:** Cloud computing, mobile cloud computing, security, privacy.

### 1. INTRODUCTION

Nowadays, information systems is used widely by organizations in various industries such as health, educations, and financial in order to provide accurate services at the real time. Information systems are involving organizations services to increase the competitive advantage through the businesses operations in daily working environment. However, information systems

development requires high expenses which increase the challenges that face the organizations budgets (Mircea & Andreescu, 2011).

Usually the organizations adopted distributed information systems. The distributed systems can be defined as separated information systems, and each system is used separately by each department or sector in working environment. The distributed information systems have many drawbacks such as information redundancy; delay the services processing time due to needed time to retrieve information from other departments, and confliction between workers responsibilities (Ferriman, 2011).

To address the above issues, cloud computing technology was developed to reduce the information systems expenses and increase the effectiveness of provided services using technology infrastructures (Malik et al., 2012). Cloud computing is the using of virtual technology infrastructure without the need to deploy real information systems in organizations (McMillan, 2011). The information and technology infrastructures are stored and deployed using central server connected with users through network. Therefore, the users can use their computer devices to retrieve, store, and update their services information using the central cloud computing infrastructure. In other meaning, all users are connected with central information system through network using computer devices, and this central system has central storage and processors to manage the service and information of all users (Behrend et al., 2011).

Cloud computing technology provide many advantages over distributed information system such as availability of services anytime and anywhere, reduce the costs that required to install and deploy many information systems, speed up the services time through central processing and storage of all depatments services and information, and avoid the confliction between working tasks inside working environemnt througuh central orgnizational structure (Erkoc & Kert, 2011).

In last decade, the technology of mobile cloud computing has grown tremendously. Mobile cloud computing is the ability of benefits from cloud computing services through using mobile devices of users. The main aim of mobile cloud computing is to portable the services that provided by organizations between users hands anytime and anywhere. This increase the

effectiveness of services accomplishing by users and decrease the costs of accomplish working services by workers (reduce salaries) (Saggi & Bhatia, 2015).

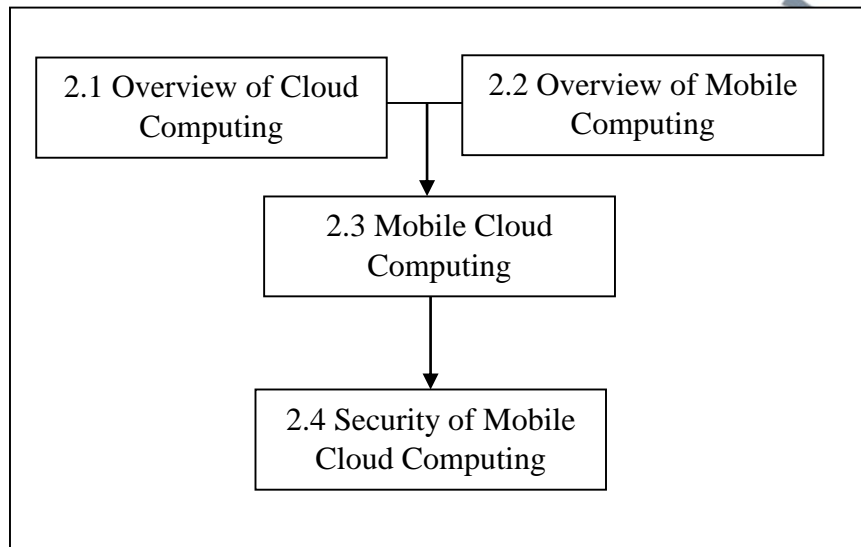
However, the security problem of mobile cloud computing is related to three main issues, which are as the following:

- i. Characteristics of mobile devices: various types of mobile devices like smart phones and IPAD features such as storage capacity, processing speed, and other features are plays important roles in mobile cloud security. The mobiles with low features delay the processing of download, upload services information which increase the opportunities of data attacking (Wang et al., 2015; Ibukun & Daramola, 2015).
- ii. Cloud computing security: the central server of cloud computing that contains all services information may be attacked when it hasn't effective security procedures like firewall and data encryption. On the other hand, the users' data authenticity of accessing cloud computing services i.e. password may stole from other people. Thus, the security procedures cannot catch the illegal accessing to cloud services (Ibukun & Daramola, 2015; Rasan & Al Shaher, 2013).
- iii. Connection between mobile devices and cloud services: the gathered services information on the connection line between mobile devices and cloud side may be attacked. Thus, it is necessary to secure the connection lines of mobile cloud computing (Khan et al., 2013; Qi & Gani, 2012).

The main aim of this research is to review the various security methods and approaches in order to increase the security performance of mobile cloud computing. Section 2 presents the literature review of mobile cloud computing and mobile cloud computing security. Section 3 provides a discussion of mobile cloud computing security. Section 4 provides some recommendations to enhance the security of mobile cloud computing. Lastly, section 5 summarizes the directions of this study and suggests future works.

## 2. LITERATURE REVIEW

Figure 1 illustrates literature review structure; sections 2.1 and 2.2 overview cloud computing and mobile computing respectively. Section 2.3 focuses on security issues of mobile cloud computing. Section 2.4 discusses related works of mobile cloud computing.



**Figure 1:** General structure of literature review.

### 2.1 Overview of Cloud Computing

According to Buyya et al. (2011), the cloud computing can be defined as management of information and services of users and organizations using central network. The aim of cloud computing is to centralized the gathered information of organization services to avoid many challenges such as data redundancy, and physical cost of services requirements. For example, if the services executed using separated 10 PCs then using cloud computing it will executed one central PC.

Rimal et al. (2009) and Juniper (2013) mentioned that there are four main cloud computing deployment approaches which are private, public, hybrid, and community (Figure 2). The description of these approaches are as the following:

- **Public Cloud:** this approach allows the deployment of cloud services via online network (McMillan, 2011). Therefore, the users' in any area around the world can benefit from cloud services based on public approach using internet network. The main advantage of public approach is the effectiveness of provide business services for global area rather than local markets which expand the targeted customers.
- **Private Cloud:** this approach allows the deployment of cloud services using internal network to deploy service inside organization (Zhang, 2010). The main aim of this approach is to reduce the infrastructure costs, and increase the information management IT system inside organization.
- **Hybrid Cloud:** this approach combine between private and public approach (Gustafsson & Orrgren, 2012; Zhang, 2010). In other meaning, some of organization services can be deployed using private approach while other are deployed using public approach. The main standpoint of this approach is that some service should be private while other services considered as non-private and can be published for any customer.
- **Community Cloud:** this approach is effective for the integrated organizations that gather services and information between each other (Amadues, 2013). Thus, many organizations within same country can share the cloud facilities due to intersection between these organizations services (e.g. governmental ministries).

Type	Properties
1. Private cloud	<ul style="list-style-type: none"> <li>• Outsource or own</li> <li>• Lease or buy</li> <li>• Separate or virtual data center</li> </ul>
2. Community cloud	<ul style="list-style-type: none"> <li>• Private cloud for a set of users with specific demands</li> <li>• Several stakeholders</li> </ul>
3. Public cloud	<ul style="list-style-type: none"> <li>• Mega scaleable infrastructure</li> <li>• Available for all</li> </ul>
4. Hybrid cloud	<ul style="list-style-type: none"> <li>• Combination of two clouds</li> <li>• Usually private for sensitive data and strategic applications</li> </ul>

**Figure 2:** Deployment approaches of cloud computing (Nilsvold, 2012)

According to IDC (2009), the main advantage of cloud computing technology is the centralization of organization services and information that involve these services. The centralization characteristic of cloud computing can provide many advantages to organizations such as reduce IT infrastructure costs, speed processing time of service through request and retrieve the needed information from various departments in real time, avoid the ambiguity of employees and departments responsibilities, availability of service anytime, and accessibility of service from anywhere.

On the other hand, Pettey (2010) had surveyed seven hundred top managers from various business companies in Euro region in order to analyze the importance of cloud computing adoring by organizations. 67% of respondents were agreeing that cloud computing technology is necessary to reduce the operational costs (including physical and management costs) of businesses. Business Cloud News (2012) argued that the main aim of organizations is to reduce the budgets that expensed on IT system of business operations. Therefore, cloud computing considered as acceptable technology to settle this issues.

Veigas et al. (2012) explained that the expense reducing is not the only advantage that gained from cloud computing implementation. The availability and accessibility of organization services and products increase customer satisfaction instead of organization that adopts traditional services and the targeted customers of provided services could be increased. Veigas et al. (2012) added that the central pool of business information allow the leaders to take accurate decisions based on effective analysis of information.

Beside the fact of cloud computing importance, there are many limitations still face this technology. Dillon (2010) argued that the main drawback of public cloud computing is the security performance. The public use online network to provide organization services which increase the oppunities of attack or stole the online gathered data. Private cloud computing can reduce the secutiy challanges of public approach due to internal deployment of cloud service (Benlian et al., 2009; Murgasan, 2011). However, the private approach is not suitable for global services which reduce the advantages of the of cloud computing implementations.

## 2.2 Overview of Mobile Computing

The mobile computing can be defined as data processing and gathering using wireless devices such as Smartphone and Bluetooth technology (Dinh et al., 2013). The main characteristic of mobile computing is the wireless connectivity between two devices anytime anywhere (wire connection is not needed). According to Fernando et al. (2013), mobile computing was developed to address many issues such as the following:

1. Connectivity architecture: people need to communicate with each other without limitations such as physical connection and connection from specific place. The communication is not only about voice calls but it involve all daily life activities such as E-learning, shopping via internet, and transportation (e.g. taxi requesting).
2. Portable machines: the traditional PCs sizes and the need of direct electric charging prevent people from own their machines anywhere. Machines like mobiles and laptops allow the users can be used anywhere and it powered using battery charging.
3. Capacity and processing: the capacity storage and CPU of modern mobiles allow the using of these devices as powerful processors.

The development history of mobile computing was started in 1990's through gather data between devices using wireless technology (bluetooth), then the researches focus on increase the bandwidth of wireless connections. Later on, the mobiles communications were developed based on wireless technology for the purpose of voice calls and SMS texting. In this stage, the positions identifications and routing of mobiles were developed (ACM, 2000). In last decade (since 2005), the mobiles as computing devices were developed and nowadays the mobiles are used widely in daily life activities as computers.

As mentioned above, there are many advantages of mobile computing such as connectivity, capacity, and portability (Traxler, 2009; Chan et al., 2006). However, there are many limitations still face the development of mobile computing. The main challenge is that the security performance of wireless connections is lower than wire connections (Dinh et al., 2013). On the other hand, the wireless devices and connections is more expansive than wire

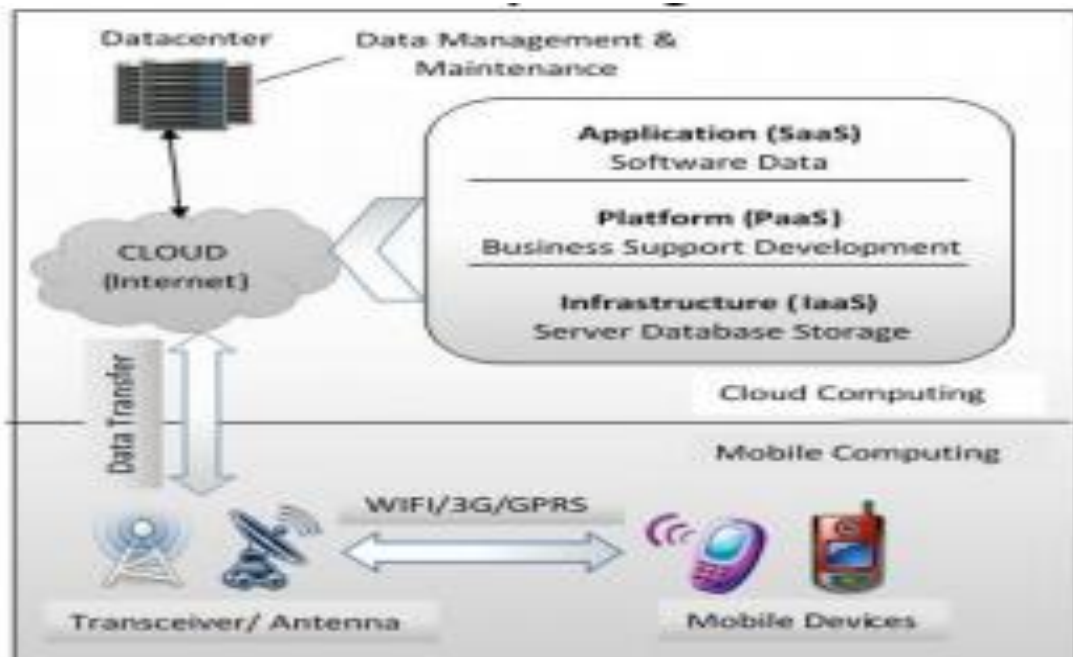
devices and connections. Moreover, the coverage space of wireless connection may weak in some places like far mountains and places outside main cities (Colley, 2010).

Wu et al. (2012) mentioned that there are two main things effects on the performance of mobile computing which are; 1) type and features of mobiles: Smart phones is more powerful than traditional mobiles because it have processors and large capacity of storage. Thus, applications such as antivirus can be installed and applied on mobiles like Smartphone. 2) feature of wireless connections such as data transferring protocols, speed of data transfer and firewall. For example, data transfer using fiber optic is more secure due to high speed of transferring processes which prevent data attacking.

### 2.3 *Mobile Cloud Computing*

The mobile cloud computing can be defined the ability to provide a convenient and rapid method for users to access and receive data from the cloud, such convenient and rapid method means accessing cloud computing resources effectively by using mobile devices.

The main architecture of mobile cloud computing (in Figure 3) is the integration between cloud computing and mobile devices. The users are connected to cloud facilities using their mobile devices. Thus, the users' interfaces are designed according to mobile characteristics rather than computer devices.



**Figure 3:** Main architecture of Mobile Cloud Computing (Donald et al., 2013)

#### 2.4 Security of Mobile Cloud Computing

There were many related works were founded to address the security issues of mobile cloud computing. Bahar et al. (2013) focused on the data security of the stored data in cloud side rather than users' mobile side. Bahar et al. mentioned that the stored data can be protected effectively using data encryption. The data should be encrypted in the cloud database and the mobiles devices should contain the decryption key of the retrieved data from cloud side. This method of encryption is called symmetric algorithm which support the data encryption using secret decoding key that owned by destination side. Bahar et al. (2013) described the data encryption in general without any details about encryption processes or methods. They also develop their work in general without reflect it on any specific environment. In the same context, Huang et al. (2011) proposed framework of mobile cloud computing for health services contains two security processes to enhance the security and privacy performance of public cloud computing. Firstly, the authentication of data is assured through generate random codes and send it to mobile number of users for login purpose. Secondly, the data are encrypted in cloud database and when the users request health information the decryption of data is conducted before send the data to users. On the other hand, Khan et al. (2013) argued that one of most effective methods to enhance the security performance of mobile cloud computing is

the signature matching between cloud side and mobile side. The mobile signature is stored in cloud database. Thus, the connection between cloud infrastructure and mobile devices considered secure if the signatures is matched. This assures that the active mobile devices in cloud environment is trusted and prevent the un-trusted mobile from use cloud infrastructure. Moreover, Ibukun & Daramola (2015) explained the importance of threats detection in the environment of mobile cloud computing. Some threats may damage or attack the users' data. The most effective method to detect the threats is installing and run antivirus application on the cloud side and on the users' mobiles. The antivirus application should suitable for mobiles (mobile specification). Most antivirus providers like Kasperky produce suitable versions for mobiles devices.

Furthermore, Qi & Gani (2012) discussed the security challenges of mobile cloud computing. The main challenges are the attacking of users' data that transferred via mobile network or stored in cloud database. Qi and Gani proposed two main solutions to enhance the security performance of mobile cloud computing which are: (1) Transfer data via secure transferring network lines such as fiber optic technology. Speed up data transferring increase the difficulty of attack users' data. (2) Assure that the cloud services are not heavy according to specifications of users' mobiles. For example, slow data download or upload due to weak specification of mobiles increase the opportunities of data attacking. Additionally, Qi & Gani (2012) mentioned that division of cloud services as public or private services reduce the security processes i.e. the public services not required security procedures.

According to Chow et al. (2010), one of the most basic privacy procurers of mobile cloud computing is assure the authenticity of mobile users through provide correct username and password. There are many rules could be applied to increase the privacy performance of users' passwords (Stajano et al., 2014); (1) Password should contain digits, letters, and special letters, (2) Password should be more than 8 characters, (3) password must not same or contain username, and (4) password should contains capital and small letters. Wang et al. (2015) agreed that the users' identification method based on users' passwords is not effective due to possibility of attack and stole the password by other people. However, this method is one of basic and important privacy procedures for mobile cloud computing. Wang et al. argued that the encryption of users' passwords could enhance the effectiveness of users' privacy. The

encryption of users passwords increase the difficulty of attack or stole the flowed passwords on mobile computing network. Rasan & Al Shafer (2013) focused on the authentication access of cloud computing using users' mobiles. They mentioned that the traditional authentication methods based on users name and passwords has many drawbacks like the ability of own or stole the users' identification by other users. They also proposed fingerprint as effective method to assure the users authentications. The users can use their mobiles to send fingerprint image to cloud side and they will access their own profiles if fingerprint image match with the stored one in cloud database.

### 3. DISCUSSION OF RELATED WORKS

According to related works, the security methods of mobile cloud computing can be classified according to many directions as the follows:

- The security processes must be conducted as two layers; (1) cloud side, (2) mobile side, and (3) connection between cloud and mobiles. Table 1 presents the security methods for each layer.

**Table 1:** Security methods according to mobile cloud layers.

Layer	Method	Source
Cloud	Data Encryption.	Rasan & Al Shafer (2013); Huang et al. (2011).
	Antivirus.	Ibukun & Daramola (2015).
	Classify the services as public or private	Qi & Gani (2012).
Mobile	- Username and passwords.	Chow et al. (2010)
	Strong passwords.	Chow et al. (2010); Stajano et al., (2014)
	Password encryption.	Wang et al. (2015).
	Fingerprints.	Rasan & Al Shafer (2013).
	Antivirus.	Ibukun & Daramola (2015).
	Suitable specifications of mobiles (i.e. speed).	Qi & Gani (2012)

Connection between cloud and mobiles	Mobile signature.	Khan et al. (2013)
	Speed network line (i.e Fiber optic).	Qi & Gani (2012)
	Agreement between gathered data amount and mobiles specifications.	Qi & Gani (2012)

It is necessary to mentioned that all above methods were developed to support the mobile cloud computing according to public deployment approach of cloud computing. The security of private mobile computing is not main issue of researchers.

- The security methods can be classified as two kinds which are data security and data privacy. The data security is about preventing other users from attack or stole user data while the privacy is about assure the users' authentication. Table 2 summarizes the security method according to security kinds.

**Table 2:** Security methods according to Security Kinds

Layer	Method	Source
Data Security	Data Encryption.	Rassan & Al Shaher (2013); Huang et al. (2011).
	Antivirus.	Ibukun & Daramola (2015).
	Mobile signature.	Khan et al. (2013)
Data Privacy	Username and passwords.	Chow et al. (2010)
	Strong passwords.	Chow et al. (2010)
	Fingerprints.	Rassan & Al Shaher (2013).
	Suitable specifications of mobiles (i.e. Speed).	Qi & Gani (2012).
	Speed network line (i.e. Fiber optic).	Qi & Gani (2012).
	Classify the services as public or private.	Qi & Gani (2012).

#### 4. RECOMMENDATIONS

Based on the discussion of related works of mobile cloud computing security, the main recommendation of this study is to develop security model for mobile cloud computing. This model can be developed based on hybrid deployment approach to increase security performance; (1) public deployment approach for general services that not require high level of security, and (2) private deployment approach for sensitive services that require high level of security. Both deployment approaches can be integrated with each other to assure the consistency of provided services by organization. This requires effective classification of all provided services as secure or general services.

Furthermore, the proposed security model can assure the security performance of mobile cloud computing through adopt effective security methods and processes for three aspects of mobile cloud computing, (1) mobile side: strong passwords, passwords encryption, fingerprint, and antivirus. (2) cloud side: encryption, antivirus, and classify the services as public or private, and (3) connection between mobile and cloud (i.e. network lines): mobile signature, Speed network line (i.e Fiber optic), and Agreement between gathered data amount and mobiles specifications.

#### 5. CONCLUSION AND FUTURE WORK

Mobile cloud computing services are useful to increase the working performance of organizations through many advantages such as availability, accessibility, reduce costs, speed up services, and central information management. The security issue is one of most important challenges that face the development and adoption of mobile cloud computing by organizations. Hence, there are many security and privacy methods can be adopted to increase the security performance of mobile cloud computing like data encryption, and strong passwords.

