

Selected Computing Research Papers

Volume 7

June 2018

Dr. S. Kendal (editor)

**Published by
the
University of Sunderland**

The publisher endeavours to ensure that all its materials are free from bias or discrimination on grounds of religious or political belief, gender, race or physical ability.

This material is copyright of the University of Sunderland and infringement of copyright laws will result in legal proceedings.

© University of Sunderland

Authors of papers enclosed here are required to acknowledge all copyright material but if any have been inadvertently overlooked, the University of Sunderland Press will be pleased to make the necessary arrangements at the first opportunity.

Edited, typeset and printed by
Dr. S Kendal
University of Sunderland
David Goldman Informatics Centre
St Peters Campus
Sunderland
SR6 0DD

Tel: +44 191 515 2756
Fax: +44 191 515 2781

Contents

Page

Critical Evaluation of Arabic Sentimental Analysis and Their Accuracy on Microblogs (Maha Al-Sakran).....	1
Evaluating Current Research on Psychometric Factors Affecting Teachers in ICT Integration (Daniel Otieno Aoko).....	9
A Critical Analysis of Current Measures for Preventing Use of Fraudulent Resources in Cloud Computing (Grant Bulman)	15
An Analytical Assessment of Modern Human Robot Interaction Systems (Dominic Button)	23
Critical Evaluation of Current Power Management Methods Used in Mobile Devices (One Lekula)	31
A Critical Evaluation of Current Face Recognition Systems Research Aimed at Improving Accuracy for Class Attendance (Gladys B. Mogotsi).....	39
Usability of E-commerce Website Based on Perceived Homepage Visual Aesthetics (Mercy Ochiel).....	47
An Overview Investigation of Reducing the Impact of DDOS Attacks on Cloud Computing within Organisations (Jabed Rahman).....	57
Critical Analysis of Online Verification Techniques in Internet Banking Transactions (Fredrick Tshane)	65

Critical Evaluation of Arabic Sentimental Analysis and Their Accuracy on Microblogs

Maha Al-Sakran

Abstract

This research paper will focus on Arabic sentimental analysis and the different methodologies used in experiments, to determine accuracy levels as well as improve the accuracy of translating social media posts through analysis and comparison of results conducted on datasets using a variety of translation tools. Translation system's tools such as classifiers and word stemming will be compared to categorize emotions and opinions in terms of positivity or negativity for both Modern Standard Arabic and Colloquial Arabic.

1 Introduction

Sentimental analysis are used widely in social media to understand opinions and emotions of user's posts and reviews, the most morphological rich Semitic language is Arabic as it's also one of the most popular languages on twitter.

In the last few years not much research has focused on sentimental analysis in Dialectal Arabic as majority of research has focused on Modern standard Arabic. Most of the work focused on analysing English texts, as such expansion of sentimental analysis needs to reach other languages including Arabic.

As stated by Santosh et. al. (2016) "Arabic text contains diacritics, representing most vowels, which affect the phonetic representation and give different meaning to the same lexical form." This can make sentimental analysis challenging, in addition to other complications such as the unavailability of words with capital letters.

Related research focused on classifying tweets by applying subjectivity, there has also been focus on classifying posts into categories such as news, events, opinions and deals. Different approaches have been applied from extracting features from texts and metadata to the use of generative models.

Recent research has mainly concentrated on the retrieval of opinionated posts against specific subjects in terms of relevance using machine

learning. Many translation systems rely on word stemming, datasets, and pre-processing, they also rely on sentiment analysis tools (Walid et. al. 2016).

Related research includes testing of classifiers by providing individual judgements through the use of expert and volunteer labellers in order to evaluate the data as well as using it as a good source for future training (Amal 2016).

As stated by Alok et. al. (2013) that they have "achieved relatively high precision, recall still requires improvement" this is in regards to sentimental analysis of micro-blogs in twitter.

This research paper will analyse experiments of which have been conducted on sentiment analysis on Arabic in social media. This research will evaluate and compare the result for the experiments in order to reach good scientific conclusions on classifiers such as SVM, Naives Bayes and N-gram training models.

This research paper will also evaluate lexicon, negation and emoticons as they require considerations in the sentiment analysis. There are challenges for example when translating English to Arabic there is a lack of resources due to the morphological and complex language being used.

Emoticons play a part as they can cause confusion, for example some sentences may seem negative but their definition could be

positive. In addition, writing from right to left making the position of the brackets () in the opposite direction in emoticons can also have an impact (Ahmad et. al. 2013).

2 Sentiment Analysis on Arabic Tweets Using Classifiers and Datasets

Two types of classifiers, SVM and Naïve Bayes were used as a preparation for the preprocessing to experiment with different stemming methods (Talaat et. al. 2015).

The methods that they have proposed were Dataset1, Dataset2, and Dataset3 along with CNB, MNB and SVM classifiers in order to identify which combination is best. Although, which configuration worked better is still not identified (Talaat et. al. 2015).

Unigrams, stemming, bigrams, filterations, word counts and IDF were used in the experiments and a percentage of the accuracy of which they have achieved for each of the datasets (D1, D2 and D3).

Bag of words model was used through datasets to detect the accuracy of the informal texts in tweets. This was tested by using three different types to identify the most suitable combination, these were N-gram training models, text pre-processing, machine learning algorithms and classifiers (Talaat et. al. 2015).

Bag of words model was tested for text classification, each of the terms is scored as ones or zeros by the vector and based on this, accuracy is determined through CNB, MNB and SVM (Talaat et. al. 2015).

Dataset	# of Tweets		
	<i>Total</i>	<i>Train</i>	<i>Test</i>
DS1	3436	2750	686
DS2	5493	4405	1088
DS3	11070	9656	1414

Table 1 Datasets Distribution (Talaat et. al. 2015)

Three different types of datasets were tested in the experiment (Talaat et. al. 2015). In the first dataset, 6000 Egyptian tweets were gathered, annotated and then categorized. The results were 2750 words for training and 686 words for testing.

In the second dataset some of the tweets were not found and this resulted in 724 positive, 1565 negative and 3204 neutral.

The third dataset was of educational terms, 1414 tweets were used for testing and the rest, 9656 tweets were used for training.

Dataset Classifier	DS1			DS2			DS3		
	Preprocessing	10 CV (%)	Test (%)	Preprocessing	10 CV (%)	Test (%)	Preprocessing	10 CV (%)	Test (%)
CNB Baseline	IDF + Unigrams + Bigrams	68.37	56.08	IDF + Unigrams + Bigrams	60.61	50.78	IDF + Unigrams + Bigrams	77.34	67.33
CNB	Baseline + El-Beltagy Stemmer	70.84	57.10	Baseline + Filtration + El-Beltagy Stemmer	65.67	51.89	Baseline + Filtration + Light Stemming + IG	76.55↓	67.90
	Base line + Light Stemming	70.62	55.49	Baseline + El-Beltagy Stemmer	65.56	51.15	Baseline + El-Beltagy Stemmer + IG	76.32↓	67.75
MNB Baseline	IDF + Unigrams + Bigrams	67.35	56.37	IDF + Unigrams + Bigrams	55.93	49.12	IDF + Unigrams + Bigrams	76.08	67.19
MNB	Baseline +El-Beltagy Stemmer	69.20	57.25	Baseline + Filtration + El-Beltagy Stemmer + IG	66.70	56.22	Baseline + Filtration + El-Beltagy Stemmer	76.95	69.81
	Baseline + Light Stemming	69.28	56.81	Baseline + Light Stemming +IG	66.31	55.48	Baseline + Filtration + Light Stemming	76.38	69.81
Baseline SVM	TF + Unigrams + Bigrams	57.41	52.71	TF + Unigrams + Bigrams	53.18	50.14	TF + Unigrams + Bigrams	66.15	63.57
SVM	Baseline + Light Stemming + IG	60.43	54.61	Baseline + Filtration + Light Stemming +IG	63.02	59.45	Baseline + Filtration + Light Stemming + IG	72.20	68.32
	Baseline + Light Stemming + Bigrams	61.41	54.03	Baseline + Light Stemming + IG	62.77	59.45	Baseline + El-Beltagy Stemmer + IG	72.47	68.15

Table 2 Datasets Classifiers (Duwairi et. al. 2014)

The results show unigrams and bigrams worked best together in terms of accuracy. CNB performed much better in comparison with MNB and SVM. Where-as SVM with word counts produced increased accuracy. And Naïve Bayes had the best accuracy with IDF. Accuracy was not affected when simple text cleaning and filtration were applied.

Different datasets were used as well as three machine algorithms which allowed variety. The data collections included 6000 tweets which is a large quantity. Appropriate terms such as educational terms were used in the tests.

Each of the three machine algorithms were selected based on previous experiments which prove that these machines outperformed other classifiers. This research has no bias as the accuracy's score is based on scientific calculations. Therefore, making the research quantitative and the tests valid.

Tweets were filtered based on number of characters and other specific criteria (Duwairi et. al. 2014). The tweets were then reduced in size through pre-processing by using

(<http://rapidminer.com>) tokens were then separated by adding commas and spaces

in order for normalization to take place (Duwairi et. al. 2014).

A dictionary was created which converted Jordanian dialect to MSA to help with the translation process in addition to two different dictionaries, Negation dictionary and Arabism dictionary (Duwairi et. al. 2014).

Different types of classifiers were used through the experiment such as NB, SVM and KNN. Each of the settings gave results of accuracy and whether stop-word filters, stemming or folds were used as well gave a score of each (Duwairi et. al. 2014).

Classifier	Stopwords filter	Stemming	Folds	Accuracy
KNN	1	1	5	55.72%
KNN	1	0	5	56.23%
KNN	0	0	5	59.99%
KNN	0	0	10	51.58%

Table 3 KNN Classifiers (Nourah et. al. 2016)

When comparing the results of NB, SVM and KNN. NB had 76% of accuracy when stop-word filters and stemming were excluded and this is by now the highest accuracy in this specific experiment. An expansion of dictionaries is needed and more advanced classifiers. There was clearly a memory problem in the Rapidminer but this will be considered for future research.

Three different types of dictionaries were used which allowed variety. Crowd-sourcing was also used in order to annotate the tweets in addition to a login option specific for the author and the user. This is a good option as over 25000 tweets were labeled, it would have not been possible for authors alone to annotate.

The data collection is large. Deliberate bias may not have been applied but using a third party website such as Rapidminer in the process has an impact on validity as this website was not tested prior to the experiment or compared with another website that does similar functions. This research has applied quantitative data through providing a score to achieve the results.

The experiment to pre-process social media posts through normalisation for the purpose of consistency through applying stemming and stop-words removals to reduce term space (Nourah et. al. 2016). For example a word can have various meanings but is still spelt the same (Nourah et. al. 2016).

	Unigram	Unigram +Bigram	Unigram +Bigram+ Trigram
Accuracy	89.553 %	88.0991 %	87.399 %
Precision	0.817	0.791	0.78
Recall	0.966	0.97	0.972
F-Measure	0.885	0.872	0.865

Table 4 Results using Normalised Tweets (Nourah et. al. 2016)

Precise, recall and F1 measure were used in the experiment for the purpose of evaluating accuracy to ensure that each dataset will be in the training and testing set. The experiment was conducted by using raw Tweets, the normalization was then applied followed by the stemmer and then the stop-words.

The results were more accurate using SVM and Naïve Bayes combined without stemmer before the pre-processing phase, as it achieved 89.553% accuracy (Nourah et. al. 2016).

The words used in Arabic had more than one meaning but in the research paper only one meaning was presented. For example the word “من” was translated as “of” which is correct but it could also mean “from” or “who”.

3500 tweets is a large amount that has been used making the experiment more wide and valuable in terms of data collection. The research was conducted on Modern Standard Arabic and dialectal Arabic but there has been no details of which dialect has been used in the experiment as there are currently 22 Arabic dialects. User details and emoticons were removed. For this specific sentiment analysis, emoticons can have an impact on the results in regards to positivity and negativity. User details remained confidential. The research has no bias as the accuracy results were calculated mathematically using quantitative data. There is no mention of incomplete data. Therefore, no negative impact on the accuracy of the results.

2.1 Comparison of Sentiment Analysis Lexicons Using SVM and NB Classifiers to Determine Accuracy

Accuracy is higher in the research that used NB, this is an achievement as there was still problems with the memory compared with the results of the datasets (Nourah et. al. 2016). The highest

accuracy of 89% was achieved when SVM and Naïve Bayes were combined.

The experiment was proposed by Ahmad et. al. (2013) a dictionary was created to convert Tweets that were collected from twitter's API through "lang:ar" query. A collection of 2300 tweets were sent to two annotators to ensure agreements on each of the tweets. Naïve Bayesians and SVM classifiers were used in the experiment to differentiate between subjective and objective through precision, recall and f-measure (Ahmad et. al. 2013).

Arabsenti lexicon was used but there were errors, which require expansion to reduce them, but surprisingly it had minimal impact on classification.

As stated by Ahmad et. al. (2013) "The expanded lexicon had much broader coverage than the original lexicon". Ahmad et. al. (2013) also claimed that expansion has improved the sentiment classification. Improvement of classification in terms of subjectivity and sentiments for Arabic tweets was achieved through the expanded lexicon rather than the original lexicon.

Experiment is considered reliable according to the method of which the tweets have been annotated, as annotators had to agree or disagree on the tweets through giving reasoned arguments and then come to an agreement, this process prevents bias. As this research is considered to be qualitative, adding more annotators can be expensive. Annotators had the expertise when choosing which tweets were positive and which were negative, adding the amount of data of which has been used. The use of at least five annotators to prevent bias further is recommended.

Leila et. al. (2016) proposed a technique for deeply mining annotated Arabic reviews. This research demonstrates the extracted features of user reviews.

	Review	Extracted pairs
Rule #1	اشترت موبائل الشاشة جيدة	{ جيدة ,شاشة }
Rule #2	اشترت موبائل جديد وفيه امكانيات كثير جيد	{ جيد ,موبائل } { جديد ,موبائل }
Rule #3	شاشة وبطارية و عمر افتراضي ممتاز	{ ممتاز ,شاشة } , { ممتاز ,بطارية } , { ممتاز ,عمر افتراضي }
Rule #4	اشترت موبائل فيه شاشة البطارية جيدة وخاصة اللبس ممتازة جدا لكن عالي	{ ممتاز ,شاشة } , { جيدة ,بطارية } , { ممتازة ,خاصية اللمس } , { عالي ,موبائل }
Rule #5	شاشة حلو و سيني	{ ممتاز ,شاشة }

Table 5 Rules Reviews (Leila et. al. 2016)

200 Arabic reviews were gathered from Facebook, forums, YouTube and Google, rule types were applied in the classification which were then extracted in pairs.