As future work, a secure mobile cloud computing model based on hybrid development approach could be developed in order to enhance the effectiveness of provided services by organizations and assure effective security level of mobile cloud computing.

## REFERENCES

- ACM, Communications of the ACM. (2000). Mobile and Wireless Networks. Vol. 43, No. 6.
- Amadues. (2013). Cloud, community and collaboration: Airline benefits of using the Amadues community cloud. Review Reports. Amadues Company.
- Bahar, A. N., Habib, M. A., & Islam, M. M. (2013). Security architecture for mobile cloud computing. *International Journal of Scientific Knowledge Computing and Information Technology*, 3(3), 11-17.
- Behrend, T. S., Wiebe, E. N., London, J. E., & Johnson, E. C. (2011). Cloud computing adoption and usage in community colleges. *Behaviour & Information Technology*, 30(2), 231-240.
- Benlian, A., Hess, T., & Buxmann, P. (2009). Drivers of SaaS-adoption—an empirical study of different application types. *Business & Information Systems Engineering*, 1(5), 357-369.
- Business Cloud News (2012). Cloud Computing in China – the greatest hurdle. Retrieved September 20, 2013, from: <http://www.businesscloudnews.com/platform-as-a-service/604-can-cloud-computing-prosper-in-china>.
- Buyya, R., Broberg, J., & Goscinski, A. (2011). *Cloud Computing: Principles and Paradigms*. New York, USA: Wiley Press. pp. 1-44.
- Chan, T. W., Roschelle, J., Hsi, S., Kinshuk, K., Sharples, M., Brown, T., Patton, C., Cherniavsky, J., PEA, R., Norris, C. & Soloway, E. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. *Research and Practice in Technology Enhanced Learning*, 1, 3-29.
- Chow, R., Jakobsson, M., Masuoka, R., Molina, J., Niu, Y., Shi, E., & Song, Z. (2010, October). Authentication in the clouds: a framework and its application to mobile users. In *Proceedings of the 2010 ACM workshop on Cloud computing security workshop* (pp. 1-6). ACM.
- Colley, A. (2010) “iPad sales hit 250,000 in Australia” : [www.theAustralian.com.au](http://www.theAustralian.com.au).
- Dillon, C. W. (2010). *Cloud Computing: Issues and Challenges*. IEEE , 27-33.
- Dinh, H. T., Lee, C., Niyato, D., & Wang, P. (2013). A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless Communications and Mobile Computing*, 13(18), 1587-1611.
- Donald, A. C., Oli, S. A., & Arockiam, L. (2013). Mobile cloud security issues and challenges: A perspective. *International Journal of Electronics and Information Technology (IJEIT)*, ISSN, 2277-3754.

- Erkoc, M. F., & Kert, S. B. (2011). Cloud computing for distributed university campus: A prototype suggestion. International Conference the Future of Education. Yildiz Technical University. Turkey.
- Fernando, N., Loke, S. W., & Rahayu, W. (2013). Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1), 84-106.
- Ferriman, R. (2011). The benefits of Cloud Computing. The Charity Cloud.
- Gustafsson, B., & Orrgren, A. (2012). Cloud computing: The adoption of cloud computing for small and medium enterprises. Jonkoping International Business School. Jonkoping University.
- Huang, D., Zhou, Z., Xu, L., Xing, T., & Zhong, Y. (2011, April). Secure data processing framework for mobile cloud computing. In *Computer Communications Workshops (INFOCOM WKSHPS)*, 2011 IEEE Conference on (pp. 614-618).
- Ibukun, E., & Daramola, O. (2015). A Systematic Literature Review of Mobile Cloud Computing. *International Journal of Multimedia and Ubiquitous Engineering*, 10(12), 135-152.
- International Data Corporation (IDC). (2009). Cloud computing forecasting. Retrieved September 2, 2013, from: <http://blogs.idc.com/ie/?p=543>.
- Juniper, (2013). Cloud-Ready Data Center Reference Architecture. Juniper network, USA.
- Khan, A. N., Kiah, M. M., Khan, S. U., & Madani, S. A. (2013). Towards secure mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(5), 1278-1299.
- Ko, S. K. V., Lee, J. H., & Kim, S. W. (2012). Mobile cloud computing security considerations. *Journal of Security Engineering*, 9(2), 143-150.
- Malik, S., Huet, F., & Caromel, D. (2012, October). Cooperative cloud computing in research and academic environment using Virtual Cloud. In *Emerging Technologies (ICET)*, 2012 International Conference on (pp. 1-7). IEEE.
- McMillan, B. (2011). Cloud computing: What it can do for your business. *Entrepreneurial Practice Review*, 2(1).
- Mircea, M. & Andreescu, A. L. (2011). Using Cloud Computing in Higher Education: A Strategy to Improve Agility in the Current Financial Crisis. IBIMA Publishing, Vol. 2011, Article ID 875547.
- Murugesan, S. (2011). Cloud computing gives emerging markets a lift. *IT Professional*, 13(6), 60-62.
- Nilsvold. (2012). Cloud basics–Deployment models. Retrieved August 26, 2013, from: <http://blog.visma.com/singletesting/2012/03/12/cloud-basics-deployment-models>.

Pettey, C. (2010). Press release: Gartner Says Worldwide Cloud Services Market to Surpass \$68 Billion in 2010. Retrieved September 4, 2013, from: <http://www.gartner.com/newsroom/id/1389313>.

Qi, H., & Gani, A. (2012, May). Research on mobile cloud computing: Review, trend and perspectives. In *Digital Information and Communication Technology and its Applications (DICTAP)*, 2012 Second International Conference on (pp. 195-202). IEEE.

Rassan, I. A., & Al Shaher, H. (2013). Securing mobile cloud using finger print authentication. *International Journal of Network Security & Its Applications*, 5(6), 41.

Rimal, B.P., Choi, E. & Lumb, I. (2009). A taxonomy and survey of cloud computing systems. *Fifth International Joint Conference on INC, IMS and IDC*. IEEE computer society, 44-51.

Saggi, M. K., & Bhatia, A. S. (2015). A Review on Mobile Cloud Computing: Issues, Challenges and Solutions. *International Journal of Advanced Research in Computer and Communication Engineering* Vol, 4.

Stajano, F., Spencer, M., Jenkinson, G., & Stafford-Fraser, Q. (2015). Password-manager friendly (PMF): Semantic annotations to improve the effectiveness of password managers. In *Technology and Practice of Passwords* (pp. 61-73). Springer International Publishing.

Traxler, J. (2009). The evolution of mobile learning. In R. Guy (Ed.). *The evolution of mobile teaching and learning* (pp 1-14). Santa Rosa, CA: Informing Science Press.

Veigas, J. P., Naik, V., & Chandrasekaran, K. (2012). Cloud computing adoption and impact in emerging markets. *IRNet Transactions on Computer Science and Engineering*, 1(1), 89-95.

Wang, Y., Chen, R., & Wang, D. C. (2015). A survey of mobile cloud computing applications: Perspectives and challenges. *Wireless Personal Communications*, 80(4), 1607-1623.

Wu, W., Wu, Y. J., Chen, C., Kao, H., Lin, C. & Huang S. (2012). Review of trends from mobile learning studies: A meta-analysis. *Computers & Education*, 59(2), 817-827.

Zhang, L. C. (2010). Cloud computing: state-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1), 7-18.

## AN APPROACH FOR EFFECTIVE VOLUNTEER PLACEMENT DURING POST-DISASTER

Basarang, S. N.,<sup>1</sup> Azizan, N.,<sup>2</sup> Bahari, M.<sup>3</sup> and Ismail, W.<sup>4</sup>

Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM),  
Bandar Baru Nilai, 71800, Negeri Sembilan, Malaysia.<sup>1,2,4</sup>  
Department of Information System, Universiti Teknologi Malaysia, Johor, Malaysia.<sup>3</sup>

nabilahbasiran@gmail.com<sup>1</sup>, nurdiana@usim.edu.my<sup>2</sup>, waidah@usim.edu.my<sup>4</sup>

### ABSTRACT

*Volunteer management is the management of individuals who will be responsible to do the day to day work of finding, matching, supporting and recognizing volunteers. Common flaw that occur during the disaster event is the failure to organize the volunteer team to perform needs distribution equally and these process is time consumed. This paper will critically examine the fitness function presently used in enhancing the performance of volunteer placement in a disaster area. In order to shorten the time taken for volunteer placement, this research will concentrate in evaluating the fitness function. From the experiments, it was concluded that the proposed algorithm outperforms existing hill climbing algorithm. The new fitness function will be used in volunteer placement structure so that volunteers can provide maximum coverage in affected disaster area.*

**Keywords:** Hill climbing, Fitness function, Post-disaster, Volunteer.

### 1. INTRODUCTION

The challenges of humanitarian organizations are the limited technologies that suit their needs. It has also been pointed out that these organizations experience coordination difficulties with volunteers that are willing to help (Falasca et al., 2009). During disasters, more number of volunteer are working less efficient due to the assignment division. Different kinds of network brokers play different roles in facilitating volunteering (Barraket et al., 2013). Currently, existing decision model to assist commodity flow in the management of volunteers are still lacking. In the event of a flood disaster, it is crucial to ensure that assistance and aid are provided to flood victims in an orderly and effective manner from the national level downwards. This approach is largely implemented to flood disasters cases (Chan, 2015).

Disaster management in Malaysia is traditionally almost entirely based on a government-centric top-down approach. National Security Council (NSC) under the Prime Minister's

Office is responsible for policies while National Disaster Management and Relief Committee (NDMRC) is responsible to coordinate all relief operations before, during and after a disaster (Chan, 2015). The optimization technique such as hill climbing algorithm is very fast and effective in extracting low level feature (Goyal, 2014). Combination of optimization technology and involvement of expertise could overcome the problem. Effective and timely relief activities are needed to maintain the stability in the region (Osa, 2013). This is because assets and economic activity can be affected due to disaster. Thus, adequate external aid is needed in the relief stage to help the impacted society to fully recover.

The rest of this paper is organized as follows: the second section will explain on the background. Research design section explains on how the process is carried out followed by hill climbing fitness function section that will explain on the hill climbing algorithm. Finally, conclusion of this paper is presented.

## 2. BACKGROUND

Disaster events should be handled collectively. When a flood starts to subside, people around the state or country are eager to help on the victims (Whittaker et al., 2015). Thus, efficient task assignment and preparation of tools and skills need to be properly managed to prevent chaos and disorganization.

In order to develop a new algorithm on volunteer management, several inputs need to be considered. In this research, volunteers are defined as the entities that play a big role in disaster relief and rescue process. The number of volunteers and their skills will be used as variables for the proposed model. The circle is used to represent the disaster area and it will include specific coordinate together with its radius. The disaster area landscape will also be included in the algorithm. In this work, an optimized solution in managing volunteer in term of volunteer placement to allocate in one target area by measuring the quality of fitness function is expected to be achieved.

Genetic algorithm is used to determine the optimum route using Travel Salesman Problem (TSP). Meanwhile, Chang et al. (2007) has developed a methodology for flood emergency

logistics preparation planning under uncertainty by using geographic information system (GIS). It is a framework that has been established for many disaster relevant databases. It usually do not provide any specific analysis module for planning emergency logistic. Aside from that, Yuan & Wang (2009) designed a modified Dijkstra algorithm to solve the issue. Based on the first model, chaos, panic and congestions in time of disaster are further taken into account. Hill climbing involve in the segmentation method should be nonparametric and takes the local and global feature distribution into consideration (Kochra & Joshi, 2012). From this approach, segmentation is a simple and fast nonparametric algorithm that detects the peaks of clusters (Ohashi et al., 2003). This method allows for scalable and efficient coordination in multi-robot system. In this research, multiple images can be managed and provided as inputs for the fitness function.

The clustering based on image thresholding according to target color area is needed. Otsu is the algorithm that assumes the image contains two classes of pixels following bi-modal histogram which are foreground pixels and background pixels. It then calculates the optimum where the sum of foreground and background spreads is at its minimum, so that their inter-class variance is maximal. Consequently, this method can roughly show the dark and bright areas that represent both affected and normal areas

Thus, hill climbing for random starting point and seeded starting point are selected to run the experiment and determine the best way to choose the starting point that will outperform the normal hill climbing algorithm.

### 3. RESEARCH DESIGN

The design for this research is portrayed in Figure 1. Based on Figure 1, this study will be carried out following four processes, which are:

- a) Read image

An image is loaded into the program to map the landscape. Several colors appear in the image and the target object will represent the roof of building in disaster area. From the landscape, testing will be carried out to see how well a circle can match into the target area.

b) Otsu method

The image will be converted into two major colors which are black and white. Thus the method by Nobuyuki Otsu is applied in testing. Otsu threshold is calculated by referring to the histogram concept. The goal of this method is to find the threshold value  $T$  that minimizes the weighted within – class variance  $I$ .

c) Random starting point in Hill Climbing

Hill climbing algorithm using random starting point is tested in the program.

d) Seeded starting point in Hill Climbing

Hill climbing algorithm using seeded starting point is tested in the program.

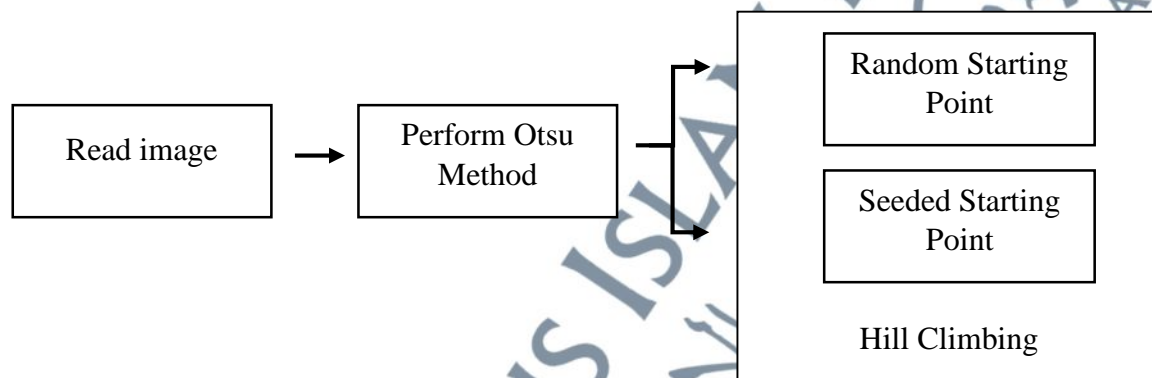


Figure 1: Research design

#### 4. HILL CLIMBING FITNESS FUNCTION

The constraints are defined based on the data collection in this work. No redundant victim is allowed to receive the basic need and no volunteer is allowed to distribute out of the circle area. In one disaster area, ‘O’ and ‘X’ marks will be used to represent the potential victims that require aid and volunteers will be sent according to number of ‘O’ with specific radius value. The basic need should be distributed and volunteers must be optimized in one or more circle area in which the area consist of number of victims per area. For example, in 5 km radius consist of 46 disaster victims, volunteer needed to distribute the basic need is calculated and assigned to the site accordingly.

#### 4.1 *Detecting Coordinates*

The initial point is necessary to locate the center of disaster site. The landscape in rectangular based with circle includes the coordinates  $(x,y)$  and parameter  $r$ . The circle is generated randomly. The result obtained is considered good if it lays on target and the coordinate between one areas to another does not overlap to indicate that there is no task redundancy assigned to one area.

#### 4.2 *Hill Climbing*

The proposed algorithm uses the concept of hill climbing. Hill climbing is a local search procedure which utilizes the concept of searching the neighborhood for a better solution. This technique iterates a number of times. In each iteration, a new point is selected from the neighborhood of the current point. If that new point provides a better value for the evaluation function, the new point becomes the current point. Otherwise, other points in the neighborhood are selected and tested again. Hill climbing naturally will stuck in local maxima during the search (Michalewicz & Fogel, 2004). This algorithm attempts to iteratively improve the current state by evaluating the function. Hill climbing and the related algorithms are shown in Figure 2.

The hill climbing technique is used for both random and seeded starting points to determine the best way in choosing the starting point.

*Input: Read the input of number of disaster points and number of volunteers*

*Coordinate landscape = VD, number of circles = N and number of iterations = ITER.*

*Create C = N circles (randomly)*

*Let Fit = F(C) applied to VD (equation 1)*

*For i = 1 until ITER*

*Create Cnew from C using change generator*

*Let Fnew = F(Cnew) applied to VD*

*If Fnew  $\geq$  Fit*

*Fit = Fnew, C = Cnew*

*End if*

*End For*

*Output: Optimum fitness function for basic needs distribution among volunteers.*

**Figure 2:** Proposed algorithm implementing hill climbing concept

## 5. EXPERIMENT AND DISCUSSION

The experiment is carried out using MATLAB software. One sample map image is used for testing purpose. Both hill climbing using random and seeded starting points uses image in Joint Photographic Expert Group (JPEG) format and the dimension for each image is 2063×867 pixels. Sample image used in this experiment is shown in Figure 3.

### 5.1 Performance Evaluation

In the test procedure, the best constructed hill climbing obtained in the learning process is applied in the image of the captured map. To test the performance of object detection, only one map image is used.

i. Determination of the landscape

At the initial stage, the landscape has to be determined.

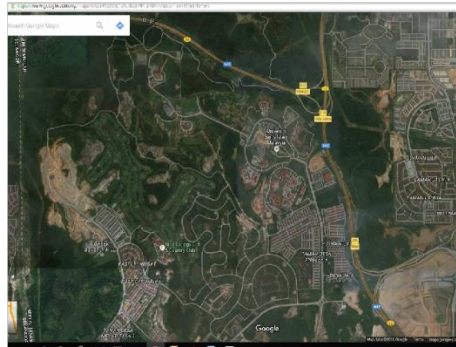


Figure 3: Sample image

ii. Determination of the fitness function

Several functions will be tested based on equation below:

$$F(C) = \frac{R(C_i) \left( \sum_{i=1}^{|C|} V(C_i) \right)}{\sum_{i=1}^{|C|} D(C_i) + 1} \quad (1)$$

where  $C$  is defined as the list of circles,  $C_i$  as the  $i$ th circle,  $|C|$  as the number of circles within  $C$ ,  $R(i)$  is the radius of circle  $i$ ,  $V(i)$  is the number of volunteers within circle  $i$  for a given area, and  $D(i)$  is the number of victim disaster point for a given image.

iii. Construction of a new fitness function

According to the test results, the fitness function will be used to evaluate the performance.

iv. Evaluation investigation of training data

Main focuses of this paper are on the construction of a new fitness function and investigation of the training data. Table 1 and Table 2 presents the outcome for Hill climbing using random starting point and Hill climbing using seeded starting point respectively.

**Table 1: Hill climbing using random starting point**

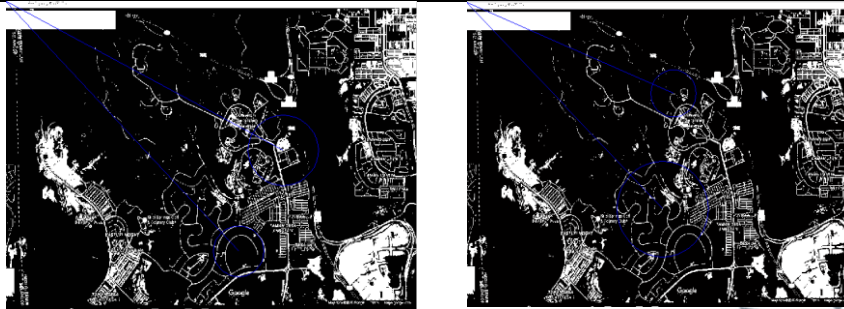


Image 1

A ( $x = 245, y = 450, r = 94$ )

B ( $x = 195, y = 330, r = 73$ )

Image 2

A ( $x = 215, y = 530, r = 69$ )

B ( $x = 124, y = 245, r = 156$ )

**Table 2: Hill climbing using seeded starting point**

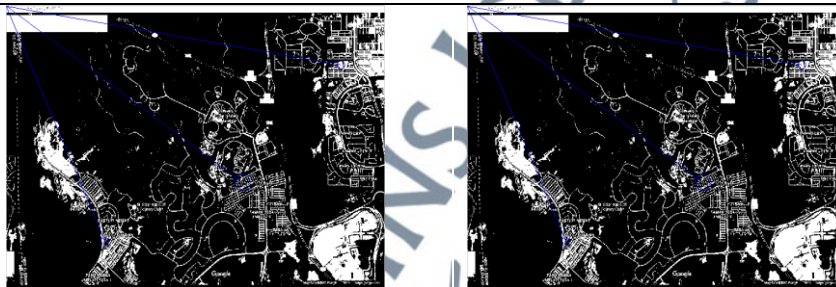


Image 3

A ( $x = 532, y = 530, r = 6$ )

B ( $x = 215, y = 330, r = 156$ )

C ( $x = 95, y = 132, r = 17$ )

Image 4

A ( $x = 215, y = 531, r = 7$ )

B ( $x = 215, y = 330, r = 156$ )

C ( $x = 95, y = 132, r = 17$ )

The hill climbing using random starting point shows that the circle is deviated from the original landscape. Referring to Image 2 in Table 1, the circle has moved to a different area, showing instability in computing the coordinate and radius. Meanwhile, referring to Image 4 in Table 2, the three circles are only 0.01% different in terms of  $x$ ,  $y$ , and  $r$  values from the initial run. Hence, the result in the evaluation using the fitness function shows that the hill climbing

algorithm using seeded starting point to determine volunteer placement in real post-disaster scene has outperformed the normal hill climbing algorithm.

## 6. CONCLUSION AND FUTURE WORK

The volunteer management for post disaster is optimized by the approach presented in this paper. Here, the preliminary work is carried out to establish the fitness function in the proposed algorithm to solve the issues. The initial work explains the constraint and situation of the disaster event. Then, the research showed that hill climbing method using seeded starting point approach provides good starting point, although several improvements are still needed in the fitness function. Hill climbing using seeded as starting point improves in showing the targeted area in the image. From this approach, the highest value of the fitness function has outperforms the original random hill climbing result. This indicates that seeding strategy works effectively in correctly pointing to the target area in the map. The approach of experimenting in different sections of the image is to show the reliability of the fitness function in different target areas. Future work will focus on defining limitation to be applied in the image where the circles generated are out of the defined coordinate landscape.

This study is part of a larger ongoing project. Results from this paper will be used in proposing a new algorithm for volunteer placement and to measure the performance of the proposed algorithm. Nonetheless, more tests, evaluations and simulations need to be conducted. It is believed that the use of fitness function in other algorithm approaches such as genetic algorithm and ant colony may produce optimum solution to address the current volunteer placement issue. The proposed algorithm and fitness function is expected to resolve management issue and could be implemented across a wide range of management problems under different domains.

## ACKNOWLEDGEMENT

This research is supported by the Universiti Sains Islam Malaysia, grant no. PPP/UTG-0213/FST/30/12213.

## REFERENCES

- Barraket, J., Keast, R., Newton, C. J., Walters, K., & James, E. (2013). Spontaneous volunteering during natural disasters. Queensland University of Technology Queensland University of Technology June.
- Chang, M. S., Tseng, Y. L., & Chen, J. W. (2007). A scenario planning approach for the flood emergency logistics preparation problem under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 43(6), 737-754.
- Falasca, M., Zobel, C. W., & Fetter, G. M. (2009). An optimization model for humanitarian relief volunteer management. In *Proceedings of the 6th International ISCRAM Conference*.
- Goyal, G. (2014). Impact and analysis of DVROFT filter on TEM image. *International Journal of Computer Science and Information Technologies*, 5(4), 5120-5124
- Kochra, S., & Joshi, S. (2012). Study on hill climbing algorithm for image segmentation. *International Journal of Engineering Research and Application*, 2(3), 2171-2174.
- Michalewicz, Z., & Fogel, D. B. (2004). Summary. In *How to Solve It: Modern Heuristics* (pp. 483-494). Springer Berlin Heidelberg.
- Ohashi, T., Aghbari, Z., & Makinouchi, A. (2003, June). Hill-climbing algorithm for efficient color-based image segmentation. In *IASTED International Conference on Signal Processing, Pattern Recognition, and Applications* (pp. 17-22).
- Osa, Y. (2013). *The Growing Role of NGOs in Disaster Relief and Humanitarian Assistance in East Asia. A Growing Force: Civil Society's Role in Asian Regional Security* (Tokyo: Japan Center for International Exchange, 2013).
- Whittaker, J., McLennan, B., & Handmer, J. (2015). A review of informal volunteerism in emergencies and disasters: definition, opportunities and challenges. *International Journal of Disaster Risk Reduction*, 13, 358-368.
- Yuan, Y., & Wang, D. (2009). Path selection model and algorithm for emergency logistics management. *Computers & Industrial Engineering*, 56(3), 1081-1094.

## A SYSTEMATIC REVIEW ANALYSIS FOR SCAM DETECTION AND RESPONSE USING APOPTOSIS

Aliyu, L. Y.,<sup>1</sup> Saudi, M. M.<sup>2</sup> and Abdullah, I.<sup>3</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>1,2,3</sup>

Islamic Science Institute (ISI), Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia<sup>2,3</sup>

yahyala1@yahoo.com<sup>1</sup>, {madihah, isbah}@usim.edu.my<sup>2,3</sup>

### ABSTRACT

*Online scam is a well-known form of fraud in which victims are tricked into making a payment either with the hope of having a better payoff or the victim being tricked into losing a lot of his or her money. There are various scamming schemes such as love scams, identity scams, auctions and merchandise scams and many more. Therefore in this research paper, a systematic analysis on existing works related with the scam detection and response techniques and apoptosis have been further investigated and evaluated. As a result, the strengths and the weaknesses have been identified and related future enhancements have been proposed in this research paper. Furthermore, this paper discusses the challenges and the potential research for future work related with the integration of scam detection and response with apoptosis. This research paper also can be used as a reference and guidance for further study on scam detection and response.*

**Keywords:** Apoptosis, malicious wares, scammers, emails, scams.

### 1. INTRODUCTION

E-mail is one of the primary ways we communicate. We not only use it every day for work, but also to stay in touch with our friends and family. In addition e-mail is how companies provide many products or services, such as confirmation of an online purchase or updates to our bank account. Since so many people around the world depend on e-mail, it has also become one of the primary methods scammers and cyber criminals use to attack others.

Scammers are known for implementing various cunning strategies to carry out their illegal activities online by deceiving innocent individuals. One of the ways these scammers use is by phishing. This type of scam is based on communication that is carried out by way of either using emails or social networks. The scammers normally send a message or messages to their

victims which will trick them into giving their login credentials which could be in the form of bank accounts, social network accounts or any personal data that can be valuable to them. These scammers will send a phishing email which will seem to have come from an official source such as financial institution or bank authorities. They trick the victims into clicking a link on the email sent to them opening a website that looks legit from the official source. The victim ends up login into the so called official site thereby giving their username and password to such phishing scammers. They will later use the obtained information to carry out illegal sales or impersonation to steal what they want, therefore putting the victim in a serious financial dilemma (Chipurici, 2016).

Another trend that these online scammers use to harm their victims is through the dating sites. Scammers are known to present one of the biggest threats regarding online dating users. For individuals that are emotionally weak, the scammers have resort to a technique called “sextortion”. The individual is lured into a webcam session with their so-called new love partner who might be an attractive man or woman. As the victim is in a video chat with the perpetrator, he records whatever activities the victim does during their session. Through this, the victims end up getting blackmailed and the scammer will try to post the video online or send it to their loved ones unless they agree to transfer a huge amount of money into their account which could also be offshore (Edmunds, 2015).

Due to the fact that these methods will require the process of detection of the signature, also leading to the distribution of the signature files to many customers on a worldwide level and making use of the new signature database for scanning, a new malicious ware such as a scam virus could be spread easily by then which will lead to a lot of damage within that time frame without being detected. Also, another setback with some scamming detection software is the problem of updating its signature database. Most of the signature databases are required to be updated by the Anti-Malware company or the user himself. But for efficient results, it is required that both parties should do the updates. The company developing the Anti-Malware products might do the update for commercial purpose but the end user must also go through such task just to stay safe from upcoming or newer Malwares. There will be no assurance that the device used will be free from viruses of the signature database is not updated (Mishra, 2010).

As a result of these numerous scam growing problems, different scam detection techniques have been developed in order to counter such issues with effective solutions. With the numerous security developments some scammers are still able to bypass the system due to the upgrades cyber criminals are also implementing.

The key problem we now face is that computer systems have become vital to many areas of our lives, but at the same time increasing in complexity and moving towards a point where we will not have the human resources to manage them (Jones, 2010). Anti-malwares and endpoint protections are mostly evaded by advanced malware developed by numerous fraudsters. It has even reached the level that Strong authentication and transaction verification solutions are also rendered useless, or also evade the system undetected. Fraudulent transaction detection systems are now facing challenges by sophisticated malware attacks (Wohl, 2014).

## 2. LITRATURE REVIEW

In recent times, there is not only interest on data fraud itself but on its detection, too. Although improvements of data fraud detection are evident, it seems that the intellectual creativity and capacity of the betrayers is unlimited. Especially, the Internet with its various services and the mobile communication opened the Pandora box for criminal acts. Furthermore, one may state the hypotheses that while the ethics behavior of people decreases over time the data fraud rate is continuously increasing. One can best unfold the complexity of data scam and fraud by classifying them into four classes. All of them are driven by three time-invariant features of the societies, business and human beings: Power, glory and money. As the saying goes “Knowledge is Power”, power is the driving force of Spying out in the military and secret service area, while knowledge and profit (money) are the main factors in business (Lenz, 2015).

In the faceless world of the Internet, online transaction fraud is one of the greatest challenges for Web merchants. Advanced solutions are needed to protect merchants from the constantly evolving problem posed by fraud. The Fraud Detection Suite arms merchants with advanced filters and tools, providing them with a customisable solution to combat the most common types of online credit card transaction fraud facing merchants today (Fraud Detection Suite, n.d).

### 3. DATA ANALYSIS

In the last few years, society have been witnessing a enormous increase in the quantity of data in the form of text, pictures, audio and video, both at global and economic level entities. This process is improved by the entry of any entity mentioned above into the virtual environment. Data comes from everywhere, from various and diverse sources such as contracts, customer interactions, call centers, social media, phones, emails, faxes, and others. The style is to use these data for the interest of the entity through conceiving strategies, opportunities identification, goodwill development, preventing and detecting fraud and other issues. The use of data analysis processes and the software devoted to these operations provide extensive and detailed analysis of such occurrences and processes of the informal economy, fraud and corruption, as the information and communication technology becomes a prerequisite instrument of registered (formal) economy (Bănărescu, 2015).

In recent times, data analysis methods are known as the most effective solution for automatic identification of scams in numerous areas. In relation to some systems, data analysis is defined or stated as the process of discovering and extracting hidden patterns from large volume of data. Majority of data analysis process is utilised in order to identify scams or fraud within the data patterns. The important benefits of data analysis and scam attack detection is the provision of the abilities to implement classes of models which could aid in identifying newer attacks which helps the professionals to be able to identify ideas in the process which could also aid in detecting newer anomalies (Nejad, 2015).

#### 3.1 *Email Scams*

E-mail is known to be one of the main ways we communicate through internet. It is not only used for every day work, but also to stay in touch with others such as friends and family. Also, e-mail is one of the main ways companies provide many products or services, such as verification of an online purchase or updates to our bank account. Since so many people around the world are dependent on e-mail for online business, it has also become one of the primary methods that cyber criminals use to attack others. In such cases, people may avoid such problems by simply avoiding or deleting the email messages (Danhioux, 2011).

Unfortunately, the scammers have upgraded their skills through social engineering which has made them come up with advance methods into trapping their victims. Phishing is a good example of social engineering. A scammer puts his or her effort into deceiving people to believing that an email message came from, example, the Outlook team, intimidating them with account closure if they didn't submit their account information. Anyone who followed such type of instructions to submit his or her account information put their identity and bank account at risk.

In October 2014, Symantec reported an escalation in a particular scam where emails were sent, regularly to a recipient working in the finance department of a company, often requesting payment by credit card or the completion of a wire transfer. The sender details were sometimes faked or made to look like they had come from the CEO or another high-ranking member of the victim's company. Money transfer details were either sent in an attachment, or required the victim to email back and request them. Web threats increased and much more aggressive in the year 2014 as weak spots in commonly used tools and encryption protocols were exposed and criminals tried to make it harder to escape their malicious grasps on target victims. The web came up with an incredibly threatening landscape in 2014, a trend that was also set to continue in 2015. Vulnerabilities and new modifications of malware underlined that website security deserves full-time, business-critical attention.

### 3.2 *Scam Detection Techniques*

Scam detection is known to be a very difficult task due to the fact that its types and natures are totally different and there are numerous types of methods. For a while, the traditional ways of data analysis have been in use to detect scams. It normally consist of hard and time consuming task that deals with different domains of knowledge like business practices, finance, economics and law. Normally Scams instances can be similar in appearance and content, but usually are not identical. So this makes scam detection a very tough task (Arif & Dar, 2015).

### 3.3 Limitations of Some Scam Detection Tools

Due to the reasons that there are numerous types of malware and scam detection tools, some of these detection tools tend to have some certain limitation due to certain requirement that needs to be fulfilled. These limitations lead to such tools to having weaknesses in their scanning processes. Even though, signature scanning is known to be an easy and definite method for detecting known specific viruses, it sometimes does not work for new or unknown viruses. Generic methods may detect new viruses but may not cure the infected files. Some of these drawbacks are stated as follows:

1. **Time lag between malware creation and detection:** - One of the serious disadvantages of signature scanning is the delay or time lag between virus creation and virus detection. As the method requires the process of detecting the signature, distributing the signature files to the customers worldwide and using the new signature database for scanning, a new virus can easily spread and cause damage during that period without being detected.
2. **Problem of updating signature database:** - A signature-scanning engine will not be useful in the absence of a virus signature database. The scanning engine has to refer to the signature database in order to scan the files. Unfortunately updating the signature database is not a simple onetime activity or task. It has to be updated from time to time on a regular basis. Updating the signature database is very important because only the viruses whose signatures have already been found out and stored in the signature database can be detected during signature scanning. But the update sometimes has to be done either by the Anti- Malware Company or the user of the anti- Malware tool.
3. **The burden of scanning time and resources:** - For the reason that the number of viruses are on the increase, the number of their signatures also increases. This condition has led to an increase in the size of the signature database and increase in the scanning time. With thousands of possible virus types and hundreds of gigabytes of file storage, a scanning operation often exceeds the acceptable limits (Mishra, 2010).

Even though there numerous types of scam detection tools, some tends to miss other types of scam attacks while other detection tools identify the problem when the damage has begun to

occur. This issue may also occur due to the fact that some scam detection tools are limited to certain aspects they are built for. Some useful detection tools includes; ThreatMetrix, Trusteer and Iovation among many.

- I. **ThreatMetrix** is known to provide visibility and insight into traffic patterns and emerging threats. The Network analyses close to two billion transactions per month, more than a third of which originate from mobile devices. These transactions are analyzed for legitimacy based on hundreds of attributes, including device identification, geolocation, previous history and behavioral analytics (Faulkner, 2016).
- II. As for **Trusteer**, it was pioneered by IBM Company. IBM Security Trusteer solutions provides detection and prevention of the full range of attack vectors which includes phishing and malware that usually target's end users and account takeovers committed by criminals making them responsible for the majority of online, mobile and cross-channel Scams. Fraudsters use malware on end user machines to steal user credentials and tamper with online banking sessions (Whol, 2014).
- III. **Iovation** Scam Prevention is a cloud-based service that dynamically manages close to a billion unique devices across a wide range of industries around the world, and has been known to protect more than 8 billion online transactions for its clients. Identifies the device used for committing scams which is an effective way to root out Scams and one that Iovation has offered its clients since 2004. Iovation also protects online businesses and their end users against Scams and abuse, and identifies trustworthy customers through a combination of advanced device identification, shared device reputation, device-based authentication and real-time risk evaluation (Iovation, n.d). The strengths and weakness of these scam detection tools can be viewed in the Table 1 below:

**Table 1:** Comparison of scam detection techniques.

Detection Tools	Strength	Weakness
<b>ThreatMetrix</b> (Faulkner, 2016)	Provides effective justification by accurately differentiating between legitimate customers and Scammers by looking holistically at behavioral trends and threat intelligence.	<ul style="list-style-type: none"> <li>- Dependence on behavioral trends might cause it to overlook or allow Morphing (changing) malwares through.</li> <li>- Requires regular Updates.</li> </ul>
<b>Trusteer</b> (IBM solution, 2014).	Pinpoint Criminal Detection including real-time correlation of events across the online and mobile channels, and provides risk assessment for mobile and online activities in both channels.	<ul style="list-style-type: none"> <li>- Open to manipulation.</li> <li>- Requires regular Updates.</li> </ul>
<b>Iovation</b> (Iovation, n.d).	Iovation's device-centric solution exposes and recognises risk unidentified by other tools, offering a critical layer of defense against cyber Scams and identity theft.	<ul style="list-style-type: none"> <li>- Being device centric might cause it to be diverted using zombie device to distract its function, leading to vulnerability.</li> <li>- Requires regular Updates</li> </ul>

### 3.4 Apoptotic Computing

In this paper, Apoptosis detection technique is a point of focus due to its unique way of dealing with these various online scamming activities.

The major advantage of Apoptosis over other computer protection software is its autonomous capability. Autonomic computing is known to be inspired and motivated by the human autonomic nervous system as a self-managing and self-regulating system which performs its functions at the background without the need of a manual or conscious effort on the part of the individual (Horn, 2001). In relating to Autonomic systems, Marshall & Dai (2005) explained that as compared to other increased system complexity, other conventional or normal systems face the problem of degradation in reliability due to accumulation of errors within the system

in which they stated that Autonomic Computing will aid in solving this kind of issues leading to production of more reliable systems also leading to their description of the autonomic properties known as Self-CHOP (Configuration, Healing, Optimisation and Protection). The general properties of an autonomic, or self-managing, computing system consist of four objectives that represent broad system requirements, and four attributes that identify basic implementation mechanisms. The autonomic system consists of the following objectives:

1. *Self-configuration*. The system must be able to readjust itself automatically, either to support a change in circumstances or to assist in meeting other system objectives.
2. *Self-optimization*: The system can measure its current performance against the known optimum and has defined policies for attempting improvements. It can also react to the user's policy changes within the system.
3. *Self-protection*: The system must defend itself from accidental or malicious external attacks, which requires an awareness of potential threats and the means to manage them
4. *Self-healing*: In reactive mode, the system must effectively recover when a fault occurs, identify the fault, and, when possible, repair it. In proactive mode, the system monitors vital signs to predict and avoid health problems, or to prevent their reaching undesirable levels.

In terms of the disciplines of Natural Computing, the Apoptosis metaphor can be viewed within the context of Artificial Immune Systems (Saudi et al., 2009). Here, the concept of Apoptosis is applied to computer system security whereby a system is built from many small units and if one becomes "damaged" e.g. through a computer virus it either initiates, or is instructed to initiate, Apoptosis, without affecting surrounding system components – the analogy being an animal cell invaded by a virus and the immune system recognising the foreign body and attacking it (Jones, 2010).

Also, according to Saudi et al. (2009), it is to carefully consider Apoptosis as a mechanism that can increase the security of numerous computer systems against attacks by malicious wares or malwares. They have shown this in their successive efforts and works on the STAKCERT Framework which they developed to control and isolate a computer attacked by a worm. Other researcher such as Sterritt (2011) also recommended a system which operates "Stay Alive"

signals to prevent agents (system) from self-destructing by default. They propose a self-regenerating system whose agents have a notion of “citizenship” in the system and are able to observe their environment. If an agent finds out a fault in itself it first tries to fix itself, and if it doesn’t work it will activate Apoptosis.

#### 4. DISCUSSION

Specific dataset related to email scams will be utilised for further study to be used in training and testing using the WEKA data mining tool.

The dataset channel to be used in upcoming research will be obtained from accessible web sources such as [www.cimining.org](http://www.cimining.org). The website contains various attributes that may be used as dataset consisting of mode of online contacts such as email address, phone numbers, and company names (through websites), ebay and many more. But specifically, scamming activities through emails tends to be one of the ways the perpetrators achieve their fraudulent schemes which could be done through phishing to harvest more information from customers. For this reason, email spam dataset will be used particularly to carry out the training and testing activities using the WEKA data mining tool.

Another task to be done in the coming research is to make use of the VM ware which will be utilised to build or develop a virtual operating system in a computer instead of a fully operational operating system.

As for detection software, even though there are available scanning detection systems, some malicious wares are able to do some damages before such threats are completely neutralised. Some scam detection software may have some limitation in its scanning method. First of all, among the most serious setbacks of malware signature scanning is the time lag between Malware creation and Malware detection.