ATKS tool was used to convert colloquial Arabic to MSA. Reviews were first annotated, processed for ATKs and POS Tagging (Leila et. al. 2016). Accuracy was affected in sentiment extraction and gave a percentage of 82.

In the second rule, accuracy was of a good level 80% whereas in the third rule accuracy was 90%, forth rule was 90% included mixed opinions. The fifth rule was not much different in comparison to rule three, four and five.

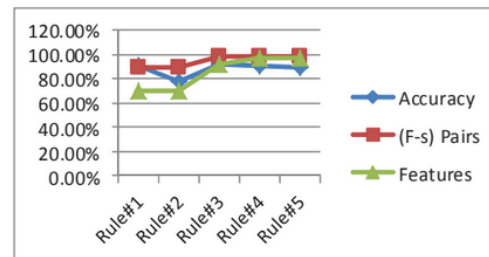


Figure 1 Accuracy Rules (Leila et. al. 2016)

Accuracy has been high, the results were of reviews written in MSA. The accuracy seemed to slightly decrease on third, fourth and fifth rule. As claimed by Leila et. al. (2016) "an English statement which is written with Arabic letters and negations are challenges for future work".

There is no description of how data was annotated to identify whether good or bad science was used, for example whether there was a professional annotator or how many annotators were used to avoid bias. Datasets were collected from various social media forms and they were specific. Therefore, data collection shows variety. In addition, 200 reviews is a large amount, expansion

of terms is recommended. Using reviews of different social media sites is good as it prevents bias for a specific company. Qualitative and quantitative research were applied. A much larger amount of data collection is recommended for improved results.

3 Comparison of ATKS Tool Using Human Annotators

The experiment proposed by Ahmad et. al. (2013) was qualitative as annotators were used, there was also errors in the Arabsenti lexicon's tool that was used thus making the research less accurate.

The experiment of Leila et. al. (2016) relied on human annotators as well as a reliable conversion tool which is the ATKS tool. In conclusion human annotators in addition to the ATKS tool presented a greater accuracy as both had reliable results as errors weren't found in experiment conducted by Leila et. al. (2016).

4 Lexicon-Based and Corpus-Based Positivity and Negativity Categorization of Comments and Reviews

The experiment's data such as comments, reviews and tweets is gathered for annotation purposes to create a model for classification as well as testing (Nawaf et. al. 2013). The annotation will be conducted by categorizing what type of tweets were used and whether the word is formal or sarcastic.

A collection of 2000 tweets were used for the experiment upon annotation. 1000 of those were negative and 1000 were positive, they were collected from two topics. All of these tweets were in MSA and Jordanian dialect.

In order to identify the semantic orientation of the tweets in order for the extraction to work these consisted of emotions and objectives.

Evaluation Metrics	Phase I	Phase II	Phase III
Precision	10.5%	52.2%	58.6%
Recall	8.9%	48.9%	64.9%
Accuracy	16.5%	48.8%	59.6%

Table 6 Lexicon's Scalability Results (Nawaf et. al. 2013)

Two types of experiments were used and these were supervised and unsupervised. Different stemming techniques were applied in the supervised experiment and these were root-stemming, light-stemming and no stemming to identify the effect on the classifier's performance.

Unsupervised techniques were applied on dataset of the collected tweets (2000). This has reported low accuracy. Experiment was conducted gradually, firstly by starting from small size and keeping the original terms and secondly, the number has gone up to 2500 words due to the 300 original stemmed words, and thirdly, the random words were combined including both positive and negative.

The results demonstrated that there is improved accuracy when the lexicon is bigger in size but increasing the lexicon in size doesn't guarantee improved accuracy, in addition this can save time and effort.

Highest accuracy is given by the light stemmed datasets. An improvement of this would be to widen the polarity case with a neutral class. This will give a more valuable results in terms of accuracy especially sarcasm as it can be misunderstood (Nawaf et. al. 2013).

A large number of datasets of the collected tweets (2000 words) was used. The number of negative and positive words were equal. Formal and informal words were chosen of only two genres. It would have been better to have multiple genres for example four or five to allow for a wider range of words.

There were two expert humans for labelling and one expert to solve any conflicts if the other two experts reached. There is a chance of bias as the expert may take one of the expert labeler's side. It is recommended to have at least two more expert consultants. It is understandable that this may be expensive. Quantitative research is applied as labelers were used.

5 Conclusions

Arabic sentimental analysis of social media posts have been analysed fully in this research paper in regards to accuracy. SVM & Naïve Bayes classifiers achieved higher accuracy (Talaat et. al. 2015). Whereas other experiments of sentimental analysis were conducted during the processing phases to determine the difference of the accuracy levels which eventually resulted of being mostly accurate in the pre-processing phase (Nourah et. al. 2016) which makes the method mostly ideal during the pre-processing phase.

Each type of datasets had its own separate experiment which then resulted in unigrams and bigrams working best together. The author Talaat et. al. (2015) identified which datasets realistically didn't seem to work well together.

Rules were compared which then determined that there has been a decrease after the third rule. Filtration supported the experiment through the collection of the samples and ensuring they followed the specified criteria which has been set by the researcher e.g. mixture of topics for the tweets (Leila et. al. 2016) results showed clearly how rules and filtrations had an impact on accuracy levels.

High accuracy has been achieved to a certain extent but it is still not 100% and nowhere near 95%. Sarcasm and emotions are both expressions that don't seem to have a way of solving at this point in time in sentimental analysis (Nawaf et. al. 2013). Therefore, determining positive, negative and neutral posts still remain a challenge of the sentimental analysis process due to Arabic being a complex language and the issue still lies in accuracy.

Sentimental analysis have been experimented on colloquial Arabic including Egyptian and Jordanian (Nawaf et. al. 2013) but there has not been research on any other colloquial Arabic. For further research, translation tools are currently needed to translate from dialect Arabic to modern standard Arabic.

Development of such tools will allow for an advanced translation system in terms of accuracy. There is a need for different systems to cover multiple dialects, as each dialect has its own complexity and unique rules of which these will

require different methods and approaches to resolve the challenges they have.

There are still 20 more dialects which will need focus on in the future specifically the Moroccan dialect as it faces many challenges.

References

- Ahmad Mourad and Kareem Darwish, 2013, 'Subjectivity and Sentiment Analysis of Modern Standard Arabic and Arabic Microblogs'. *4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 55-64, Atlanta, Georgia, June.
- Alok Kothari, Walid Magdy, Kareem Darwish, Ahmed Mourad, and Ahmed Taei, 2013, 'Detecting Comments on News Articles in Microblogs'. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, pages 293-302, June.
- Amal Abdullah AlMansour, 2016, 'Labeling Agreement Level and Classification Accuracy'. *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems*, pages 271-274, Ieee.
- Laila Abd-Elhamid, Doaa Elzanfaly, Ahmad Sharaf Eldin, 2016, 'Feature-Based Sentiment Analysis in Online Arabic Reviews'. *Computer Engineering & Systems (ICCES), 2016 11th International Conference*, pages 260-265, ieee.
- Nawaf A. Abdulla, Nizar A. Ahmed, Mohammed A. Shehab and Mahmoud Al-Ayyoub, 2013, 'Arabic Sentiment Analysis: Lexicon-based and Corpus-based'. *Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEEECT)*, pages 1-6, ieee.
- Nourah F. Bin Hathlian, Alaaedin M. Hafezs, 2016, 'Sentiment - Subjective Analysis Framework for Arabic Social Media Posts'. *Information Technology (Big Data Analysis) (KACSTIT), Saudi International Conference*, pages 1-6, ieee.
- R.M Duwairi, Raed Marji, Narmeen Sha'ban, Sally Rushaidat, 'Sentiment Analysis in Arabic Tweets'. *2014 5th International Conference on Information and Communication Systems (ICICS)*, pages 1-6, ieee.