These are some of the few limitations faced from other Malware detection tools. But with the advent of innovation in this security sector, overcoming these limitations is gaining a lot of

ground in reducing the scammers and the perpetrators of such fraudulent activities gaining access to people's private information.

## 5. RECOMMENDATION AND CONCLUSION

One of the recommended solutions to prevent scammers from using their Malwares and also overcoming the update limitations is using a detection tool with automated technique abilities for identifying malicious signatures. An automatic update is a common technique adopted at the client machine to check and download the new updates from the anti-virus developer's website (Mishra, 2010). This will also aid in the reduction of the time lags between creation of malicious wares and distribution of their signatures. Such drawback of signature scanning could also be rectified by making use of generic methods of malware detection.

As a matter of fact, Apoptosis detection software should be a recommendation in order to obtain a satisfactory and efficient protection from various scamming activities the masses are facing in this information era as the detection tool tends to fulfill of the objectives above.

Fortunately Apoptosis should be considered as a mechanism to improve the security of computer systems against attacks by such malwares in coming research. Apoptosis should also be applied as an agent base within an Autonomic environment for testing and training using email scam dataset in order to develop a new model for scam detection which will aid as a blueprint for developing an efficient system against future online threats.

## ACKNOWLEDGEMENT

The authors would like to express their gratitude to Universiti Sains Islam Malaysia (USIM) and Institute Science Islam (ISI) for the support and facilities provided. This research paper is supported by USIM grants: [PPP/USG-0216/FST/30/16916].

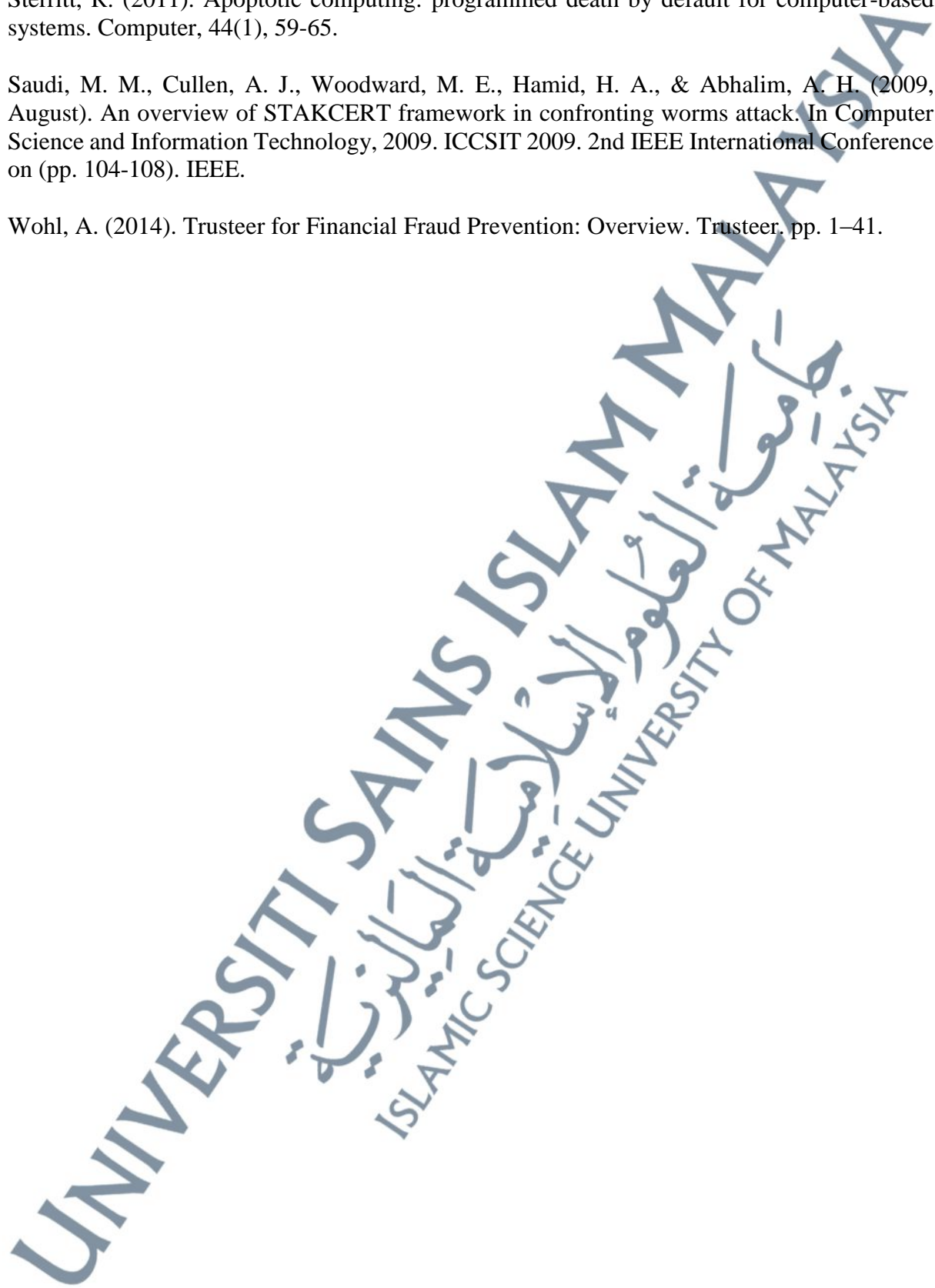
## REFERENCES

- Arif, M., & Dar, A. R. (2015). Survey on Fraud Detection Techniques Using Data Mining. *International Journal of u-and e-Service, Science and Technology*, 8(3), 165-170.
- Bănărescu, A. (2015). Detecting and Preventing Fraud with Data Analytics. *Procedia Economics and Finance*, 32, 1827-1836.
- Chipurici, C. (2016). Top Online Scams Used by Cyber Criminals to Trick You, Hiemdal Security, viewed August 14<sup>th</sup> 2016, <https://heimdalsecurity.com/blog/top-online-scams/>
- Danhieux, P. (2011). E-mail Phishing and Scams, the Monthly Security Awareness Newsletter for Computer Users, 07 December
- Edmunds, S. (2015). Scammers & Online Dating Frauds, *Global Dating Insights*, pp. 02.
- Fraud detection Suite, n.d, Regal Technologies: The future of Payments, Severna Park, MD 21146, p. 01.
- Faulkner, A. (2016). ThreatMetrix. Available at: <http://www.threatmetrix.com> (Accessed: 10 July 2016).
- Horn, P. (2001). AUTONOMIC COMPUTING: IBM's Perspective on the State of Information Technology. IBM Corporation.
- Internet Security Threat Report. (2015). Vol. 20, pp. 59.
- Jones, D. (2010). Implementing biologically-inspired Apoptotic behaviour in digital objects: An aspect-oriented approach. Master Degree. Open University.
- Lenz, H. (2015). Data Fraud Detection: A first General Perspective. Springer International Publishing Switzerland, pp. 14.
- Marshall, T., & Dai, Y. S. (2005, July). Reliability improvement and models in autonomic computing. In *Parallel and Distributed Systems, 2005. Proceedings. 11th International Conference on* (Vol. 2, pp. 468-472). IEEE.
- Mishra, U. (2010). Methods of Virus detection and their limitations. Bangalore India, 25 August, pp. 8-9.
- Nejad, R. F. (2015). The fraud detection in the bank payments and its methods. *Asian Journal of Information Technology*, 14(6), 239-245.
- Online Fraud: Your Guide to Prevention, Detection and Recovery, n. d, Microsoft, p. 01.

Sterritt, R. (2011). Apoptotic computing: programmed death by default for computer-based systems. *Computer*, 44(1), 59-65.

Saudi, M. M., Cullen, A. J., Woodward, M. E., Hamid, H. A., & Abhalim, A. H. (2009, August). An overview of STAKCERT framework in confronting worms attack. In *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on* (pp. 104-108). IEEE.

Wohl, A. (2014). Trusteer for Financial Fraud Prevention: Overview. *Trusteer*. pp. 1-41.



## A REVIEW ON INTRUSION PREVENTION SYSTEM (IPS) BASED ON BIO-INSPIRED ALGORITHM

Nadir, A. M.,<sup>1</sup> Saudi, M. M.<sup>2</sup> and Touhid, B.<sup>3</sup>

Faculty of Science & Information Technology, Daffodil International Unviersity (DIU),  
102 Shukrabad, Mirpur Road, Dhanmondi, Dhaka, Bangladesh <sup>1,3</sup>  
Faculty of Science & Technology, Universiti Sains Islam Malaysia (USIM)  
Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia <sup>2</sup>

it@daffodilvarsity.edu.bd

### ABSTRACT

*Security has been a vital issue in the enterprise network. With the innovation and flow of new technology such as ubiquitous computing, enterprise mobility, e-commerce, university ERP and cloud computing, network security still remains as a challenge. Network Intrusion Prevention System (NIPS) currently is a growing concern. Intrusion Prevention Systems provides an inline mechanism focusing on identifying and blocking malicious network activity in real time. As the packets within the network pass through the NIPS device or software, they are inspected for the presence of any attacks. In this paper, a review analysis on existing works related to the Intrusion Prevention System (IPS) has been further investigated and evaluated. The aim of this paper is to present a background of IPS and the research gaps that need to be addressed. From the review analysis, the strengths and the weaknesses of existing IPS have been identified and related future enhancements have been proposed in this paper. A proof of concept is carried out and the results for the proposed algorithm detection rate is 98.94%. The accuracy rate results is 0.997% and false positive rate is 99.00%. This paper discusses the challenges and the potential research for future work related with the integration of Intrusion Detection System (IDS) with bio-inspired algorithm. This paper could be used for reference and guidance for further studies on IDS and IPS.*

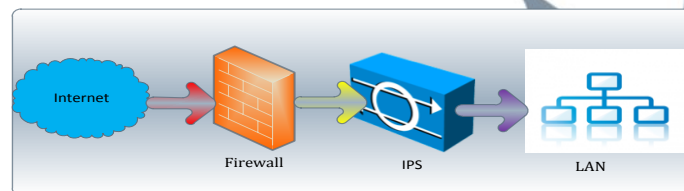
**Keywords:** *Bio-algorithm, Intrusion Prevention System (IPS), Intrusion Detection System (IDS).*

### 1. INTRODUCTION

Intrusions or network attacks have become very rampant in the Internet world nowadays. Several tools and devices have been developed to combat malware attacks or any sort of network malicious activities in order to ensure computer and network security. However, this has been a kind of “cat and mouse” game between the network inventers and the hackers or malicious network spy systems. For example, firewall is a network security device where one can run their own rules and regulations but it can be too weak to protect attack from inside and anomaly attacks. There are many anti-virus software devices available in the market but they

are unable to recognize the new viruses without updating the new rules by the vendor. Hence, the hackers or malicious network spy systems will always have the opportunities to attack the network system.

Intrusion Detection System (IDS) is a new solution which is effective in detecting any malicious threat. This system is able to send the attack alert but does not have the capability to protect the system from attacks. Therefore, there is a need for a new approach to allow the system to identify and prevent attacks. This approach is called Intrusion Prevention System (IPS) as shown in Figure 1.



**Figure 1:** Firewall and IPS

Due to lack of cyber security measures, small business organizations fall into becoming victims of the attacks hence losing huge amount of money. In order to secure the system, the enterprise uses several technology security systems, and almost 69% of them use intrusion prevention to defense from threat and attack (Kumar & Dalal, 2014).

### 1.1 IPS Classification

A Network Intrusion Protection System (NIPS) is a combination of IT systems that protects computer networks from hacking or unauthorized access. A NIPS monitors an organization's computer networks and generates event logs. Along with this alerting systems, NIPS stops malicious activities which can potentially damage and deactivate the organization's network systems.

Generally, IPS triggers an alarm. An alarm can be either a false positive or a false negative. False positives alarm occurs when the IPS reports certain damaging activity as malicious. This

requires human intervention to diagnose the event. False negatives occur when the IPS does not detect and report actual malicious activity.

## 1.2 *IPS Detection Techniques*

NIPS commonly utilizes one of three detection methods, which are signature-based, anomaly-based detection, and protocol state analysis detection.

### A. *Signature-Based Detection*

Signature-Based detection method monitors the network traffic and compares it with the preconfigured signatures to find a match. On successfully locating a match, the NIPS takes the next appropriate action. This type of detection fails to identify zero-day error threats. However, it has been proven to be effective against single packet attacks.

### B. *Anomaly-based Detection*

Anomaly-based detection method creates a baseline based on the average network conditions. Once a baseline has been created, the system intermittently create samples of network traffic using statistical analysis and compares the sample to the created baseline. If the activity is found to be out of the baseline parameters, NIPS takes the necessary action. This anomaly-based detection determines the normal network activity, such as commonly used bandwidth, protocols and interconnected ports and devices and alerts the administrator or user when anomalous traffic is detected.

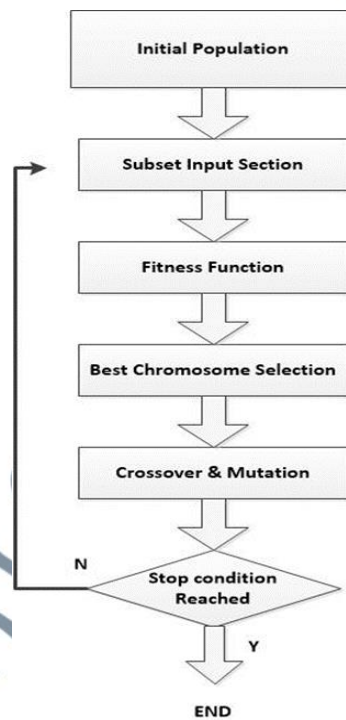
### C. *Protocol State Analysis Detection*

Protocol State Analysis Detection method identifies deviations of protocol states by comparing observed events with predefined profiles.

## 1.3 *Bio-algorithm Overview*

A Genetic or bio-algorithm is an encoding technique that mimics organic evolution as a problem-solving strategy. This is the principles of Charles Darwin of "survival of the fittest" among individuals over sequential generation for solving a problem. Genetic algorithms use a

chromosome-like data structure and evolve the chromosomes using selection, crossover, and mutation operators. The procedure generally begins with randomly generated population of chromosomes, which represent all probable clarification of a dilemma that are considered as candidate solutions. Chromosomes are divided into genes, encoded as bits, characters or numbers. A “fitness function” is used to compute the fitness of each chromosome according to the preferred solution. During evaluation, two basic operators which are crossover and mutation are used to simulate the natural reproduction and mutation of species. The selection of chromosomes for survival and combination is biased towards the fittest chromosomes (Kumar & Dalal, 2014). Figure 2 shows the structure model of simple Genetic Algorithm (Cateni et al., 2012).



**Figure 2:** Genetic algorithm structure model

The three genetic operators which are selection, crossover, and mutation are applied to each individual during the generation process. A group of suitable chromosomes are selected using a fitness function initially eliminating the other individuals. The process continues by selecting a number of individuals and making pairs with each other. The chromosome pair generates one off-string which exchanges their genes around selected cross points. Finally, some individuals

are identified and the mutation operations are applied on it. The sub attack labels such as smurf, mailbomb, saint and warezmaster are recognized according to the fitness criteria by selecting the best-fit chromosomes capable of detecting the attacks from every population (Uppalaiah et al., 2012). This research will improve the original genetic algorithm to develop an Intrusion Prevention System.

## 2. RELATED WORKS

In computer network security, there are several techniques to detect and protect attacks by using Intrusion Prevention Systems. Genetic Algorithms have also been implemented to combat such attacks.

It is identified that maximizing the detection rate accurately and minimizing the false alarm are the main problems in IDS (Hashemi et al., 2013). In order to solve these problems, they have proposed a Genetic Algorithm approach. Gupta et al. (2014) mentioned that 'pattern matching algorithm' could be implemented as the heart of IDPS. The proposed algorithm works by recognizing signatures of known attack (Gupta et al., 2014). After analysis, it was found that Boyer-Moore and KMP algorithm are more efficient than other algorithms. RobinKarp algorithm could also work efficiently if good hash function is provided (Gupta et al., 2014). Another algorithm was proposed to be used as classification algorithm of the new NIDS where it is trained and tested using the KDD data set (Soliman & Rassem, 2014). The new NIDS is compared with another detection system based on particle swarm optimization (PSO). An alternative algorithm was also proposed using a Hash Function which uses SRAM that creates fingerprints of the packet payload which are then compared to the pattern signatures (Vinod & Sagar, 2013). In addition, a NIPS design was presented with a novel pattern matching algorithm (Tzur-David, 2005). The algorithm uses a Ternary Content Addressable Memory (TCAM) and is capable of matching multiple patterns in a single operation.

Most common problems of IPS have been identified which is the detection of false positives or false negative (Ierace et al., 2005). Another problem identified occurred in IPS is related to network performance. If multiple IPSs are used in a network, every packet of data needs to stop several times from its source to destination, thus decreasing network performance. The Genetic

Algorithm was proposed and a set of rules was generated for the Intrusion detection system for detecting DoS, R2L, U2R, Probe from KDD99CUP data set population (Uppalaiah et al., 2012).

Previous studies show that IDS can only be responsible to detect the attack and IPS is responsible to detect and protect from the attack. There are several algorithms proposed by different researchers, which are based on prescribed rules. Security manager can twists such rules so that the system respond in the way they want. In this paper, a new enhance algorithm will be proposed to improve the performance of the existing solution.

### 3. METHODOLOGY

This research deals with network security using IPS. In order to secure network, a new model of IPS inspired by bio algorithm is proposed. Initially, existing research are reviewed. In this research, KDD CUP 99 (kddcup.data\_10\_percent) dataset is used for the experiment. Table 1 shows the list of tools used for this experiment.

**Table 1:** Experiment Tools.

Name	Function
Python v2.7	For designing and algorithm simulation.
Microsoft Visio	To create design.
Microsoft Excel	For creating graphs.

In this research, Python v2.7 has been used to simulate both of the existing research and proposed algorithm. The attacks implemented in the experiment are Smurf, Probe-FP, Probe and Smurf-FP. The proposed enhanced algorithm for detecting attacks is shown in Figure 3.

```

train_normal_count=train_attack_count= test_normal_count = test_attack_count = 0
read train data
read test data
for each record of train data:
    if it is a normal train
        train_normal_count <- train_normal_count + 1
    else
        train_attack_count <- train_attack_count + 1
for each record of test data:
    if it is a normal train
        test_normal_count <- test_normal_count + 1
    else
        test_attack_count <- test_attack_count + 1
get_fitness(rule):
    if rule is None
        return 0.0
    a = b = 0
    A <- train_attack_count
    B <- train_normal_count
    for each record of train data:
        if train is equal to rule but is not a normal train
            a <- a + 1
        else
            b <- b + 1
    return ( a / A ) - ( b / B )
tournament_selection(pop, k):
    best <- None
    for i <- 0 to 2
        ind <- any random value from pop list
    if best == None or get_fitness(ind) > get_fitness(best)
        best <- ind
    return best

#define initial population of 200 rules
population[]
for i <- 0 to 200:
    duration <- any random number between 0 to 2
src_bytes <- any random number between 1025 to 1050
dst_host_srv_error_rate <- randomly choose a floating point number between 0.00
to 1.00

for i <- 0 to 1000:
    parent1 <- tournament_selection(population, 2)
    parent2 <- tournament_selection(population, 2)
    child1[0] <- parent1[0], child1[1] <- parent2[1], child1[2] <- parent2[2]
    child2[0] <- parent2[0], child2[1] <- parent1[1], child2[2] <- parent1[2]
    if generated random number between 0.0 to 0.99 < 0.01:
        child1[0] <- child1[0] + any random number between 1 to 5
        child1[1] <- child1[1] + any random number between 1 to 100
        append child1, child2 to population

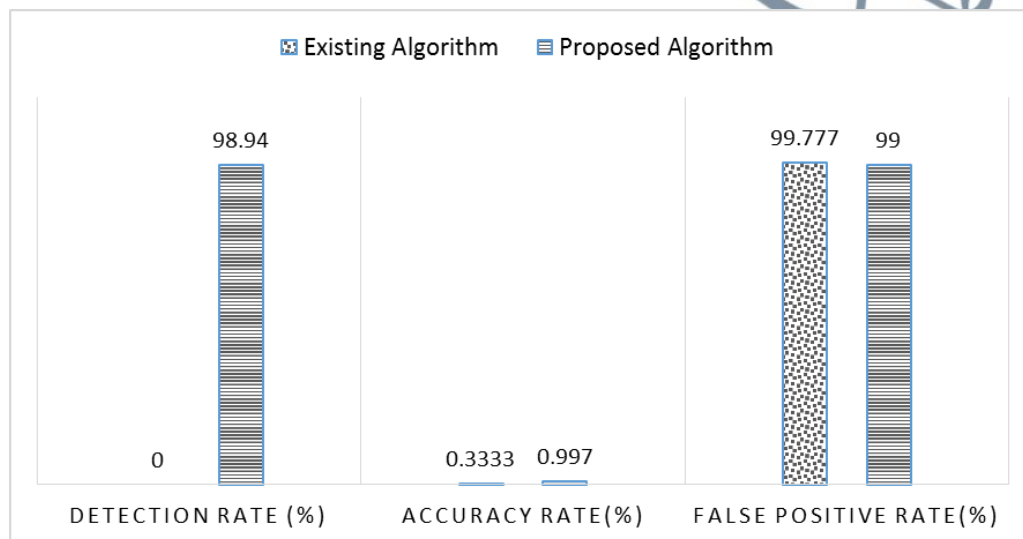
for each record of population:
    calculate best_fitness and best rule using get_fitness()
#TP <- True Positive, TN <- True Negative, FN <- False Negative, FP <- False Positive
TP = TN = FP = FN = 0
rule <- best_rule
for each record of test data:
    initially set is_intrusion <- False
    if test is equal to rule
        is_intrusion <- True
    if test is normal train
        if is_intrusion == True: FN <- FN + 1
        else: TP <- TP + 1
    else
        if is_intrusion == True: TN <- TN + 1
        else: FP <- FP + 1
detection_rate <- ( TN / (test_attack_count * 100.0))
false_detect_percentage <- (( FP / test_attack_count)* 100.0)
accuracy <- (TP + TN) / (TP + TN + FP + FN)

```

**Figure 3: Proposed Algorithm**

#### 4. FINDINGS AND RESULT

Based on the simulation result, the detection rate of the existing research by Uppalaiah et al. (2012) is 0%. The accuracy rate of this method is 0.3333% and false positive accuracy rate is 99.777%. As for the proposed algorithm, detection rate results is 98.94%, which is higher than the result obtained by the existing research. The accuracy rate of the proposed algorithm is 0.997% and false positive accuracy rate is 99.00%. Figure 4 shows the detection rate, accuracy rate dan false positive rate for both algorithms.