Santosh K.Ray, Khaled S. (n.d.). 'A Review and Future Perspectives of Arabic Question Answering Systems'. Ieee 2016 Transactions on Knowledge and Data Engineering Vol.28, No.12, pages 3169-3190.

Talaat Khalil, Amal Halaby, Muhammad Hammad, and Samhaa R. El-Beltagy, 2015, 'Which configuration works best? An experimental Study on Supervised Arabic Twitter

Sentiment Analysis'. 2015 First Internaional Conference on Arabic Computational Linguistics, pages 86-93.

Walid Cherif, Abdellah Madani, Mohamed Kissi, 2016, 'A combination of Low-level light stemming and Support Vector Machines for the classification of Arabic opinions'. Intelligent Systems: Theories and Applications (SITA) 2016 11th International Conference pages 1-5, ieee.

Evaluating Current Research on Psychometric Factors Affecting Teachers in ICT Integration

Daniel Otieno Aoko

Abstract

There are various Instruments used to assess the numerous aspects of Technology in learning. This study was aimed at establishing the psychometric factors affecting teachers using technology to enhance learning. One of the modern ways used to substitute conventional methods of teaching by embracing digital learning as modern learning tools (Tinio, 2017). The focus is on current study on the various psychometric factors affecting teachers in ICT integration. The outcome shows that access to technology by most of our educators is not necessarily a proof to determine active usage of this platform. Therefore corrective pointers must be instilled to restore confidence and positive attitude among tutors. The outcome is used to establish and improve on ICT integration leading to new findings drawn and appropriate recommendations made. This is based on tangible evidence contained in the paper and proposal of further research in future to improve on the same.

1 Introduction

Many countries have identified the significant contribution ICT in providing quality interactive learning and have put in infrastructure by investing on digital learning devices and Networking of learning centers (Pelgrum, 2001). Majority researchers have envisioned that a digital content in curriculum is almost becoming mandatory as the modern teaching tool and therefore usage will continue to increase. The biggest challenge is a reality the seamless integration of e-content in learning remains a myth among some of the tutors (Anderson, 2002).

Having looked at several research papers evidence it has come out clearly that to achieve successful use of ICT in educational sector is subject to the attitude and the participation of the educators. It is paramount that the user's perception on the use of e-contents for learning be tamed to avoid a possible resistance to use ICT gadgets in class.

The real impact of ICT is effective when used in content within a confined environment (Parr, 2010). Research evidence shows that unprogressive reforms frustrated by the teachers

beliefs, skills and attitudes were not taken into account. Teachers behaviour, abilities, attitudes not withstanding the existing environment had a far reach consequences to make ICT Intergration both in developing and progressive nations a reality (Mumtaz, 2018). It is a fact that the diversity of people within an Institution from different cultural background, age attitudes and beliefs are key in determing tha ratios of acceptance and sets up percentage score in what can be termed as setting social mood in ICT intergration within our learning centres.

Leadership is very outstanding factor that influences the usage of ICT in Institutions, in schools where principals encourages collaborations between one or more students, teachers and pupils with other schools by means of technology for Educational Exchange a significant success is shown as both learners and teachers makes more effort to adopt and conform to this requirements by actively participating in these activities (Alkahtani, 2016).

The outcome of using technology for intergration purposes varies from one learning centre to another, study has shown that there are gaps yet to be filled on students learning from technology. It

is from this factor that the survey was carried out on psychometric factors affecting teachers using technology (Patnoudes, 2014). Such factors were termed as

- a) Teachers statistics in using ICT at personal level.
- b) Teachers level of skills in using ICT equipment.
- c) Teacher perceptions towards technology
- d) Teachers views of using ICT as an additional classroom learning tool.
- e) Frequency of teachers using ICT for developing classroom resources.
- f) School environment responsiveness to technology.

The study seeks to determine psychometric properties providing statistics of reliable and valid evidence using an examination of the items enumerated.

2.0 Methodology

The approach in the current research was drawn from several researches and outcomes relations formed the conclusions.

2.1 ICT intergration to working experience.

Papanastasiou & Angeli (2008) used a sample of 578 tutors who were teaching in Cyprus junior schools during the year of study 2003 – 2004. The age group of participants tutors engaged on average was 32 years old and the least age was 22 years while the highest age was 59 years. Most of the participants had an average work experience of slightly above 10 years and on the higher side 39 years of work. Five teachers were notably on their first year of posting (0.9% of identified teachers). It is estimated that close to 78% sampled were female representing gender parity variances expected at the elementary levels in Cyprus and estimated 22 % were male at same levels.

Gorder (2008) carried out a study and made conclusion that teachers experience determined the usage of ICT. She further reveals that effective technology usage depends on the personal skills. Those with good skills are used ICT more as opposed to those with inadequate skills.

In relation to computer usage and exposure 96% of the identified tutors acknowledged that they had access to ICT equipments either at work or home. While 70% were identified as having completed preferred professional courses in fundamental ICT skills. Looking at the analysis carried out in these contents there is a clear relationship between gender and ICT integration with high percentage of Male teachers showing more confidence in computer usage than their female counter parts, this variance is as a result of men's ability to take bigger risk than women. This translates again into less usage of ICT in classroom where over 70% of teachers teaching elementary class in Cyprus are women. The conclusion made is that success of ICT integration in classroom relies upon teacher's willingness and the working environment.

Papanastasiou & Angeli (2008) study contradicts (Rahim, 2008) showing that tutors with more working experience had more confidence as opposed to the young, reason given was the fact that long serving teachers were able to know exactly when and where ICT integration is applicable.

While focusing on the same I found out that age is relative and may not necessarily be used to determine ICT Integration because there are incidences where people with different have similarities in competences .

Lau (2008) stated that female teachers are diligent and very positive in accepting and using technology to their male counterparts as perceived by to (Papanastasiou & Angeli, 2008).

Bauaneng-Andoh (2012) carried an analysis that showed that personal characteristics which include gender, age, educational Backgrounds and teaching experiences play a great role when it comes to effective ICT implementation.

The study showed that most of the younger teachers used ICT often as opposed to the elderly ones. (Rozell, 1999) relates tutors attitudes towards ICT; the study outcome was that tutors experience in using computers is most likely to influence their attitudes in employing use of ICT integration.

In my view various study conducted failed to agree that age was a qualifying factor to determine the frequency at which tutors employ the use of ICT

in the classroom. There are cases where the old were seen to be doing better and in some cases the young where experience is equated to age while the young teachers are seen to be doing well because they have just left college where trainings on ICT skills are conducted.

2.2 Social factors to ICT intergration

Buabeng-Andoh (2012) established that technological and institutional factors significantly contributes to encouraging he cited lack of appropriate skills, self-confidence, ideal learning programs, poor ICT infrastructure and rigid curriculum severely interfered with the tutors in using ICT integration in teaching and enhancing learning in classroom. The study concluded that the Institutions were better placed in addressing these barriers if ever ICT integration were to become effective.

Keengwe (2008) carried out a research whose aim was to make ICT Integration more effective the outcome shows that teachers' supports and attitudes will either positively or negatively affect use of computers in learning process. The conclusion arrived at was that beliefs and behavior of tutors would determine the success of ICT integration.

Kandasamy (2013) conducted a study on ICT usage in educational learning centers in Malaysia with 60% of the participants acknowledging that they employ the use of ICT in learning and collaborations among tutors and pupils, while 80% of the responded cited lack of time in their respective school as a major barrier in employing use of ICT in teaching.

Yunus (2007) carried out a research in Malaysia to establish how ESL tutors used ICT in their learning centers. This study was conducted in technical Institutions through surveys and partial interviews with the tutors. The study aimed at finding attitudes and factors associated to impact of teaching using ICT. Technology Acceptance Model (TAM) was employed in carrying out this analysis. The results termed majority of educators had access to computers at home and were positive about ICT. An estimated 76% of the educators could only access one ICT lab and therefore ICT integration becomes a big challenge due to constraints of the available facilities. 75% of

teachers identified poor quality of computer hardware as a major barrier in Integration.

Davis (1989) used Technology Acceptance Model to test the ICT usage and attitudes among educators. The aim of the research was to establish the percetion of the participants with regard to ICT usage and adequate skill among the tutors. The model shows how user accept and apply the use of ICT. The evidence of the outcome is that when users are subjected to new systems there factors that determines perceived usefulness. He concluded that the level of ones perception in the usefulness of the program is most likely to influence the use of ICT the evidence shows that professional skills is a contributing factor to ICT usage .

Figure 1. This structure was used to Investigate the teachers acceptance level of using ICT

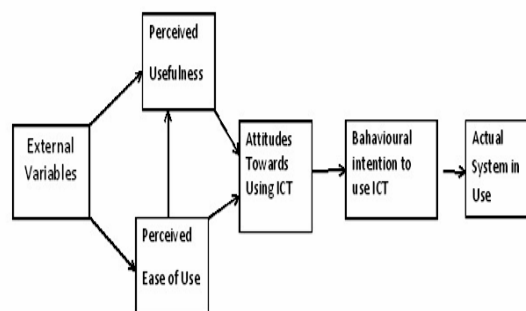


Figure 1 Technology Acceptance Model (Davis, 1989).

Kandasamy (2013) carried out experiments conducted to determine effects Gender relationships in all the items of study.

The study established that High numbers of male tutors have sufficient skill in running programs as opposed to their female counterparts rated ($F=6.28$, $p=0.012$), Male tutors appear to be to be more conversant with regular programs than the female ($F=21.69$, $p<0.000$) and tailored programs ($F=13.75$, $p<0.000$). While measuring the self-esteem males were found to be more attracted to using technology as a teaching tool as opposed to their female counterparts. ($F=24.69$, $p<0.003$).

Item of consideration	Type II sum of squares	df	Mean square	F	sig.	R ²	Females	Males	Reliability
Knowledge of Regular programs	8.118	1	8.118	6.28	.012*	0.011	3.185	3.475	0.9
Knowledge of tailored programs	6.812	1	6.812	11.894	.001*	0.021	1.687	1.954	0.83
Use of regular applications	14.428	1	14.428	21.69	.000*	0.038	2.629	3.015	0.84
Use of tailored applications	2.292	1	2.292	13.753	.000*	0.024	1.216	1.369	0.78
ICT self esteem	4.356	1	4.356	4.966	0.026	0.009	2.948	3.16	0.89
Sensitization from peers	0.8	1	0.8	1.289	0.257	0.002	3.211	3.302	0.86
Desire to use ICT	16.711	1	16.711	24.698	.000*	0.042	3.511	3.927	0.82
Personal beliefs and Values of ICT	2.449	1	2.449	8.878	.003*	0.016	3.747	3.906	0.76
Technology as a tool for change	0.001	1	0.001	0.004	0.951	0.000	3.949	3.953	0.59
Technology Physical environment	1.542	1	1.542	2.132	0.145	0.004	2.988	3.115	0.76

Figure 2 Shows experiments comparing results of various items of this study (Kandasamy, 2013).

Krishnan (2015) carried out an experiment by employing SEM algorithm the result was a good fit. Using path evaluation ten hypotheses were tested. The highest level of satisfaction was recorded in technology effectiveness. An average score was recorded on behavior in ICT usage.

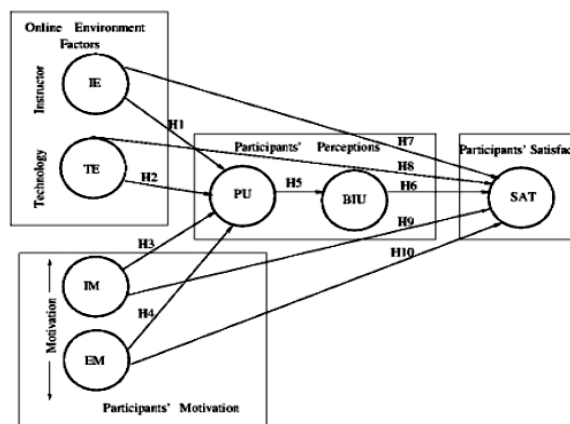


Figure 3 Research Model (Kannan, 2015).

KEY:

- H1 – Tutor will be positive on usefulness of ICT workshop.
- H2 _The workshop will positively improve technology effectiveness among tutors.

- H3 and H4 _ Tutors motivation will have a positive impact in technology.
- H5_ Perceived usefulness to impact positively to change in attitude.
- H6_ Change of attitude will positively impact on use of ICT.
- H7_ Tutors effectiveness leads to positive ICT usage.
- H8_ Technological reliability will lead to satisfaction in ICT Intergration.
- H9 and H10 – Tutors motivation will positively influence ICT intergration.

This test conducted by Kannan (2015), H1, showed that with access to training teachers had positive attitude towards ICT intergration. H2 provides evidence that proficiency trainings contributes significantly to educators skills in ICT intergration. H3 and H4 shows that teachers attitudes were changed when the were motivated as a result a positive outcome was registered in utilizing ICT. H5 provides the evidence that teachers perceptions largely contributes to change in attitude using ICT. While H6_ justified that a change in attitude positively lead top increased usage of ICT by the educators as shown in figure 3. Study shows that tutors with adequate skills oftenly use ICT more H7. The condition of ICT equipments was also cited to have an impact in ICT usage. In Institutions where high frequency of repairs is recorded teachers tend to be discouraged compared to similar Institutions where ICT equipments are in good condition and sufficient H8. H9 and H10.

The test carried was sampled from 100 teachers data collected from teachers who were operating from different geographical locations. The study adopted the online survey method. The outcome of the findings is sufficient to prove that teachers perception influences ICT usage.

Volman (2005) conducted a study that showed that the female showed less effort in learning ICT at high school and after secondary compared to their male counter parts. Contrary to (Watson, 2006) conducted a study in Queensland state schools on use of ICT from 929 educators revealed that female teachers are least participating in ICT integration compared to their male counter parts. In comparison with US Mid-western schools (Breisser, 2006) findings showed that female teachers drastically improved in their perceptions as opposed to their male counter parts that

remained dormant, (Adam, 2002) accepted an outcome of a study that concluded that female teachers applied ICT more than their male counterparts. (Yukselturk, 2009) justifies that gender parity is not a determinant on use of ICT among teachers' facts more female teachers were seen using internet technologies. (Kay, 2006) study concluded that although male teachers had high ability and attitude but variances existed between female and male teachers after implanting ICT his conclusion is that training played vital role in realigning the disparities.