**Figure 4:** The detection rate, accuracy rate dan false positive rate for both algorithms.

#### 5. CONCLUSION AND FUTURE WORK

Firewalls, anti-virus, and IDS are used for system and network security, with each having unique features. The implementation of firewalls, anti-virus and IDS will ensure each organization to meet the basis to mid-level of security. IPS defense provides an in-depth approach to security and is an evolution of IDS technology. IPS proactive capabilities will help to keep system and network safe from harmful attacks by combining the firewall technique with the Intrusion Detection System. In this paper, relevant works on IDS algorithm have been reviewed and based on the review, an algorithm has been proposed. In the proof of concept, this algorithm's detection and accuracy rate results are higher than the existing algorithm by

population (Uppalaiah et al., 2012). Future work will focus on enhancing accuracy of new prevention algorithm to obtain better performance against real-time attack on networks.

## REFERENCES

Cateni, S., Vannucci, M., Vannocci, M., & Colla, V. (2012). Variable selection and feature extraction through artificial intelligence techniques. *Multivariate Analysis in Management, Engineering and the Science*, 103-118.

Gupta, V., Singh, M., & Bhalla, V. K. (2014, September). Pattern matching algorithms for intrusion detection and prevention system: A comparative analysis. In *Advances in Computing, Communications and Informatics (ICACCI), 2014 International Conference on* (pp. 50-54). IEEE.

Hashemi, V. M., Muda, Z., & Yassin, W. (2013). Improving intrusion detection using genetic algorithm. *Information Technology Journal*, 12, 2167-2173.

Ierace, N., Urrutia, C., & Bassett, R. (2005). Intrusion prevention systems. *Ubiquity*, 2005(June), 2-2.

Kumar, S., & Dalal, S. (2014). Optimizing intrusion detection system using genetic algorithm. *International Journal of Research Aspects of Engineering and Management*, 1(1), 42-45.

Soliman, O. S., & Rassem, A. (2014). A network intrusions detection system based on a quantum bio inspired algorithm. *International Journal of Engineering Trends and Technology (IJETT)*, 10, 371-379.

Tzur-David, S. (2005). *A Network Intrusion Prevention System (NIPS) for High-speed Networks*. Hebrew University of Jerusalem.

Uppalaiah, B., Anand, K., Narsimha, B., Swaraj, S., & Bharat, T. (2012). Genetic algorithm approach to intrusion detection system. *International Journal of Computer Science And Technology (IJCSST)*, 3(1).

Vinod, O., & Sagar, B. M. (2013). Hash-based string matching algorithm for network intrusion prevention systems. *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, 2(2), 2319 – 2526.

## CLUSTERING ALGORITHM TO IMPROVE NETWORK CONNECTIVITY IN WSNs FOR CORRELATED NODES BEHAVIOUR: A SYSTEMATIC REVIEW

Marsal, K. A.,<sup>1</sup> Azni, A. H.<sup>2</sup> and Ridzuan, F.<sup>3</sup>

Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia,  
Nilai, Negeri Sembilan, Malaysia<sup>1,2,3</sup>

Islamic Science Institute (ISI), Universiti Sains Islam Malaysia,  
Nilai, Negeri Sembilan, Malaysia<sup>2,3</sup>

khaledmrs17@gmail.com<sup>1</sup>, {ahazni, farida}@usim.edu.my<sup>2,3</sup>

### ABSTRACT

*Correlated node behaviour plays an important role in network performance for wireless sensor networks. In sensor networks, node may change its behaviour from normal node to misbehaved node leading to correlated node behaviour that may affect nodes connectivity. The effect of misbehaved nodes on nodes connectivity is directly proportional to the number of failed neighbouring nodes in the network cluster. The higher number of failed neighbouring nodes resulting bigger traffic load originally forwarded by those failed neighbours. Thus, the nodes in the cluster might fail faster due to excessive energy consumption. This situation is also applicable to malicious nodes where they are more likely to be compromised by their malicious neighbours. Eventually, misbehaved node leads to network partitioning hence degrading the network's performance caused by the unavailability of the nodes. Thus, finding the best clustering algorithm to be implemented for nodes partitioning is the key to solve this problem. The objective of this paper is to review the existing clustering algorithm for network connectivity which can be implemented to improve network performance in wireless sensor network. Based on the comprehensive reviews on clustering, the strengths and the weaknesses of each algorithm are identified. Additionally, this paper also discusses the parameters that contribute to connectivity in the presence of correlated node behaviour. This review on clustering algorithms may be used as a future reference for research related to clustering algorithms as well as in the area of correlated node behaviours for wireless sensor networks.*

**Keywords:** *Clustering algorithm, network connectivity, wireless sensor networks, correlated node behaviours.*

### 1. INTRODUCTION

One of the fundamental properties of the wireless sensor network, WSN is its connectivity. However, maintaining connectivity in WSN is extremely difficult due to the dynamic changing of network topology. Several techniques that can be used to understand the connectivity level for a given network topology are shown by Jain & Chand (2016). This paper reviews the

existing literature with respect to network connectivity and clustering algorithm in WSN. Network connectivity is an important property of WSN because the nodes frequently change in the network topology. Hence, if the network becomes disconnected, the data may not be sent to the desired destination. Most of the existing works have considered unrealistic scenarios however minimal works have discussed the major factors contributing to network disconnectivity as well as highlighting the important parameters that affect network connectivity. Meanwhile, the main issues in network connectivity are mobility prediction, node behaviour, energy consumption and connectivity efficiency and cost. As a result, these issues need to be explored further to understand network connectivity.

In this paper, the author conducted Systematic Literature Review (SLR) on clustering algorithm in WSNs by using the guideline for performing SLR published by Kitchenham et al., (2009). SLR is described as a mean of identifying, evaluating and interpreting all obtainable research relevant to a particular research question, or topic area, or phenomenon of interest. Systematic literature reviews can be an effective means for triangulation multiple sources of data studies to answer Research Questions, RQ with a considerable standard of the confidence of research. According to Garousi & Mäntylä (2016) the objective is to systematically categorize the secondary studies in software testing. In the same paper, the authors proposed that tertiary studies can serve as a summarizing reference which facilitates finding the most relevant information from secondary studies and thus to support evidence-based decision research. In this case, the results are derived from scientific experiments making it implementable in any given area of telecommunications and software engineering in Garousi & Mäntylä (2016). In SLR state of affairs, evidence means the best quality scientific studies on a specific topic and uses a statistical method to integrate the results. For instance of studies that apply SLR are Portocarrero et al. (2014), Farias et al. (2016), Kushwaha & Gupta (2014) and Ibukun & Daramola (2015).

The objectives of this SLR in this chapter are threefold, 1) to review systematically the clustering algorithm in WSNs, 2) to examine the methods and parameters used to formulate clustering algorithm and 3) to highlight open issues for future research in clustering algorithm to improve network connectivity in WSNs of correlated node behaviour.

## 2. REVIEW METHOD

This section describes the review process undertaken in this study as illustrated in Figure 1. The first stage involves a process of constructing research questions. Then, the search process is presented which includes a source of selection and search keywords. In this case the goal of the review is to assess SLR which is referred to as secondary studies, so this study is categorised as a tertiary literature review. Next stage is to apply the inclusion-exclusion criteria towards the selected studies, as a means to further filter the relevant studies. This then followed by determining the quality criteria in primary articles. During data extraction stage, the information in the articles is extracted, collected and organized in table.



**Figure 1:** Stages involved in performing Systematic Literature Review

### 2.1 Formulating Research Questions

The review objectives are defined to answer the following research questions in Table 1.

**Table 1:** Research Questions in SLR

No RQ	Research Questions
RQ1	What is the current knowledge on clustering algorithm to improve network connectivity?
RQ2	What are the methods used in clustering algorithm to improve network connectivity?
RQ3	What factors influence the network connectivity?
RQ4	What are the parameters required to measure network connectivity?

For RQ1, the research question tries to obtain research trend in clustering algorithm on how to improve network connectivity in WSNs for correlated nodes behaviour. To present RQ1, the number of published journals and conferences dated since 2010 until 2016 are identified. To answer the research question, the main topic and research problems in the literature are analysed. In order to study on clustering algorithm to improve network connectivity in WSNs for correlated nodes behaviour, RQ2 evaluates the methods used in clustering algorithm to improve network connectivity. Furthermore, RQ3 is developed to find the factors that influence network connectivity, so that the recommendations for future research in the area of network connectivity can be used. Finally, RQ4 considers the parameters required to measure network connectivity.

## 2.2 Search Processes

In stage 2, the search process involves manual search of specific conference proceedings and journal papers since 2010. To retrieve articles that are best suited to the research area, the search was done in five difference stages from the following digital databases:

- IEEE Digital Library
- ACM Digital Library
- Science @Direct
- SpringerLink
- GoogleScholar

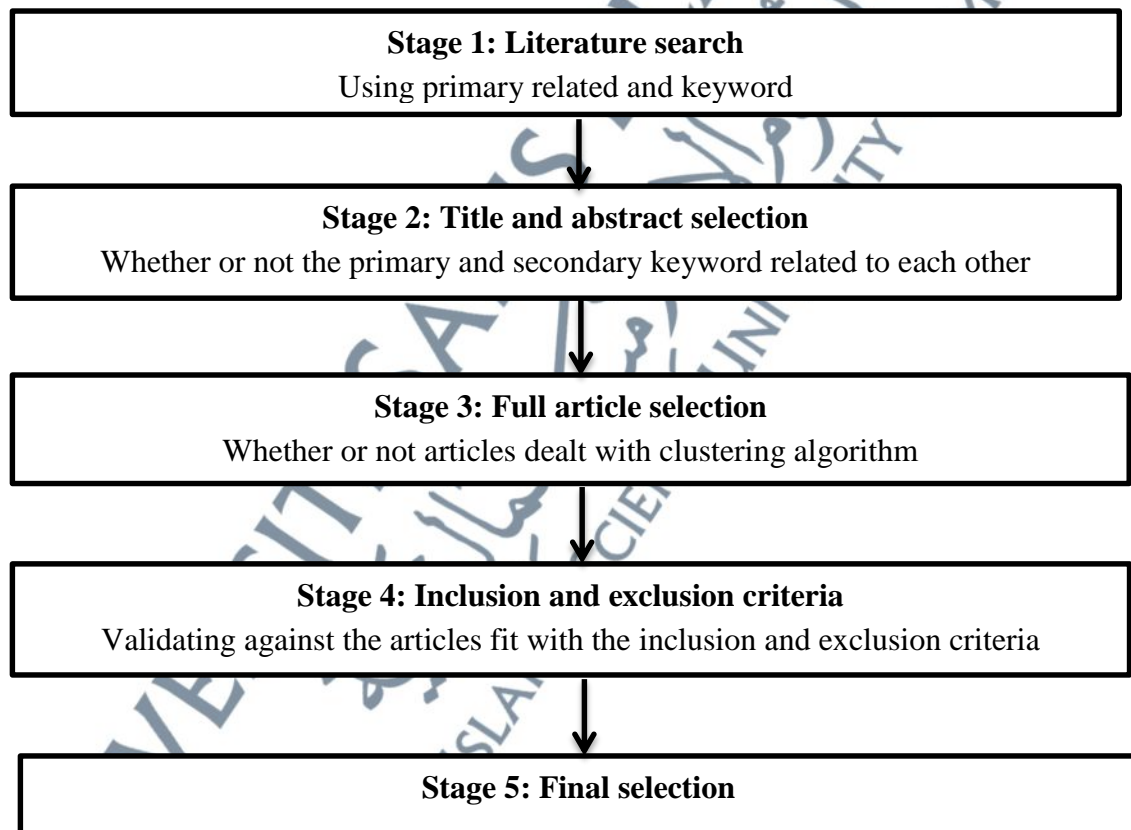
The outline of the search process is detailed out in Figure 2. In stage 1, five databases mention above is used to search articles. The initial search was using keyword in the article title and abstract using primary keywords “Clustering algorithm”, and “network connectivity” or secondary keyword “wireless sensor networks” and “Correlated node behaviour”, in digital databases. In stages 2, the title and abstracts of the retrieve articles were read and relevant articles were downloaded at stage 2. Then, thorough readings were done to select relevant articles in clustering algorithm to improve network connectivity at stage 3. Articles selected from stage 4 will undergo inclusion exclusion process based on the criteria mention in section 2.3 to be categorized as final selection. The final selection of articles at stage 4 will be evaluated and analysed to answer research questions.

### 2.3 Inclusion and Exclusion Criteria

Peer-reviewed articles on the study, published between 2010 and 2016, were validated against the inclusion and exclusion criteria. This is to ensure that articles selected are answering the specified RQs.

For inclusion criteria, the articles must fulfil the following criteria:

- Included articles where the topic only related to network connectivity a clustering algorithm in WSN.
- Only journals and conference papers were selected.
- Paper discussing the procedures used for WSN.
- Only peer reviews papers are selected.



**Figure 2:** Stages in articles selection process for SLR

In contrast, exclusion criteria are as follows:

- Papers not describing clustering for WSN.
- Papers not written in English.
- Papers that have no validation reported.
- Duplicate reports of the same study when several reports of a study exist in different journals.

#### 2.4 Quality Assessment

In this stage each article was evaluated based on York University, Centre for Reviews and Dissemination (CDR) Database of Abstracts of Reviews of Effects (DARE) criteria. The criteria are created based on four quality assessment (QA) questions:

- **QA1.** Are the review's inclusion and exclusion criteria described and appropriate?
- **QA2.** Is the literature search likely to have covered all relevant studies?
- **QA3.** Did the reviewers assess the quality or validity of the included studies?
- **QA4.** Were the basic data or studies adequately described?

The questions were scored based on Quality Assessment Criteria listed in Table 2.

**Table 2:** Quality Assessment Criteria

Quality Assessment	Decision	Description
QA1	Yes	the inclusion criteria are explicitly defined in the study
	Partly	the inclusion criteria are implicit
	No	the inclusion criteria are not defined and cannot be readily inferred
QA2	Yes	the authors have either searched 4 or more digital libraries and included additional search strategies or identified and referenced all journals addressing the topic of interest
	Partly	the authors have searched 3 or 4 digital libraries with no extra search strategies, or searched a defined but

		restricted set of journals and conference proceedings
	<b>No</b>	the authors have search up to 2 digital libraries or an extremely restricted set of journals
<b>QA3</b>	<b>Yes</b>	the authors have explicitly defined quality criteria and extracted them from each primary study
	<b>Partly</b>	the research question involves quality issues that are addressed by the study
	<b>No</b>	no explicit quality assessment of individual primary studies has been attempted
<b>QA4</b>	<b>Yes</b>	Information is presented about each study
	<b>Partly</b>	only summary information about primary studies is presented
	<b>No</b>	the results of the individual primary studies are not specified

The scoring procedure was Y = 1, P = 0.5, N = 0, or unknown if the information is not specified.

### 3. RESULTS AND DISCUSSION

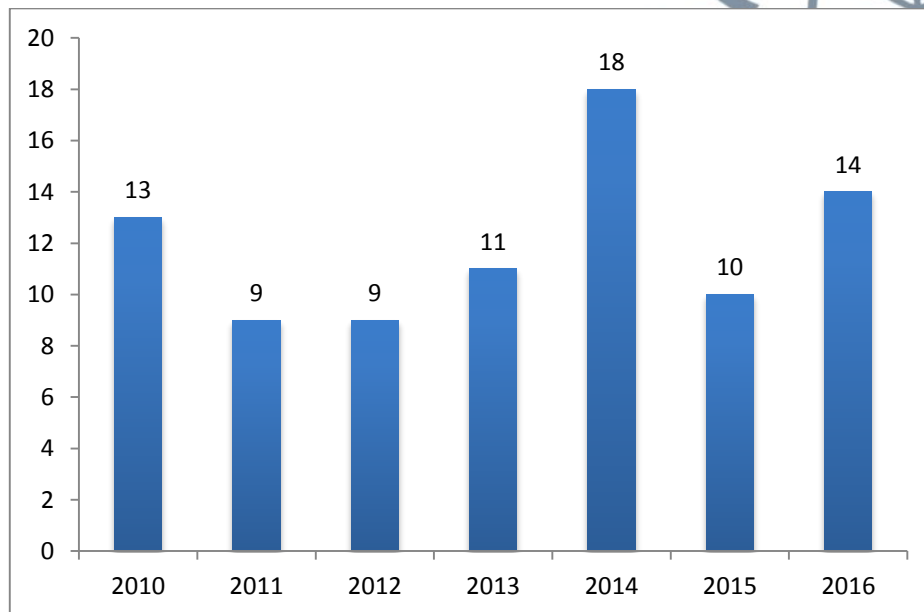
This section summarizes the results of the study for all of the research questions mention above.

#### **RQ1: What is the current knowledge on clustering algorithm to improve network connectivity?**

To answer this question, the analysis will be based on publication years and the quantities of the articles. Table 3 and Figure 3 show 84 articles published in 2010 to 2016 before the quality assessment was done. It is noted that clustering algorithm research in network connectivity was not the main focus in early years of wireless research. The research of this topic becomes more demanding, mainly because there are many applications or systems implemented in wireless network and wireless technology in improving nowadays. For this reason, it is worth to consider on improving network connectivity quality, which may be affected due to several factors including the power and battery of wireless devices used in wireless technology.

**Table 3:** Sources Searched for Years 2010-2016

Year	2010	2011	2012	2013	2014	2015	2016	Total
Number of sources found	13	9	9	11	18	10	14	84
Number of selected sources	-	5	3	11	6	4	6	34



**Figure 3:** Sources Searched for Years 2010-2016

After a thorough quality assessment only the articles with the quality score of 3.5 were selected for this research as shown in Figure 4. This process resulted only 34 articles that will be used to answer research questions RQ2, RQ3 and RQ4.