3.0 Conclusions

Integration of ICT has continued to grow increasingly ambitious that it is almost becoming mandatory for every teacher and student to live with it. This process has equally been met by certain social barriers mainly from the educators like appropriate skills, physical ICT environment, attitudes and tutors motivation to use ICT in teaching of other areas as a modern class room tool (Shah, 2013).

Having analysed, compared and contrasted the psychometric factors affecting teachers in ICT integration that impede their energy to teach using ICT equipment's, my findings revealed that majority of the researchers agreed that ICT integration is inexistence but its success will depend on the efforts that will be taken to turn around the underlined issues. Among the issues I established affecting the tutors are knowledge in using regular applications, tutors' behaviour and value in ICT integration, use of tailored programs, tutors' self-esteem, sensitization by the peers, and attraction in using ICT equipment's, technology physical environment, ICT as tool of change in learning. My Investigations shows that various responses have a reliable outcome that highly depicts the actual situation. In my view from the various evidence gathered teachers have come out to play an important role in making ICT integration yet a reality. Therefore there is a need to create strategic plans in employing corrective measures to counter various challenges opposed to making ICT integrations a reality.

The need for sensitization of educators', training and leadership that fosters positive attitudes to learners and tutors, team work and self-drive are among the key pointers that can be used to reverse these trend (Singhavi, 2017).

Having gone through various journals and conferences, I carried out various analysis with sufficient evidence and concluded that the presence of ICT equipment's in learning centres does not necessary translate to their usage in enhancing learning activities. The reality is that by addressing the psychometric factors discussed in this paper and applying corrective pointers will certainly change learning using ICT.

4.0 Further Work

Since the report was conducted as a self-report it could be possible some of our respondent fully with regards to social responsibility. Therefore it is highly recommended that a cross validation may be essential to establish the tutors' behaviour, skills and attitudes in ICT integration by conducting a further investigation to make the way ICT integration in our learning centres a reality

References

- Abd. Rahim, B. S., 2008, 'Teaching Using Information Communication Technology: Do trainee teachers have the confidence?' *International Journal of Education and development Using ICT*, 4(1).
- Adam, N., 2002, 'Educational computing concerns of postsecondary faculty', *Research on Technology in Education*, 285 - 303.
- Alkahtani., 2016, 'ICT in teaching in Saudi Secondary Schools.' *International Journal of education and Development using Information and Communication Technology.*, pg 34.
- Anderson, K., 2002, 'Factors Affecting Teaching with Technology.' *Educational Technology & Society.*, pg 69-86.
- Bauaneng-Andoh, C., 2012, 'Factors influencing teachers' adoption and intergration of information and communication technology into teaching', *International Journal of Education and Development using Information and Communication Technology*, 8(1), 136-155.
- Breisser, S., 2006, 'An examination of gender differences in elementary constructionist

classrooms using Lego/Logo instruction', *Computers in the Schools*, Vol 22 pp. 7-19.

Davis, 1989, 'Technology Acceptance Model' *Information seeking Behaviour and Technology Adoption*, pp. 319-340.

Gorder, L., 2008, 'A study of teacher perceptions of instructional technology intergration in the classroom.' *Delta Pi Epsilon Journal*, Vol 50, pp. 21-46.

Kandasamy, 2013, 'Knowledge, Attitude and Use of ICT among ESL Teachers', *GSE Journal of Education* (p. 185). Malaysia: National University of Malaysia.

Kannan, K., 2015, 'A structural Equation Modelling Approach for Massive Blended Synchronous Teacher Training', *Educational Technology & Society*, Vol 18.

Kay, R., 2006, 'Addressing gender differences in computer ability, attitudes and use: The laptop effect', *Journal of Educational Computing Research*, Vol 34, pp 187-211.

Keengwe, J., 2008, 'Computer Technology intergration and student learning: Barriers and Promise.', *Journal of Science Educational and Technology*, 560 - 565.

Krishnan, K., 2015, 'A Structural Equation Modelling Approach for Massive Blended Synchronous Teacher Training.', *Educational Technology & Society*, 1-15.

Lau, T., 2008, 'Learning behaviour of university's business students in regards to gender and levels of study-An exploratory research.', *Proceedings of International Conference Sciences and. Malaysia*.

Mumtaz., 2018, 'Factors Affecting teachers' use of information and communication Technology.', *Journal of Information Technology for Teacher Education*, pg 319-342.

Papanastasiou., 2008, 'Factors Affecting Teachers Teaching with Technology.', *Educational Technology & Society*, Pg 69-89.

Parr, W., 2010, 'Prospective EFI Teachers' Perception of ICT Intergration', *Educational Technology & Society*, pg 185-196.

Patnoudes, 2014, 'Assessment of Teachers Abilities to Support Blended Learning Implementation in Tanzania.', *Contemporary Educational Technology*.

Pelgrum, 2001, 'Factors affecting teachers teaching with technology.', *Educational Technology & Society*, Pg 69-86.

Rozell, E. G., 1999, 'Computers in human Behavior.', *Computer-related success and failure, a longitudinal field study of factors influency computer-related performance.*, Vol 15, 1-10.

Shah, K., 2013, 'Knowledge, Attitudes and Use of ICT Among ESL teachers.', *GSE Journal of education (worldconferences.net)*, pg.186.

Singhavi, B., 2017, 'Factors Affecting teachers perceived in Using ICT in the classroom.', *IAFOR Journal of Education*, pg 69.

Tinio., 2017, 'Centralised Learning and assessment tool for Department of Education.', *International Journal of computing sciences*, vol No1 Pg 21-35.

Volman, M., 2005, 'New technologies, new differences. Gender and ethnic differences in pupils' use of ICT in primary and secondary education.', *Computers & Education*, 45, 35- 55.

Watson., 2006, 'Technology Professional development: Long-term effects on teacher self-efficacy.', *Journal of Technology and Teacher education*, vol. 14 151 -166.

Yukselturk. E., 2009, 'Gender differences in self-regulated online learning environment.', *Journal of Educational Technology & Society*, 12-22.

Yunus. M. M., 2007, 'Malaysian ESL teachers' use of ICT in their classrooms: expectations and realities.', *European Association for Computer Assisted Language Learning*, ReCALL, 19(1):79-95.

A Critical Analysis of Current Measures for Preventing Use of Fraudulent Resources in Cloud Computing

Grant Bulman

Abstract

Economic Denial of Sustainability (EDOS) attacks could have huge financial implications for an organisation, whether this EDOS attacks where renting server space within The Cloud on a Pay-As-You-Go basis, or DDOS attacks. This paper discusses current DDOS/EDOS prevention algorithms in place, as well as provide a critical evaluation of these algorithms. Furthermore, a comparison is made between each algorithm based on the experiments performed. Penultimately, methodologies will then be fully examined in order to propose the best solution from the algorithms evaluated. Finally, conclusions will be provided and recommendations made based on the critical evaluation of these current algorithms.

1 Introduction

Cloud Computing is revolutionising the way modern businesses store their data and the way in which their services are provided. Threats posed from DDOS and EDOS attacks are increasing dramatically. In 2016, we saw the highest amount of DDOS attacks in history. Saied, A et. al. (2015) state that “DDOS attacks are serious security issues that cost organisations and individuals a great deal of time, money and reputation, yet they do not usually result in the compromise of either credentials or data loss.”

Idziorek J et. al., (2012) state that use of fraudulent resources “is a considerably more subtle attack that instead seeks to disrupt the long-term financial viability of operating in the cloud by exploiting the utility pricing model over an extended time period”.

The Cloud offers businesses the flexibility of renting server space/bandwidth on a Pay-As-You-Go basis, meaning they only pay for the bandwidth used. Somani G et. al. (2016) state that “Economic aspects are affected because of the high resource and energy usage, and the resultant resource addition and plugging, thus generating heavy bills owing to the “pay-as-you-go” billing method”.

Somani G et. al. (2016) go on to develop a system to better understand a DDOS attack and conclude that “this model has also detailed the resource

overload state of a virtual machine under attack and its possible spread using vertical scaling, horizontal scaling and migrations”. They also go on to differentiate and relate DDOS and its economic version EDOS.

An EDOS Attack is very subtle in the way in which it is performed and is very hard to identify, unlike a DDOS attack. The attacker would generally be motivated to perform this attack at a specific organisation. The attacker would ping the server, consume as much bandwidth data as possible without any dramatic traffic being identified by the client. Alosaimi W et. al. (2015) create and test a new algorithm to protect the cloud environment from both DDOS and EDOS attacks.

Over the course of this research paper we will analyse different methods of detection of Denial of Service (DDOS) and Economic Denial of Service (EDOS) attacks and compare these. The experiments used, as well as their claims will be critically evaluated in order to select the most practical method for detecting these types of attack.

2 Current Measures for Preventing Fraudulent use in the Cloud

The following section provides an evaluation of current algorithms in place for preventing fraudulent use of cloud resources.

2.1 DDOS Attacks

Somani G, et. al. (2016) state that DDOS attacks target the victim server by sending a high volume of service requests to the server by using a bot. Nowadays, Botnets can easily be obtained for free online and are amongst the most popular types of cybercrime.

They propose an algorithm known as Victim Service Containment, which aims to minimise the effects a DDOS attack can have both physically and financially. They do this by using a system called DDOS Deflate which will identify if an attacker has made more than 150 connections to the client's server (fig1).

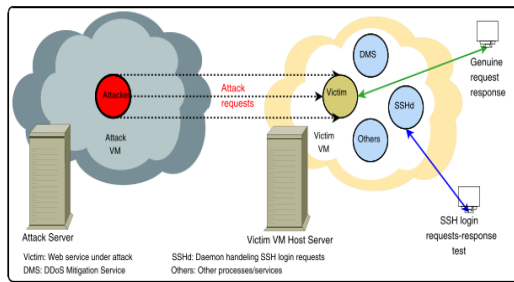


Figure 1: Experimental setup (Somani G et. al. 2016)

They perform the experiment by hosting two VM's (Virtual Machine), one is the victim VM and the other is the attacker. They then send 500 Secure Shell (SSH) requests, 100 genuine requests (each request logs out of the session before the next request is sent) and 500 concurrent attack requests all simultaneously. The results are as follows:

Table 2

Various Metrics: DDoS attack experimental study.

Attack reporting time	Victim service unavailability time	SSH unavailability time	Maximum response time (SSH)	Minimum response time (SSH)	Average response time (SSH)
39 s	945 s	517 s	186.830 s	0.129 s	1.226 s

Figure 2: Experimental results (Somani G et. al. 2016)

From the results of this experiment they then create an algorithm which calculates resource in order to contain resource contention.

The authors conducted the experiment in a controlled environment which helps reduce the risk of bias. They also repeated the experiment to get accurate data before they published the results. Additionally, they compared this technique to other techniques in order to check if their algorithm is more effective. Therefore the experiment performed was of a high level as the work done involved 500 attack requests and it was made clear in their conclusions that more research is needed and further work is needed before the algorithm is to be applied in the real world.

In order to detect a DDOS attack sooner, Hoque N, et. al. (2017) propose a new method to actively detect a DDOS attack as it is happening, known as NaHIDverc. The experiment is performed by capturing raw data from the router as "TCP/IP network layer packets, which are subsequently sent to the pre-processor module." Hoque N, et. al. (2017).

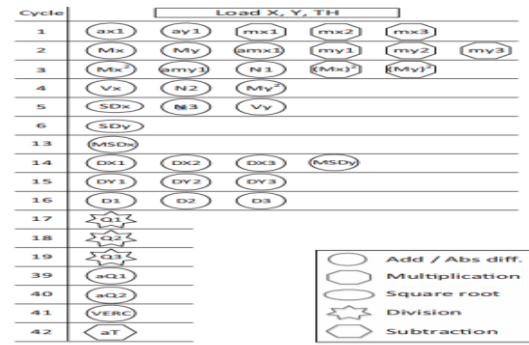


Fig. 3. Implementation model of the detection module.

Figure 3: Implementation model (N et. al. 2017)

In order to evaluate the results, they use three network intrusion datasets: CAIDA, DARPA and TUIDS.

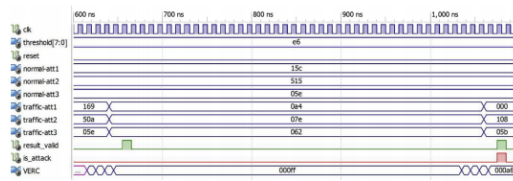


Fig. 11. Simulation waveforms demonstrating the operation of the DDoS attack detection module.

Figure 4: Simulation waveforms demonstrating the operation of the DDOS Attack Detection Model (N et. al. 2017)

The results from this experiment showed a 99% detection rate on the CAIDA dataset, 100% accuracy on the DARPA dataset and 100% detection using the TUIDS dataset, therefore they conclude that they fully met the hypothesis of the research, which was to detect all DDOS and sooner.

The experiment was carried out by the authors was performed fairly and under a controlled lab environment to reduce the risk of bias. Additionally, they tested their algorithm under three different intrusion datasets which proved some very promising results when it comes to detecting this type of attack sooner, including detecting an EDOS and FRC attack which has financial impacts. Although the work demonstrates higher levels of testability than other work in this area, the validity of the results should be questioned because it could be argued that the authors could have manipulated the variables to better the results. Conclusively, the experiment should be repeated to get a mean average of the results in order to get a better understanding of its accuracy and this should be performed in a secure, controlled environment to ensure these results are valid.

Wang C et. al. (2017) also propose a new algorithm for detecting DDOS attacks effectively, based on RDF-SVM. Their algorithm is developed in Python and aims to detect unusual incoming traffic and validates the precision rate of this. They compare this to two other algorithms: SVM and RF and SVM to compare the results.

They compare all three algorithms and the results show that the RDF-SVM algorithm has an 82.5% detection precision rate and overall 80.09% recall rate – which is the highest of the three.

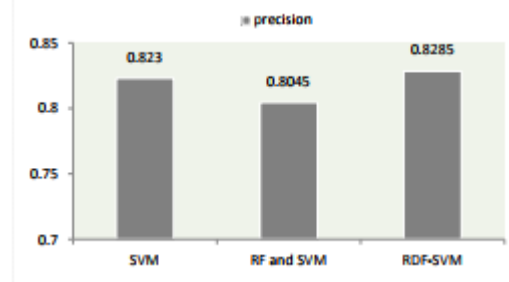


Figure 5: Precision rate three methods (Wang C et. al. 2017)

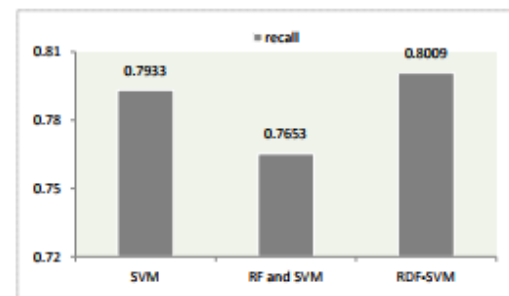


Figure 6: Recall rate three methods (Wang C et. al. 2017)

The authors conclude that the RDF-SVM algorithm detects DDOS attacks, both known and unknown effectively.