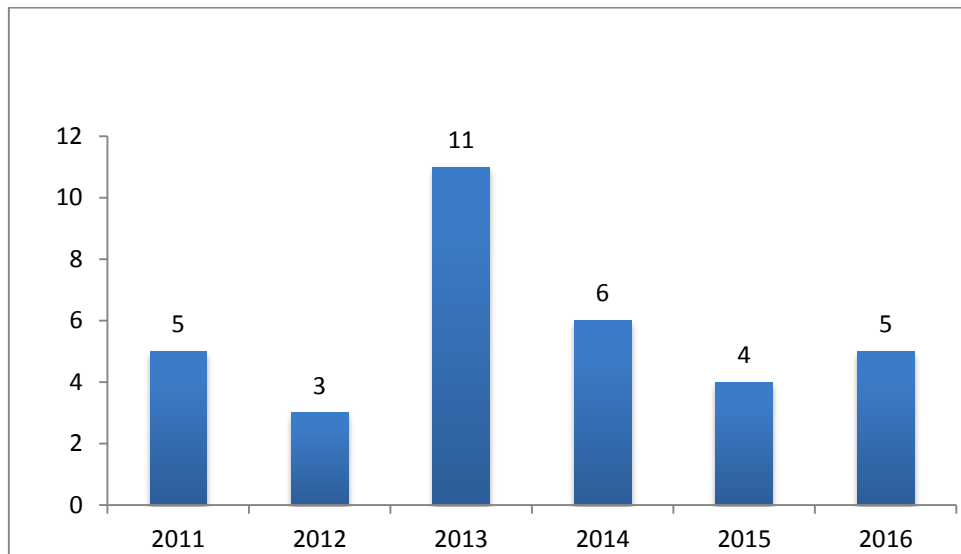


Figure 4: Selected Articles with Quality Score 3.5

**RQ2: What are the methods used in clustering algorithm to improve network connectivity?**

In recent years, wireless communication technology and the rapid development of the miniaturization and low cost sensor nodes has accelerated the development of the wireless sensor network. WSN is a large number of static or mobile sensor nodes which form the wireless network using self-organization and multi-hop method. The purpose of WSN is to collaborate detection, processing and transmitting the object for monitoring information in the areas where the wired network is impossible. The major objective of this study is to present clustering algorithm method as the initial significant developments of clustering approach in WSNs. Based on the 34 selected articles, there are ten methods to be reviewed and analysed as shown in Table 4. The methods listed below.

**Table 4:** Clustering Approach in WSNs

Clustering	Authors	Methods	Advantages	Disadvantages
LEACH	Lee & Cheng, 2012; Attea & Bara'a, 2011; Muthulakshmi, 2013.	Network lifetime Aggregation/fusion cluster count cluster sizes energy consumption,	All nodes can transmit with enough power to reach the base station.  Each node has enough computational power to support different MAC protocols.	Nodes located close to each other have correlated data.
TL-LEACH	Abidoeye et al 2011; Nayyar & Gupta, 2014.	network lifetime Aggregation/fusion cluster count cluster sizes energy consumption,	It reduces the amount of cluster head that need to transmit to the sink, effectively reducing the total energy usage.	It however may not be energy efficient if the CH is at range from the sink
EEC	Kumar, et al 2011; Liu, 2012; Driss & Abdelilah, 2013; Zhang, 2013.	network lifetime Aggregation/fusion cluster count cluster sizes energy consumption,	Improved energy utilization throughout the network, full connectivity can be achieved for a longer duration.	Algorithm that resolves the problem that clusters at a greater distance from the base station requires more energy for transmission
Load Balancing	Farahmand et al., 2016; Mirsadeghi et al., 2014.	energy consumption,	Intra-cluster management duties	Algorithm more issue in WSN where CHs are picked from the available sensor.
Fault-Tolerance	Virmani et al., 2015; Khoozani, 2013	network lifetime Aggregation/fusion cluster count cluster sizes	others nodes (backup nodes or other active nodes)	Cluster failure

		energy consumption,		
Highest-Connectivity Cluster Algorithm	Gupta et al., 2015.	network lifetime Aggregation/fusion cluster count cluster sizes energy consumption,	The connectivity of a node is considered. The node with the highest connectivity connected to the most number of nodes is elected CH.	The node with the lowest ID prevails.
Weighted Clustering Algorithm (WCA)	Li et al., 2011; Lee, 2012; Karimi et al., 2014; Afsharfarnia & Karimi, 2014.	network lifetime Aggregation/fusion cluster count cluster sizes energy consumption,	weight is calculated by each node and transmission across the entire neighbours	This is algorithms only use limited features of the nodes
R-HEED	Gupta et al., 2015; Mardini et al., 2014.	network lifetime cluster count cluster sizes energy consumption,	Residual energy of each node and intra-cluster communication cost is to determine the node to join the cluster.	At the beginning of every round the CHs wait a pre-defined period of time to receive a re-clustering message from the sink. If they do not receive the re-clustering message, they will continue rotating the cluster head within the same cluster.
PEGASIS	Madani et al., 2012; Peiravi et al., 2013; Rana & Vhaktar, 2015;	network lifetime Aggregation/fusion cluster count	node communicates with its nearest neighbour	

	Gupta et al., 2015.	cluster sizes energy consumption,		
GROUP	Vishwkarma et al., 2013; Wang et al., 2013.	network lifetime Aggregation/fusion cluster count cluster sizes energy consumption,	larger number of nodes	data is transmitted short ranges

**RQ3: What factors influence the network connectivity?**

Network connectivity is affected by mobility prediction of the node that determines the estimation of its future locations. The definition of location depends on the link of wireless networks. In infrastructure networks, location refers to the access points to which mobile terminal is connected. In infrastructure less networks, the location refers to geographical coordinates (Jain & Chand, 2016). Its main advantage is to predict the link expiration time to improve the node connectivity and routing performance. Many location prediction methods are discussed in Kanniche & Kamoun (2010). However, no studies discuss the prediction of future locations considering realistic scenarios. This is one of the biggest challenges in network connectivity. For example, the mobility of the nodes can affect the outcome of the wireless network. Where for the group mobility wireless network, the nodes move from cluster head groups, so the mobility of the nodes within the same cluster head group depends on the mobility of that cluster head group.

Static node can be the best solution to connect two clusters in highly dynamic environment. In order to deploy static nodes in place of critical nodes, deployment cost of static nodes must be considered. While placing such nodes attention must be paid on to maximize the connectivity and to have minimum cost. But to achieve both connectivity efficiency and cost objective while deploying the critical points is an NP-complete problem. Authors attempted this part in Romoozi & Babaezi (2011) but no work has done to find the critical links and critical nodes

considering the both factor cost objective and connectivity efficiency. Since it is NP-complete problem, more optimized technique needs to be followed in order to achieve this goal. Due to long distance data transmission, the nodes can consume more energy to neighbours. Besides, in WSNs, sensor nodes have limited battery resources. Hence, sensor nodes need to extend the network lifetime via balanced energy consumption.

Energy Consumption is the key challenges to save energy to prolong the network lifetime considering the network connectivity constraints. Since the energy is the most valuable resources in network connectivity, its status should continuously be monitored after network deployment (Kumar et al., 2010). Researchers have considered the random waypoint mobility model, which is an unrealistic model. Due to assuming that required energy for data transmission of the node is proportional to the square of transmission distance, the best clustering algorithm to minimize energy consumption for data transmission must minimize the sum of the square of data transmission distance in a network.

**RQ4: What are the parameters used to measures network connectivity?**

There are several methods described in primary studies that define and quantify the parameters to evaluate clustering algorithm to improve network connectivity. To design a wireless network it is necessary to identify the key parameters in WSNs such as network life, data gathering and security. In addition, the parameters that affect the performance of wireless networks are such as node degree, distance, transmission range, mobility energy consumption, bandwidth, packet loss, link lifetime, and failure rate. Furthermore, wireless network performance will be degraded when a network is under attacked or there is a node or link failure. Fundamentally, network connectivity also determines QoS based on throughput, energy consumption, bandwidth, and packet loss.

To achieve connectivity network, there must be a wireless multi-hop path from each mobile node to other mobile node. Therefore the connectivity depends on the number of nodes on area on the node density and of their transmission range. Each mobile node contributes to the connectivity of the entire network. Bandwidth (throughput) measures the speed in how quick data can be transmitted once it starts to flow. In other words, throughput is the amount of data

that is sent over a certain amount of time or the amount of time consumed to transfer a certain amount of data from mobile node to another mobile node. In wireless network, the packet loss refers to the failure of packets when in transmission data transfer to each node due to weak signal strength, interference by nature or human, noise, hardware, software failure. The loss of a packet in networking causes evident performance issue, which will affect the wireless network (Gebreyohannes, 2014). There are also studies on the security aspect of ad hoc network as network connectivity parameters such as confidentiality, integrity, and authenticity. These parameters are used in qualitative and normally immeasurable are listed in Table 5.

**Table 5:** Index and parameters (Azni et al., 2015).

Index	Parameters
Connectivity	Node degree
	Distance
	Transmission range
	Mobility
Performance	Throughput
	Energy consumption
	Bandwidth
	Packet loss
Fault-tolerance	Link lifetime
	Failure rate
	Transmission range
	MTTF
	MTBF
Security	Confidentiality
	Integrity
	Authenticity

#### 4. CRITICAL ANALYSIS AND SUMMARY

LEACH provides the following key areas of energy savings of overhead is wasted making the decision of which node become CH as each node decides independent of other nodes. CDMA allows clusters to operate independently, as each cluster is assigned a different code. Each node

calculates the minimum transmission energy to communicate with its CH and only transmits with that energy level. LEACH provides the following improvements over conventional networks by Lee & Cheng (2012) and Attea & Bara'a (2011). LEACH reduces transmission energy by a factor of 8 versus MTE and direct-transmission. The first death occurs in LEACH 8 times later than that of MTE, direct-transmission and static clustering. In addition the final death of a node occurs more than three times later than that of the other listed protocols.

TL-LEACH improves upon LEACH by utilizing a two-level hierarchy (Abidoeye et al., 2011). The energy improvements are achieved from smaller transmission distance for the majority of nodes. This network configuration requires that merely a few nodes transmit large distances. Simulations have shown that the addition of the two-level hierarchical algorithm TL-LEACH results in an improvement of network lifetime by around 30% versus its basis algorithm LEACH.

PEGASIS: The minimization of energy in this algorithm is achieved from four areas. During a given round, only one node in the network is transmitting data to the base station. Since the transmission range to the base station is large, this can result in an improvement with regards to energy savings. Since each node communicates with its nearest neighbour, the energy utilized by each node is also minimized. Each node performs data-fusion, effectively distributing the energy required for this task across the network. The overhead associated with dynamic cluster formation during each round is eliminated. Based on their simulations in C have showed that PEGASIS can result in a 100% to 300% improvement over LEACH for a variety of different network sizes and topology (Gupta et al., 2015).

EEC: Minimization of energy consumption in EEC is accomplished in a similar manner to that of LEACH (Driss & Abdelilah, 2013). Zhang (2013) however used the algorithm to improve on LEACH. This is accomplished by creating dynamic cluster sizes which are a function of the distance from the base station to the cluster. This addresses the larger transmission power requirements for nodes at a greater range from the sink. The energy utilization rate  $\eta$  is used as a comparison measurement for energy in the simulations of this algorithm. It is the ratio of the total energy consumed in the network at the time the first node dies, to the total initial energy. This measurement is related to the efficient spread of energy in the network. The EEC protocol

has shown a 35% improvement in network lifetime versus the original LEACH in their simulation.

GROUP: Energy conservation is achieved by the lower transmission distance for upstream data. In LEACH, a CH must transmit data to the base station directly, while in GROUP, data is transmitted short ranges along the upstream path. Based on their simulations showed that energy consumption is related to the number of nodes in the network. With nodes, the energy consumption is comparable to that of LEACH, however for a larger number of nodes, GROUP can offer a reduction in the maximum energy consumption per node by a factor (Wang et al., 2013). It can also be seen that the energy distribution for a larger number of nodes is more consistent with GROUP than it is with LEACH.

In this HEED algorithm, the network life time is prolonged through reducing the number of nodes that compete for channel access CH updates, regarding cluster topology through an overlay among CH which has a small network diameter. Comparing HEED to a generic weight-based clustering (GC) protocol such as WCA (Gupta et al., 2015). When using a GC algorithm, the number of iterations grows quickly as the cluster radius increases, so each node has more neighbours. Implying a node has to wait longer for higher weighted nodes to decide which cluster to join the node. The protocol is assumed to propagate the node residual energy throughout the network; and node selects a CH in its range proximity, which is not assumed to span the entire network area. The first feature will allow for better selection of CHs than original LEACH and it also prolongs the network lifetime. The second feature allows for multi-hop networks. HEED improves network lifetime over generalized LEACH, because generalized LEACH randomly selects CHs, thus resulting in a faster death of some nodes. HEED is determined based on two important parameters which are residual energy of each node and Intra-Cluster Communication Cost is to determine the node to join the cluster. An enhanced version to HEED, namely Rotated HEED (RHEED), conducts the setup phase according to certain rules and schedule, with HEED performs this step at the beginning of some rounds. The results showed that the RHEED outperforms the HEED protocol by more than 20% in term of network lifetime and residual energy in Mardini et al. (2016).

In terms of energy consumption, the algorithm tries to achieve the most stable cluster architecture, meaning after the first iteration the algorithm is executed only when there is a demand. This reduces system updates and hence computation and communication costs. The algorithm performs better than existing heuristic (Karimi et al., 2014) in terms of reaffiliation in time. Another important feature of this algorithm is that the CHs are chosen among the nodes that have enough energy available. This leads to a fair CH distribution amongst nodes; avoiding the problem of power drainage for nodes that serve as CHs for long periods of time.

Load balancing is a more pressing issue in WSNs where CHs are picked from the available sensors. To find the optimal set of cluster heads members, a new methodology (LEGAC) using two-stage GA was proposed (Farahmand et al., 2016). The first stage is dedicated to select the optimal cluster heads. The optimal assignment of cluster members to the selected cluster heads is done in the second stage. The objective of the proposed GA methodology is to minimize energy consumption, minimize number of orphan nodes and maximize coverage. Thus, the lifetime of the network will increase. Using a benchmark case study, the performance of this technique is compared with similar methods UCFIA, GCA, SCP and a method that proposed by Mirsadeghi et al. (2014).

Highest connectivity clustering algorithm a node with maximum number of neighbour's maximum degree is chosen as CHs. The neighbours of a CH become members of that cluster and can no longer participate in the election process. Since no CHs are directly linked, only one CH is allowed per cluster. Any two nodes in a cluster are at most two hops away since the CH is directly linked to each of its neighbours in the cluster. Basically, each node either becomes a CH or remains an ordinary node. This system has a low rate of CH change but the throughput is low. Typically, each cluster is assigned some resources which are shared among the members of that cluster. As the number of nodes in a cluster is increased, the throughput drops. The reaffiliation count of nodes is high due to node movements and as a result, the highest-degree node the current cluster head may not be re-elected to be a CH even if it loses one neighbour.

#### 4. CONCLUSION

SLR was conducted in this study to understand the issues related to clustering algorithm to improve network connectivity in WSNs of correlated node behaviour. This study has explained the process of selecting and reviewing literature according to guidelines for performing Systematic Literature Review published by Kitchenham et al. (2007) style. The paper shows the lack of clustering algorithm approach to avoid partitioning in WSN. Partitioning will disconnect the network and connectivity will not be achieved. Furthermore, the method for clustering algorithm focus on individual nodes which make the measurement for connectivity is not commutative and network performance not accurate. The review result includes parameters for performance measurement and the factors that affects network to be in critical condition. It is noted from the literature that network connectivity practices still fixates the theoretical works, with much work on ensuring connectivity of nodes to guarantee network performance will not degrade. From this point of view, each one of these method are only suitable for a specific application, rather than any arbitrary wireless network. In this, clustering algorithm is proposed to cater partitioning issues by taking factors and parameters define in the selected papers. Finally, misbehaved node leads to network partitioning hence degrading the network's performance caused by the unavailability of the nodes. Thus, finding the best clustering algorithm to be implemented for nodes partitioning is the key to solve this problem. As wireless technology keep growing, more challenges need to be addressed. Therefore, the key points to resolve this problem are to reduce the effect network connectivity, enhance method to improve network connectivity, and enable the reliable transmission of field information. Thus, this paper proposed promising improvement over conventional clustering algorithm. Many energy improvements thus far have focused with minimization of energy associated in the cluster-head, in which will examine some future research problems and draw conclusions on the current state of wireless network clustering.

#### REFERENCES

Abidoeye, A. P., Azeez, N. A., Adesina, A. O., & Agbele, K. K. (2011). Ancaee: A novel clustering algorithm for energy efficiency in wireless sensor networks. *Wireless Sensor Network*, 03(09), 307–312.

- Afsharfarnia, A., & Karimi, A. (2014). A New Clustering Algorithm Using Links' Weight to Decrease Consumed Energy in MANETs. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 12(2), 411-418.
- Azni, A. H., Ahmad, R., Noh, Z. A. M., Hazwani, F., & Hayaati, N. (2015). Systematic Review for Network Survivability Analysis in MANETS. *Procedia-Social and Behavioral Sciences*, 195, 1872-1881.
- Driss, E., & Abdelilah, M. (2013). Clustering algorithms based on energy efficiency in wireless sensor networks: Survey, *ARPN Journal of Engineering and Applied Sciences*, 8(10), 785–795.
- Farahmand, E., Sheikhpour, S., Mahani, A., & Taheri, N. (2016, March). Load balanced energy-aware genetic algorithm clustering in wireless sensor networks. In *Swarm Intelligence and Evolutionary Computation (CSIEC), 2016 1st Conference on* (pp. 119-124). IEEE.
- Farias, C. M. D., Li, W., Delicato, F. C., Pirmez, L., Zomaya, A. Y., Pires, P. F., & Souza, J. N. D. (2016). A systematic review of shared sensor networks. *ACM Computing Surveys (CSUR)*, 48(4), 51.
- Garousi, V., & Mäntylä, M. V. (2016). A systematic literature review of literature reviews in software testing. *Information and Software Technology*, 80, 195-216.
- Gebreyohannes, M. B. (2014). *Network Performance Study on Open Stack Cloud Computing*. Master's Thesis Spring 2014.
- Gupta, S. K., Ranjan, A., & Sharma, G. (2015). Clustering Techniques for Wireless Sensor Networks. *International Journal of Emerging Trends in Science and Technology*, 2(06), 2694–2698.
- Ibukun, E., & Daramola, O. (2015). A systematic literature review of mobile cloud computing. *International Journal of Multimedia and Ubiquitous Engineering*, 10(12), 135-152.
- Jain, M., & Chand, S. (2016). Issues and challenges in node connectivity in mobile ad hoc networks: a holistic review. Division of Computer Engineering, Netaji Subhas Institute of Technology, NSIT, University of Delhi, New Delhi, India.
- Kanniche, H. & Kamoun, F. (2010) Mobility prediction in wireless ad hoc networks using neural networks. *Journal of Telecommunications*, 2, 95-101.
- Karimi, A., Afsharfarnia, A., Zarafshan, F., & Al-Haddad, S. A. R. (2014). A novel clustering algorithm for mobile ad hoc networks based on determination of virtual links' weight to increase network stability. *The Scientific World Journal*, 2014.
- Khalil, E. A., & Bara'a, A. A. (2011). Energy-aware evolutionary routing protocol for dynamic clustering of wireless sensor networks. *Swarm and Evolutionary Computation*, 1(4), 195-203.

- Khoozani, P. A. (2013). *A Low-Complexity Architecture and Framework for Enabling Cognition in Heterogeneous Wireless*. Queen's University.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and Software Technology*, 51(1), 7-15.
- Kumar, V., Jain, S., & Tiwari, S. (2011). Energy efficient clustering algorithms in wireless sensor networks: A survey. *IJCSI International Journal of Computer Science Issues*, 8(5), 1694-0814.
- Kumar, N., Kumar, M., & Patel, R. B. (2010). Coverage and connectivity aware neural network based energy efficient routing in wireless sensor networks. arXiv preprint arXiv:1003.3558.
- Kushwaha, V., & Gupta, R. (2014). Congestion control for high-speed wired network: A systematic literature review. *Journal of Network and Computer Applications*, 45, 62-78.
- Lee, W. C. Y. (1995). Design Aspects of a CDMA Cellular Radio Network. In *Code Division Multiple Access Communications* (pp. 269-282). Springer, Boston, MA.
- Lee, J. S., & Cheng, W. L. (2012). Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication. *IEEE Sensors Journal*, 12(9), 2891-2897.
- Li, Q. Q., Gong, H., Liu, M., Yang, M., & Zheng, J. (2011). On prolonging network lifetime through load-similar node deployment in wireless sensor networks. *Sensors*, 11(4), 3527-3544.
- Liu, X. (2012). A survey on clustering routing protocols in wireless sensor networks. *Sensors*, 12(8), 11113-11153.
- Madani, S. A., Hayat, K., & Khan, S. U. (2012). Clustering-based power-controlled routing for mobile wireless sensor networks. *International Journal of Communication Systems*, 25(4), 529-542.
- Mardini, W., Yassein, M. B., Khamayseh, Y., & Ghaleb, B. A. (2014). Rotated Hybrid, Energy-Efficient and Distributed (R-HEED) Clustering Protocol in WSN. *Transactions on Communications*, 13, 275-290.
- Mirsadeghi, M., Mahani, A., & Shojaee, M. (2014). A Novel distributed Clustering Protocol using fuzzy logic. *Procedia Technology*, 17, 742-748.
- MuthuLakshmi, M. V. (2013). Advanced leach protocol in large scale wireless sensor networks. *International Journal of Scientific & Engineering Research*, 4(5), 248-254.
- Nayyar, A., & Gupta, A. (2014). A comprehensive review of cluster-based energy efficient routing protocols in wireless sensor networks. *IJRCCT*, 3(1), 104-110.

- Peiravi, A., Mashhadi, H. R., & Hamed Javadi, S. (2013). An optimal energy-efficient clustering method in wireless sensor networks using multi-objective genetic algorithm. *International Journal of Communication Systems*, 26(1), 114-126.
- Portocarrero, J. M., Delicato, F. C., Pires, P. F., Gámez, N., Fuentes, L., Ludovino, D., & Ferreira, P. (2014). Autonomic wireless sensor networks: a systematic literature review. *Journal of Sensors*, 2014.
- Rana, J., & Vhaktar, S. (2015). Comparative study of PEGASIS and PDCH protocols in wireless sensor network. *International Journal of Computer Applications (0975-8887)*.
- Romoozi, M. & Babaezi, H. (2011). Improvement of connectivity in mobile ad hoc networks by adding static nodes based on realistic mobility model. *International Journal of Computer Science Issues*, 8, 176-183.
- Virmani, D., Kaur, S., & Jain, S. (2015). Secure and fault tolerant dynamic cluster head selection method for wireless sensor networks. *Procedia Computer Science*, 46, 989-996.
- Vishwkarma, R. K., Kumar, A. & Arora, S. (2013). A Survey on Energy-Efficient protocol for Wireless Sensor Networks. *International Journal of Advanced Research in Computer Engineering & Technology*, 2(2), 756–760.
- Wang, Y. H., Lin, Y. W., Lin, Y. Y., & Chang, H. M. (2013). A grid-based clustering routing protocol for wireless sensor networks. In *Advances in Intelligent Systems and Applications-Volume 1* (pp. 491-499). Springer, Berlin, Heidelberg.
- Zhang, M., & Yungang, L. (2013). Energy-balanced distributed clustering algorithm in wireless sensor networks. *Journal of Software*, 8(8), 2008-2014.

## QUALITY EFFECT ON STUDENTS' SATISFACTION TOWARDS STUDENT INFORMATION SYSTEMS (SIS) IN A UNIVERSITY

Mohamed, Y. O.,<sup>1</sup> Fabil, N.,<sup>2</sup> Alwi, N. H. M.<sup>3</sup> and Azizan, N.<sup>4</sup>

Faculty of Science & Technology, Universiti Sains Islam Malaysia (USIM),  
Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia<sup>1,2,3,4</sup>

yomd007@yahoo.co.uk<sup>1</sup>, norasikin.fabil@usim.edu.my<sup>2</sup>, najwa@usim.edu.my<sup>3</sup>,  
nurdiana@usim.edu.my<sup>4</sup>

### ABSTRACT

Nowadays, Student Information System (SIS) are used for various academic purposes in many universities. This is a case study on the system quality of SIS in a university. This paper aims to investigate the factors that might influence the system quality of SIS which are usability, functionality, flexibility, convenience, data quality, responsiveness, accessibility and timeliness towards students' satisfaction. The study population consists of the target respondents of the sampling frame which include users of USIM's SIS. This quantitative study is carried out based on purposive sampling method in which the postgraduate students have been randomly chosen as the research samples. Based on the data analysis carried out on the 186 completed questionnaires, the result shows that system qualities have a strong impact on the SIS output and affect the level of satisfaction of users to use the systems frequently. This paper recommends that the developers of SISs should pay high attention to the development process and focusing on the dimensions of SIS quality to enhance the system's usage. The main contribution of this paper is to extend the current knowledge on the dimensions of system quality of SIS in academic application.

**Keywords:** Student Information Systems, Information Quality, Usability, Satisfaction.

### 1. INTRODUCTION

A Student Information System (SIS) is a software application applied for managing student data in universities. SIS provides capabilities for entering the database of the university and check student test results, discipline records, and other assessment scores associated with the students that enhance the communication between the students and university management (Gagliardi, 2014).

SIS is used to build student schedules, track student attendance in the university, and the most advantageous use of this system is to manage other student-related data needs as well as enabling feedback from parents of the student for the management of the university that could

enhance the management of the whole educational process in the university (Mansour & Mohamed, 2009).

The system quality of SIS is assumed to be the most influential factor for users to use SIS for data exchange and data request. Thus, providing these two factors in any information system is essential to increase the efficiency of the information systems. System quality can be measured in terms of usability, functionality, flexibility, convenience, data quality, responsiveness, accessibility and timeliness (DeLone & Mc Lean, 2003).

This paper will focus on system quality as one of the success factor of information systems. System quality is assumed as the independent factor that affects user's satisfaction. Specifically, the purpose of this paper is to investigate the influences of the dimensions of SIS's system quality which are: usability, functionality, flexibility, convenience, data quality, responsiveness, accessibility and timeliness towards the satisfaction of postgraduate students.

## **2. RESEARCH METHODOLOGY**

This research is a quantitative based and by using purposive sampling method, the postgraduate students have been randomly chosen for the research samples. The population of the research is USIM SIS users. The data instrument is questionnaire used for collecting the primary data from the respondents. Data analysis is conducted using simple linear regression.

## **3. STUDENT INFORMATION SYSTEM**

Student Information Systems development involves with a variety of fields such as: the design and analysis of systems, information security, computer networking, decision support systems, and database management (Mansour & Mohamed, 2009). SIS is an academic field that deals with the generation, collection, organization, storage, retrieval, and dissemination of recorded knowledge. Furthermore, it is a collection of related components designed to support operations, management, and decision making in an organization. Information System supports people or users in making intelligent decisions based upon the information derived from reliable data. With the growing demand for accountability brought up by different state and

federal government initiatives, many university districts have adopted a number of student information systems (SISs) to manage student academic data. Educational experts contend that proper usage of student academic data subsequently leads to better decision-making.

#### 4. QUALITY OF INFORMATION SYSTEMS

System quality represents the quality of information processing itself, which is characterized by the employment of state-of-the-art technology, a system offering key functions and features (denoted as IS excellence), and software that is user friendly, easy to learn, and easily maintainable (denoted as IS value) (Gorla et al., 2010).

Some of the recent studies that have validated IS success models are Rai et al. (2002), Iivari (2005) and Wu & Wang (2006). Rai et al. (2002) validated two models of IS success, DeLone and McLean (1992) and Seddon & Yip (1992), with student as the users of an integrated student information system. The research used five variables (system quality, information quality, perceived usefulness, user satisfaction, and IS use) to test the models at the individual level. The research found that both IS success models exhibited a reasonable fit to the data. Iivari (2005) tested the DeLone and McLean model (1992) using a mandatory system with employees of a municipal corporation as subjects. They used individual impacts and found significant results for most paths of the success model. Wu & Wang (2006) tested the DeLone and McLean model (2003) without service quality and found good support for the success of knowledge management system.

The researcher argues that providing easy access alone is insufficient in attracting users to continue in using IS. It is important to understand all elements that enhance the systems quality such as usability, responsiveness, information quality, and timing of getting the output from the system. In addition, user's satisfaction is found to be a critical factor and difficult to be achieved without understanding the IS success factors.

## 5. USER’S SATISFACTION

User’s satisfaction refers to “a subjective measure of system success and its usefulness to the user”. This means that user’s satisfaction on information system will be different in meaning and interpretation from individual to other (Ang & Koh, 1997). For example, user who is evenly satisfied with the same system following a certain measurement may not be equally satisfied according to another one. User satisfaction measurement has a critical role in informing service improvement and quality. It allows the organization to identify users need, system’s effectiveness and actions to be taken to improve the service delivery. This is because high quality service leads to more satisfied users and satisfaction is considered to be based on expectations of user and perception of service quality (Ekinci, 2003; Christou & Sigala, 2002).

## 6. QUALITY OF SIS

The quality of any information system is a critical matter for both system developers and IS vendors. Today, system quality of SIS is the key to success of education institutes. The quality of SIS allows universities to keep up and meet current quality levels, meet the user’s requirement and fulfil user’s satisfaction (Karwowski, 2011; Albertas and Jelena, 2005).

In the field of information systems, there are several elements contributes in developing a quality system. The following systematic literature review shows the main quality elements of SIS:

**Table 1:** Findings of system quality elements from literature.

<b>Authors</b>	<b>Quality Element</b>	<b>Finding</b>
Karwowski et al., 2011	Usability	<p>Usability includes methods of measuring the interface of information system in human-computer interaction and computer science.</p> <p>Usability represents the elegance and clarity of a computer application and its design.</p> <p>Usability differs from user satisfaction and user experience because usability also considers usefulness.</p>

<p>Caplinskas &amp; Gasperovic (2005)</p>	<p>Functionality</p>	<p>One of the main elements of system quality is usability which refers to the ease of use and learnability of a human-made object.</p> <p>It is the degree to which the designed product or system will perform to meet its intended purpose.</p> <p>Functionality is defined as the set of features necessary to describe requirements of a future system. Functionality is one of the most important characteristics of internal quality of information system.</p> <p>The functionality of information systems in general and SIS in particular is very important to ensure that students use the system frequently and gain their acceptance and satisfaction on the overall functionality of the system.</p> <p>Users facing complicated functionality may not use the system again due to complexity and difficulty in obtaining the output from the system.</p>
<p>Boh &amp; Tang (2003)</p>	<p>Flexibility</p>	<p>It is the ability of IS to respond to different requirements by users, where IS should be flexible enough to cater various demands of users.</p> <p>IS must be able to accommodate a certain amount of variation regarding the requirements of the supported business process and fulfill the demand of users</p> <p>Insufficient flexibility can reduce the overall lifespan of a system in cases where it prevents system changes in order to accommodate changes of the supported business process.</p>
<p>Miller (2010) Wang (2007)</p>	<p>Data Quality</p>	<p>Data quality or the term information quality is used to express the content's quality for information system. It is the capability of the system to communicate with the user in a method he will understand.</p> <p>Data quality is frequently practically identified as the fitness to use the information provided.</p> <p>Information content is a key factor in application success and an application</p>

		information system should serve as a decision-support system by providing detailed information about the transaction support and product.
Larry (2009) Robert (2013)	Responsiveness	<p>Responsiveness is defined as the particular capability of the system or functional part for completing assigned tasks through a given time. A fast response application is a proven system for dealing with the ever mutable user's requirements and request, and valuable information system is one of the spines for having a successful program.</p> <p>The responsiveness of SIS is among the most important factor that determines information system quality.</p>
Moreno & Martinez (2013)	Accessibility	<p>Information systems should be easily accessible by users regardless of their experience with IS.</p> <p>Companies can increase the number of possible customers or users to their online applications by making their web site more accessible.</p> <p>When designing accessible webpages and computer applications, professionals must take into account the diverse ways, determined by various personal, technological and contextual factors, in which individuals' access web content.</p>
Bouzeghoub (2004)	Timeliness	<p>Timeliness in information system refers to whether information requested by the user is up-to-date and available to the user in an accepted time without being outdated.</p> <p>"Timeliness" is used to represent the system ability to deliver valuable information in time to the information seeker or the end user of the system that store or generate the information.</p> <p>Timeliness of information systems means having information immediately or in the right time when information requester needs it.</p>

## 7. RESULTS AND DISCUSSION

This paper investigated the correlation strength between SIS quality and user's satisfaction of postgraduate students in the university campus. The data is collected using questionnaire which consist of a set of items. A questionnaire is often designed for statistical analysis pre-defined constructs of relationships between variables.

The population in this study is defined as USIM's SIS users. USIM postgraduate students are randomly chosen as sample for this research. 186 completed questionnaires are used in the data analysis phase. In the correlation analysis, the study estimates a sample correlation coefficient ( $r$ ), more specifically the Pearson Product Moment correlation coefficient. Correlation matrix is important because it is used to test the degree of association between the variables predefined in the theoretical framework below:



**Figure 1:** Theoretical model between System Quality and User's Satisfaction

Before the correlation between system quality and user's satisfaction is examined, it is important to test the reliability of the questionnaire. The total score of all items is computed to estimate the consistency of the whole questionnaire. This test is necessary before measuring the reliability of each variable (Parsian & AM, 2009). Table 2 shows that the reliability coefficient Cronbach's Alpha of all items is 94.40%. Based on the above orderings of alpha values, it is found that the overall reliability of the questionnaire is "Excellent". Therefore, the reliability of the current questionnaire is highly consistent and can be further used for statistical analysis.

**Table 2:** Cronbach's Alpha for all items of the questionnaire.

<b>Cronbach's Alpha</b>	<b>Cronbach's Alpha Standardized Items</b>
<b>All items</b>	94.40%
Usability	88.20%
Functionality	72.50%
Flexibility	89.44%
Responsiveness	85.45%
Accessibility	76.82%
Timeliness	81.20%
Data quality	90.12%

The next step is to measure the correlation coefficient, which is a measure of linear association between two variables (system quality and user's satisfaction). Values of the correlation coefficient are always between -1 and +1. The correlation between two variables can be either positive or negative. A correlation coefficient of +1 indicates that two variables are perfectly related in a positive linear relationship (i.e., higher levels of one variable are associated with higher levels of the other). A correlation coefficient of -1 indicates that two variables are perfectly related in a negative linear relationship (i.e., higher levels of one variable are associated with lower levels of the other). Instead, a correlation coefficient of 0 indicates that there is no linear relationship between the two variables. The output of the correlation matrix table is shown in Table 3.

**Table 3:** Pearson Correlation Matrix of research variable

<b>SIS_quality</b>	<b>Pearson Correlation</b>	<b>1</b>	<b>.725**</b>	<b>.774**</b>
	Sig. (2-tailed)		.000	.000
	N	186	186	186
	Sig. (2-tailed)	.000		.000
	N	186	186	186

<b>User_satisfaction</b>	Pearson	.774**	.779**	1
	Correlation			
	Sig. (2-tailed)	.000	.000	
	N	186	186	186

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

Table 3 shows that the correlation between the independent variable (system quality) and the dependent variable (user's satisfaction) are strong ( $r^{**} > 0.30$ ,  $\rho < 0.05$ ,  $\rho = 0.000$ ). Pearson correlation between system quality and user's satisfaction = .774\*\* (statistically significant), positive and strong. This result implies that as the quality of the SIS increased, the users' satisfaction will also increase. Therefore, SIS developers should focus on the factors of SIS in order to increase user's usage and improve user acceptance to the system. Moreover, SIS developers should realize the strong impact of system quality elements in future development of SIS.

## 8. CONCLUSION

This paper examined the effect of system quality and its dimensions on user's satisfaction of SIS. Previous studies showed that it is essential to use information system for communication between students and university management in order to improve the academic achievements of student. In addition, scholars indicate that SIS technologies have brought significant changes in the way universities handle student data and interact with the community. The use of SIS is related in improving student academic achievement and increasing academic collaboration. A large body of literature supports the view that SISs affect student performance. The finding from this research shows that system quality affect user's satisfaction significantly.

## REFERENCES

Ang, J., & Koh, S. (1997). Exploring the relationships between user information satisfaction and job satisfaction. *International Journal of Information Management*, 17(3), 169-177.

Bouzeghoub, M. (2004, June). A framework for analysis of data freshness. In Proceedings of the 2004 International Workshop on Information Quality in Information Systems (pp. 59-67). ACM.

Caplinskas, A., & Gasperovic, J. (2005). Functionality of information systems specification language: concept, evaluation methodology, and evaluation problems. *Information Systems Development*, 341-351.

Christou, E., & Sigala, M. (2002). Conceptualising the measurement of service quality and TQM performance for hotels: the HOSTQUAL model. *Acta Turistica*, 14(2), 140-169.

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30.

Ekinci, Y. (2003). An investigation of the determinants of customer satisfaction. *Tourism Analysis*, 8(2), 193-196.

Iivari, J. (2005). An empirical test of the DeLone-McLean model of information system success. *ACM Sigmis Database*, 36(2), 8-27.

Karwowski, W., Soares, M. M., Stanton, N. A. (2011). *Human Factors and Ergonomics in Consumer Product Design: Methods and Techniques (Handbook of Human Factors in Consumer Product Design): Needs Analysis: Or, How Do You Capture, Represent, and Validate User Requirements in a Formal Manner/Notation before Design* (Chapter 26 by K Tara Smith), CRC Press.

Larry, P. (2009). *Information Quality Applied*, Wiley Publishing, Indianapolis.

Moreno, L., & Martinez, P. (2013). Overlapping factors in search engine optimization and web accessibility. *Online Information Review*, 37(4), 564-580.

Mansour, A. & Mohammed, A. A. (2009). *Information systems analysis using a computer*, 5<sup>th</sup> edition, Criterion Publishing and Distribution, Amman, Jordan.

Miller (2010), Holmes (Spring). *Information Systems Management* 13 (2):79–82.

Gorla, N., Somers, T. M., & Wong, B. (2010). Organizational impact of system quality, information quality, and service quality. *The Journal of Strategic Information Systems*, 19(3), 207-228.

Parsian, N., & AM, T. D. (2009). Developing and validating a questionnaire to measure spirituality: a psychometric process. *Global Journal of Health Science*, 1(1), 2.

Gagliardi, N. (2014). US universities at greater risk for security breaches than retain and healthcare: BitSight. ZDNet, August 21, 2014.

Rai, A., Lang, S. S., & Welker, R. B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. *Information Systems Research*, 13(1), 50-69.

Robert, S. (2013). Global Initiative for Inclusive Information and Communication Technologies. G3ict Publications, UK.

Seddon, P., & Yip, S. K. (1992). An empirical evaluation of user information satisfaction (UIS) measures for use with general. Journal of Information Systems, 6(1), 75-92.

Wu, J. H., & Wang, Y. M. (2006). Measuring KMS success: A respecification of the DeLone and McLean's model. Information & Management, 43(6), 728-739.

UNIVERSITI SAINS ISLAM MALAYSIA  
جامعة العلوم الإسلامية  
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

## CONCEPT, PRACTICALITY AND REALITY IN MALAYSIA GOVERNMENT PROJECT PAYMENT CERTIFICATION OF FINAL ACCOUNT

Waluyob, S. N.,<sup>1</sup> Darus, Z. M.<sup>2</sup> and Ismail, M. N.<sup>3</sup>

Faculty of Science and Technology (FST), University Sains Islam Malaysia  
Nilai, Negeri Sembilan<sup>1,2</sup>

faithayu@yahoo.com<sup>1</sup>, zmdarch@gmail.com<sup>2</sup>, noorizhar@usim.edu.my<sup>3</sup>

### ABSTRACT

*By principle, the Malaysian Government Procurement Policies undeniably set the supreme objective as to support whole Government programs especially on public project developments aiming the benefits for all. Lack of understanding despites of expertise area, indefinite in respect of responsibilities between parties involves and poor project team management has resulted from numbers of project failures especially when it comes to project certification of payment. Hence, this study is attempted to identify the current payment procedure practiced, understand the challenges and the remedies in payment certifications on the project completion i.e. final account. This study adopts a qualitative case study approach of school projects based on reported final account issuance of 51 projects under Southern Region in RMK-8 to RMK-10 up to the year 2016 task force under Ministry of Education (MOE) which cost approximately RM487.7million where the total 412 projects cost about RM8.7billion. The data were collected and analyzed using case study qualitative approach through detailed, in-depth data collection involving multiple sources of information that involved observations, interviews, documents reviews and reports. Findings show that challenges were not only related to project management and roll-out strategies, but also to procedures, communication failures, confusion in understanding and compliance with the contract. Outcomes show the importance of forming an effective framework was essential as to coordinate the whole cycle of ministry's development project. The theoretical effective framework draws and was further developed during data analysis. The finding of a study particularly later will be compared with both current and proposed framework for payment certification on final account issuance. Therefore, it is expected from this findings would be benefitted the ministry in ensuring the late payment or failure of final account closing account can be minimized and applied for future ministry's project development.*

**Keywords:** HPM, Fractional operator, Fredholm-Volterra integral equation, Discretization.

## 1. INTRODUCTION

Procurement decisions have a reflective effect on the balance of risk and reward on projects, and the role of each part in involves are count. With the expectation of some government and local government bodies which have implemented, the whole process of procurements and certification for payments are found rather somehow way behind the actual framework set by the Government. At present, similar to another type of procurements, Government procurements continue to face countless of problems (thou controlled) that threaten to hinder its developments and might risk the fundamental and principles if not addressed and managed effectively. A number of influence factors, particularly related to productivity, quality, time and cost have post-challenge to the development in this industry.

Payment delay is defined as failure of a paymaster to pay within the period of honoring of Certificates as provided in the contract (Harris & McCaffer, 2013). The parties involved in the process of a payment claim such as client, contractor, superintending officer, architect, quantity surveyor, banker, and other construction players may cause a payment to be delayed. Payment under Enactment of the Construction Industry Payment & Adjudication Act (CIPAA) is described as a payment for work done or Services rendered under the express terms of a construction contract. Government procurement discipline has received great attention from the eyes of the public interest. Most of the literature review concentrates on causes, effects of payment delay; also it talks about methods to avoid payment delay effects and payment delay on construction in general.

Understand that, the implementations of public projects not only aim to achieve production output, outcome and impact of the project even seen as disciplines which should be practiced by all project managers. In the management of the implementation of public projects, especially in the construction phase, elements of compliance with the specifications must be given full attention (Implementation Coordination Unit, 2015). Besides, the establishment of Implementation Coordination Unit (ICU) itself was as a support system for efficient, effective and resources saving via extensive application of the principles and practices of good project management, planning and implementations of public policy resolutions. This was supported by the Procurement Guidelines Book, The Financial Procedures Act 1957, The Treasury

Instructions (TI) and the Treasury Circular Letter (TCL) which unconditionally by right understood, applied and exercised by all government legal frameworks across federal, state government and semi-government agencies.

However, these legal frameworks are insufficient in ensuring the accountability and transparency of the project procurement. Therefore, by having effective and good project governance it could assist in building trust and cooperation between top management, project sponsor and project team members. Efficiency in project procurement is an important issue in cases where public sector procurement accounts for a large portion of economic activity. By ensuring transparency in the procurement procedure especially in an essential determinant of efficiency, as it enhances the competitiveness of public sector plan procurement. By improving the transparency of project procurement planning this could divert public sectors expenditure away from activities that could provide an opportunity for illegal activities. This is important as the Malaysian government needs to take the appropriate action to overcome numerous loopholes in the public project procurement activities to ensure a more transparent and efficient system.

As far as government perspective is concerned, this underachieving performance has led to many problems including public complaints, loss of reputation financially and physically, safety issues and public interest. Payment issues in the construction industry are considered a factor of significant concern to all the players in the industry eventually affect the whole chain of the project development itself from the government (the issuance) to the end users (students, school teachers, and administrations team). Yet, the elements of bureaucracy otherwise red tapes were shackled and dragged the governing system of project procurements down to the alley of delay in decision-making practices. Unnecessary intrusion by numbers of parties involves direct and indirectly to the project procurement is said to be seriously undesirable impact specifically in developments and infrastructures particularly school projects as a nucleus.

The main factors for late and non-payment in the construction industry identified from the study include: delay in certification, paymaster's poor financial management, local culture/attitude, paymaster's failure to implement good governance in business, underpayment

of certified amounts by the paymaster and the use of 'pay when paid' clauses in contracts (Danuri et al., 2006). As far as Malaysia construction development or other projects are concerned, delay in final account issuance which is exceedingly synonym has been the most intolerable issue. Agreeable with (Zakaria et al., 2012) indicated that its impacts were so significant that it tends to decelerate the implementation of Malaysia Plans. Thus, by reforming and thus executing an established framework, these matters in questions can be possibly minimized and accountability exercises are unmistakably well defined.

For the purpose of this study, selections are made and represent on behalf of numbers of development procurement projects under 8th Malaysian Plan towards 10th Malaysian Plan for Ministry of Education (MOE) which adopts a qualitative case study approach of school projects based on reported final account issuance of 51 projects under Southern Region cost approximately RM487.7million where the total 412 projects cost about RM8.7billion. The justification for choosing MOE as the ground study is because its dominos impact directly towards its stakeholders, public interest and the exclusive end users i.e. the students. The scope of the study will cover the development projects specifically the school building project under 8th Malaysia Plan (RMK-8) towards 10th Malaysia Plan (RMK-10) under Ministry of Education (MOE).

## 2. METHODOLOGY

This study is needed to evaluate the level of understanding the cause and effect of 51 projects specifically under South Region based on pending final account cost approximately RM487.7million. Interview sessions were conducted to the management team including project managers, supervisors, site engineers and others. The data analysis gathered from the official preliminary reports from court cases which picked randomly, Auditor General Reports and other reliable sources. The analysis was constructed as to distinguish the reality of the concept of project procurement management and its practicality. Examination of the effect and impacts due to formation and transformation on the provision of a settlement of government procurement and established the recommend strategies and policies to suit government payment procedures to ensure the whole process are significant, relevant and efficient align with standard government view and practice.

### 3. LITERATURE REVIEW

Procurement decisions have a profound effect on the balance of risk and reward on projects, and the role of each part in involves are count. With the expectation of some government and local government bodies which have implemented, the whole process of procurements and certification for payments are found rather somehow way behind the actual framework set by the Government. In Malaysia, the importance of plan procurement is supported and encouraged by project management experts and government departments which are Ministry of Finance, Public Work Department, Economic Planning Unit and Treasury Board. This is because the procurement contract can lead to dissatisfaction, ethical issues, and time-consuming detours when it is not properly done.

Previous studies mentioned above were generally focused on finding delay causes. Some of these studies identified very lacking factors or ignored some important categories. This may be misleading or may result in wrong analysis. In this paper, through the comprehensive literature review and interviews with different highly experienced top management of the Ministry, the execution teams, construction professionals, the preliminary reports, findings and current practices, the study will attempt to use the comprehensive resources for the qualification of relative importance indices of a comprehensive list of delay factors and effects in overall performance.

Project governance is a subset of corporate governance focusing on the areas of corporate governance related to project activities, including portfolio direction, project sponsorship, project and program management and efficiency and disclosure and reporting (Garland, 2009). The important focus of project governance is to ensure that the project objectives are aligned with the organization portfolio and objective. Therefore, there are three main goals of project governance: choose the right project, deliver the chosen project efficiently and to ensure projects that have been chosen can be sustained.

Meyer & Rowan (1977) formulated a new trend in project management: "Project governance has only recently become an issue of importance community and literature. Over the last ten years, there has been more interest in the governance of projects in general and the governance

of large complex public projects in particular". Project Governance is the framework around selection, prioritization and project oversight for continued adherence to organization objectives. When the public sector applied this concept to the implementation of their projects, it actually could assist in changing project's scope base on project oversight. There must have a clear framework to differentiate between the organizational governance and project governance so that there will be no interference among them in making decision.

However, in Malaysia, there is a lack of methods for governing the project procurement as there are layers of bureaucracy and red tape involved in it that caused delay and problems in the decision-making process. When there is too much interference from another party it could cause a problem to the decision makers for that particular project. This happened between Works Ministry and Public Works Department as there is interference in handling the projects. This interference has caused delay to project implementation and decision-making process. By implementing a project governance framework, these problems can be hindered as the line of accountability is clearly defined.

The research focuses on public sector projects (Aliza et al., 2011), therefore it is very important to understand the project environment in public sector projects. According to Patrick (1995) transparency and accountability are important in the public sector because the sector has a profound and pervasive effect on the lives of citizens and on the activities of the private sector. Therefore, transparency and debate to improve the analysis need to be implemented. Unaccountable decision making could increase the danger of corruption. This includes the diversion of public resources, a risk of the costly project and project rejection is the whole of government subject to media attention which can be politically harmful.

#### **4. OVERVIEW CONCEPT AND PRACTICALITY**

In principle, a final account in construction contracts is the agreed statement of the amount of money to be paid at the end of a project contract by the employer to the contractor. A final account brings about a sense of finality to the negotiations leading up to the agreement of the final account between the parties to the contract. In construction contracts, it is generally a common practice for either the employer (or the employer's representative) and the contractor

to sign the final account statement to signify that the final account figure represents the full and final settlement of all claims. The closing of final account negotiations between the contractor and the architect, engineer or quantity surveyor on behalf of the employer will in due course trigger the issue of the final account statement and ultimately, enable the issuance of the final certificate. The major factor leading to a project's success is the practice of efficient and timely payment in construction projects. The cause of the problem to closing the final account is because of dispute and delay. The dispute occurs in project management is inevitable, yet dispute is manageable.

Generally practice, construction final account is tackled at the usually less pressing and less resourced post-completion stage and has been often neglected or even treated by a simpler sense that the settlement of final account is just a matter of time or decision. Stated in Standard Form of Contract to be used where bills of Quantities Form Part of Contract P.W.D Form 203A; that as soon as is practicable but later than 3 (three) months after the issuance of the Certification of Practical Completion (CPC), the contractor shall submit full particulars complete with receipts, vouchers records that would substantiate the contractor's claim under Clause 44 together with any documents, supporting vouchers and any explanations and calculations including documents relating to the accounts of Nominated Sub-Contractors or Nominated Suppliers, which may be necessary to enable the Final Account to be prepared by the S.O. provided always the contractor has given the notice of claim in writing the stipulated time or times in the said provisions (Cawangan Kontrak dan Ukur Bahan, 2010).

As far as said clause is concerned, the construction players are obliged to settle the final account and issue a final payment certificate to the contractor within a specified period of time after the receipt of the contractor's final account statement. Even though it is clearly stated on standard forms of contract, delays in final account closing remain occurred. Although the backlogs in the closing of final account have existed for decades where the important factors significantly affecting the closing of the final account of construction projects have been identified since then, namely contractual-related, contractor related and management-related, these factors usually attract less attention and priority by project participants and contracting parties. Undue protracted inaction makes the closing of final account more difficult and frequently leads to the emergence of unnecessary dispute (Kwok, 2009).

Payment problems are old age issues that permeate the Malaysian construction industry. So often, contractors and parties in the construction industry complain either not getting paid or payments have been unduly delayed by the employer (Azmana et al., 2014). Cases are most likely addressed under Government construction projects especially highly impacts, high profile and involved billions of ringgits. According to the records provided by CIDB (2012), the total number of projects until March 2012 is 655 with the value of RM9,399.09 million. The project was awarded to local contractors is 652 with a total value RM9,109.53 million and foreign contractors are three with a total value is RM289.56 million. The number of private projects generally exceeds the number of government projects, regardless the status of the contractor. A failure of the contractor in getting regular and timely payment could result in project delay, reduced profitability and in the extreme case the company may go into liquidation. It will also have a knock-on effect on the entire construction value chain because when clients do not pay the main contractors on time, the sub-contractors, suppliers, hirers and everyone in the construction value chain will suffer (Azmana et al., 2014).

There are a lot of problems faced by Malaysia construction industry, which will effect to the productivity of construction industry in terms of manpower, quality of productivity, time and payment issue especially during the closing of final accounts in construction projects.

## **5. REMEDIES FOR DELAY IN FINAL ACCOUNT PAYMENT**

Research by Zakaria et al., (2012) stated that client or paymasters should practice an efficient system to make sure that the contractor receives the payment accordingly. Hence, delays of final accounts closing can be resolved at the early stage if the documents are kept properly. Delay in the closing of final accounts may also cause problems to the contractor in making working capital and eventually lead to bankruptcy. In additional, Zakaria et al., (2012) stated that the rapid increase of the delay in the closing of final account nowadays has gone to the point that it tends to decrease the good image of the construction industry in Malaysia. For that reason, personal in charged in the project should be maintained especially the project manager until then the closing of final accounts is completed. Thus, the project managers need to have a better understanding of critical success and failure factors in the construction project and how to measure them.

Understood consultant's responsibility is started once receiving the Letter of Acceptance (LOA) until the Final Certificate contractor is finalized. Consultants should refer to the Term of Reference (TOR) in Contract Sum Adjustment (CSA) to apply any additional provision, the extension of time, etc. if there is a requirement that the duration of the appointment. Whereas the responsibility of the contractor starting from the receipt of the Letter of Acceptance to the Final Certificate is been finalized which covers the period of construction, Testing and Commissioning (T&C), Defect Liability Period (DLP) and eventually signed the Final Account Certification. Conversely, based on interview session with Technical Team, the major cause of delay in an assurance of final account for listed 51 projects under Southern Region is insufficient documents submitted to Technical Unit for final payment payments process. Reminder letters have to be sending out and more discussions need to be taken to a place in order to make sure payment made accordingly. Understanding, cooperation, and commitment from various parties involve triggered a massive traffic where challenges concerning the understanding of each other's the role.

This issue eyed and highlighted by Chief Secretary where through Ministry of Finance (MOF) directed that the total of 412 projects affected by pending for the final account must be solved entirely within three years by MOE which scheduled ended in December 2017. Thus, as to tackle this issue collectively, the task force lead by the Secretary of Education Department was formed in 2015 with a prime objective to identified issues and problems occur in implementation as well as project managements in Education Development Department, MOE. Therefore, the determination of the core processes in the implementation of the project (deliverables) until the project payment process and the final account is made and officially closed between the manager, consultant, and contractor completed and reported periodically to the MOF.

Understand that task force regime or known as a quick-effective-action commonly taken into consideration action is like a treatment for healing the sickness that has been identified which a common last resort. Supposedly, distinguishable strong plan procurement should plug into the project framework. Plan Procurements (formerly known as procurement planning and plan purchases and acquisition) is one of the first tasks that need to be performed on a project (Aliza

et al., 2011). This is the stage to decide which goods and services will be done internally and which goods and services will be carried out by suppliers and contractors. Planning is the foundation for all management fields including project procurement. It is a function that forms the foundation for the rest of the management functions. When planning is properly conceived and implemented, it can serve as an important mechanism for extracting, distributing and allocating resources (James, 2004). Plan procurement calls for the early involvement of the purchasing office so that options and alternatives can be explored with the requesting user. Issues such as purchase estimates, product specifications, make or buy decisions, and outsourcing opportunities may be very well being on the agenda (Matthews, 2005).

In Malaysia, the importance of plan procurement is supported and encouraged by project management experts and government departments which are Ministry of Finance, Public Work Department, Economic Planning Unit and Treasury Board. This is because the procurement contract can lead to dissatisfaction, ethical issues, and time-consuming detours when it is not properly done. At the end of the day, projects and procurements should be well planned for reasons other than exposure or reputational risks; they should be well planned to ensure the establishment of compliant and effective contract mechanisms that will ensure the provision of qualified, capable and professional results to departments (Deme, 2009). Mullins (2013) has argued that procurement planning is a process of determining the procurement needs of an entity and the timing of their acquisition and their funding such that the entities operations are met as required in an efficient way. However, if this stage is not properly governed, therefore it is difficult to meet all the requirements and objective of project procurement. Thus, this could cause high risk to the project completion.

An outcome of the research piloted by (Riazi et al., (2013) has developed a conceptual guideline for the utilization of Supply Chain Management (SCM) tools and techniques towards both mitigating and reducing delays in Malaysian public sector projects. Based on the research findings, a conceptual SCM framework which aims at ratifying the problem is proposed. It is anticipated that the SCM framework can be used as a preliminary guideline towards understanding and strategizing solutions to address the problems of the industry. The important focus of project governance is to ensure that the project objectives are aligned with the organization portfolio and objective. Therefore, there are three main goals of project

governance: choose the right project, deliver the chosen project efficiently and to ensure projects that have been chosen can be sustained.

## 6. CONCLUSION

Final account closing success in construction projects is defined as a certified statement of final cost which is effectively closed through systematic project administration procedures, where an on-time and within budget construction is a major criterion of final closing and management success as a whole (Syuhaida, 2014). Fundamentally, an ideal project procurement management should be clearly understood by all parties involve and translated into a proper framework. The structured framework consists of continuation each and every step binding from one process of project procurement management to another towards the closing of final account issuance should be clearly recorded. Terms, rules and regulation as stated in a contract agreed ought to be able to accommodate transparency, competency and accomplishment of the project procurement particularly for education physical development that carries great sensitivity.

Government procurement has always attracted public attention. Maybe because of media coverage, particularly about the Auditor General (AG) Report which normally highlights cases of irregularities and mistakes in government procurement which might affect public perception about the lack of efficiency, integrity and transparency as well as the presence of corruption in the procurement process or systems used for awarding major Government projects (National Audit Department, 2012). Transparency in the procurement procedure especially in an essential determinant of efficiency, as it enhances the competitiveness of public sector plan procurement cost a high impact to overall project procurement management. By improving the transparency of project procurement planning this could divert public sectors expenditure away from activities that could provide an opportunity for illegal activities. This is important as the Malaysian government needs to take the appropriate action to overcome numerous loopholes in the public project procurement activities to ensure a more transparent and efficient system (Othman et al., 2010).

Therefore, existing deficiency or fault in government project payment certification of final account should be taken as a lesson learned by related parties involved. Management team i.e. the government, contractors, consultants, superintending officer, architect and quantity surveyor must play their effective role and act cooperatively to ensure that the future project development will be accomplished within target timeframe. Thus, in order to express a sense of obligation and accountability, the issue raised in this paper should be duly addressed.

## REFERENCES

- Azmana, M. N. A., Dzulkalnineb, N., Hamidb, Z. A., & Bingc, K. W. (2014). Payment Issue in Malaysian Construction Industry: Contractors' Perspective. *Jurnal Teknologi*, 70(1), 57-63.
- Aliza, A. H., Stephen, K., & Bambang, T. (2011). The importance of project governance framework in project procurement planning. *Procedia Engineering*, 14, 1929-1937.
- Cawangan Kontrak dan Ukur Bahan, J. K. (2010). *Buku Pentadbiran Kontrak Kerja Raya Edisi Ketiga* (3rd ed.). Kuala Lumpur: Jabatan Kerja Raya (JKR).
- Danuri, M. M., Munaaim, M. C., Rahman, H. A., & Hanid, M. (2006, November). Late and non-payment issues in the Malaysian Construction Industry-Contractor's perspective. In *International Conference on Construction, Culture, Innovation and Management (CCIM)*.
- CIDB, C. I. M. P. (2012). *Market Watch 2010; Construction Industry*. Kuala Lumpur, Malaysia.
- Deme, A. (2009). Why is it important to undertake good planning before undertaking a procurement process? *FMI Journal*, 21(1), 11-15.
- Garland, R. (2009). *Project Governance: A practical guide to effective project decision making*. Kogan Page Publishers.
- Harris, F., & McCaffer, R. (2013). *Modern Construction Management*. John Wiley & Sons.
- Implementation Coordination Unit. (2015). *Managing the Expectations of the Public Project Management Concepts, Practical and Reality*. Putrajaya, Malaysia, W.P Putrajaya, Malaysia: Implementation Coordination Unit (ICU, JPM).
- James, U. V. (2004). *Public Policy and the African Environment: An Examination of the Theory and practice of the planning process of the continent*. *Agenda Setting and Public Policy in Africa*, 23.
- Kwok, C. K. (2009). *Study of important factors affecting final account settlement satisfaction of Hong Kong civil engineering projects: contractor's perspective* (Doctoral dissertation, City University of Hong Kong).

Matthews, D. (2005). Strategic procurement in the public sector: A mask for financial and administrative policy. *Journal of Public Procurement*, 5(3), 388.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340-363.

Mullins, D. R. (2003). *Accountability and coordination in a decentralized context: institutional, fiscal and governance issues*. Washington, DC: American University.

National Audit Department. (2012, October 10). *Excellence in Government Procurement: Lessons Learnt. "Understanding the Integrity Pact, the Corporate Integrity Pledge and the Do's and Don't's of Government Procurement"*, 1-14. Hotel Istana, Kuala Lumpur, W.P Kuala Lumpur, Malaysia.

Othman, R., Zakaria, H., Nordin, N., Shahidan, Z., & Jusoff, K. (2010). The Malaysian Public Procurement's Prevalent System and its weaknesses. *American Journal of Economics and Business Administration*, 2(1), 6.

Patrick, P. (1995). *Managing Trust: Transparency, Accountability and Ethics in Malaysia*. Kuala Lumpur, Malaysia: ISIS Malaysia : Goethe-Institut.

Riazi, M., Riazi, S., & Lamari, F. (2013, May). Public sector project delay: the Malaysian perspective and the way forward. In *Proceedings of the 19th CIB World Building Congress, Brisbane 2013: Construction and Society*. Queensland University of Technology.

Sakamon, A. B. 2006. *Manual Pakej Penilaian Tender. Cawangan Kontrak dan Ukur Bahan, Ibu Pejabat JKR Malaysia*.

Syuhaida, I. (2014). *Construction-Final Account Closing In Project Management Perspective*. 16-18.

Zakaria, Z., Ismail, S., & Yusof, A. M. (2012). Cause and impact of dispute and delay the closing of final account in Malaysia construction industry. *Journal of Southeast Asian Research*, 2012, 1.