Overall, the experiment conducted by the authors was compared with other algorithms in place to detect better accuracy and give better results, therefore this gave a better understanding of the results. Additionally, the authors also made a comparison among three methods, to ensure the validity of the results given.

Wang B et. al. (2015) also propose a similar DDOS mitigation technique, called DaMask. This technique consists of three layers: the network switch, the network controller and the network application. The purpose of this mitigation technique is to detect DDOS attacks quickly and react instantly.

The test is performed in the hybrid Cloud, again using the Amazon EC2 service and the authors use Mininet to create a virtual network to emulate the SDN setting used during the experiment.

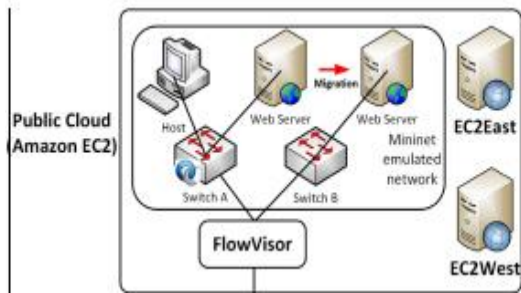


Figure 7: Experimental setup on Public Cloud Amazon EC2 (Wang B et. al. 2015)

They then compare the results of this experiment with the Bayesian technique as well as the Snort mitigation technique, which is a free open-source detection system. The results from the experiment show that the new DaMask technique is similar to that of the Bayesian technique.

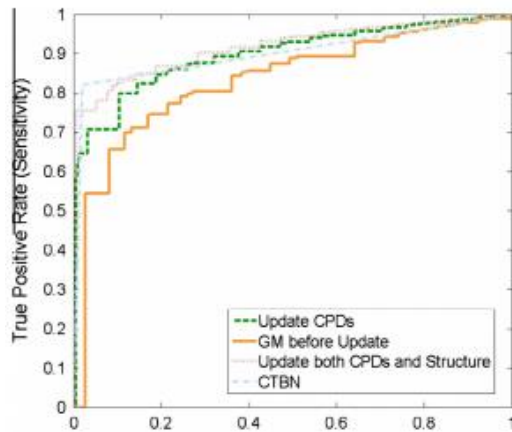


Figure 8: Experimental setup on Public Cloud Amazon EC2 (Wang B et. al. 2015)

The authors conclude that whilst the DaMask technique is similar to current techniques, it requires little effort from the cloud provider, meaning minimal changes required from the computing service architecture. Overall, the authors have conducted a fairly good experiment. They firstly checked the bandwidth speed and logged this to prevent bias, as well as conducted the experiment in a controlled environment. Additionally, they compared their technique with others and concluded that the results are similar.

Conclusively, the authors carried out extensive testing during the experiment, which eliminated bias from the results. It should be noted however

that they use Mininet to create a virtual network which is rather outdated. They could have improved the validity of the results by using a more recent version to get better results rather than ones similar to the Bayesian and Snort techniques.

2.2 EDOS Attacks

Wang H et. al. (2016) state that “Distributed Denial of Service (DDOS) attacks have evolved to a new type of attack called Economic Denial of Sustainability (EDOS) attack”. Unlike a DDOS attack, an EDOS attack aims to financially impact the victim through use of the Cloud’s Pay-As-You-Go model.

They perform their controlled experiment by building a website in the Amazon Cloud and “The website hosts various recourses including images with sizes from 38KB to 40MB, videos with sizes from 2MB to 171MB and documents with size from 10KB to 10MB” Wang H et. al. (2016). The attack laptop then calls each provider to bring the targeted resources 2000 times with a request interval of 10 seconds.

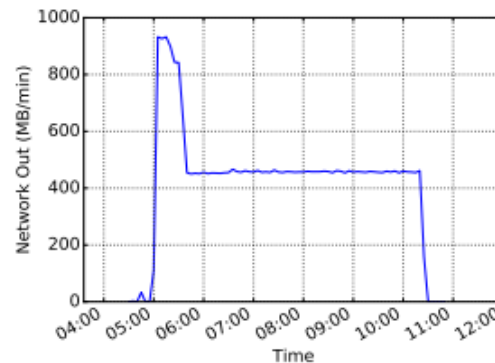


Figure 8: Average network out traffic during experiment (Wang H et. al. 2016)

The results show that the attacker incurred charges of \$11.87 to the victim in this short test alone.

In order to keep costs to a minimum, Wang H et. al. (2016) propose a ‘Redirection-Based Defense Mechanism’ which aims to redirect third-party services to the URLs with a valid cache and checking it has a cache hit. They go on to state that the victim then experiences much less traffic when this algorithm is tested.

The research conducted by Wang H et. al. (2016) is of a very high quality. They analyse current algorithms for preventing EDOS attacks and present their conclusions. However, their own experiment is only performed once and they jump to conclusions from this. They then state terms such as “We imagine...” which is not justified nor is it good science.

Baig Z et. al. (2016) propose a mitigation technique for detecting EDOS attacks sooner. To do this, they deployed two Cloud servers and the upper and lower limit of auto-scaling are set to 80% and 30%.

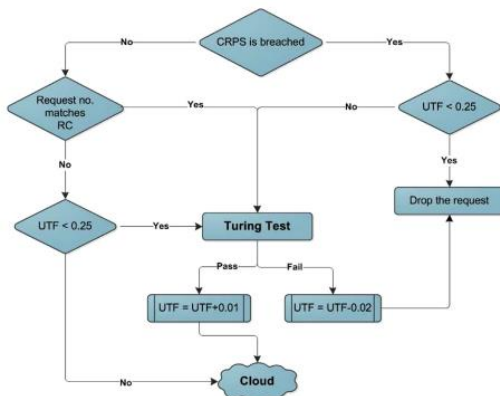


Fig. 4. VMInvestigator workflow.

Figure 9: Parameters used in experiment (Baig Z et. al. 2016)

To perform the experiment they assumed a normal CPU usage rate of 40% is legitimate usage and the cost of a VM instance as \$0.03. They then send between 200 and 400 requests per second to the victim server, up to a maximum of 1200 requests (see Fig 6) before the CPU peaks its consumption.

The results show that without their mitigation technique in place the costs billed to the victim sever would increase dramatically. The mitigation technique in effect identifies suspicious requests targeting the victim sever through use of a Firewall which filters the incoming traffic and any unusual request are added to a blacklist.

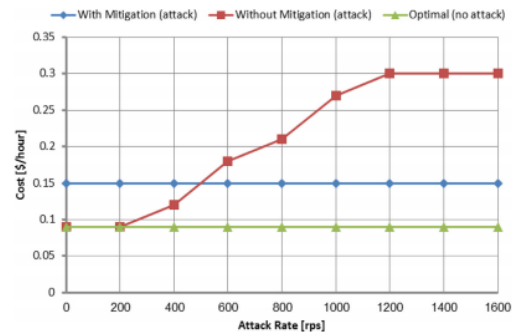


Figure 10: EDOS attack effect against CPU usage (Baig Z et. al. 2016)

The results from the experiment show that with the mitigation technique in place, the cost stays at a steady rate (Fig 6) and therefore the victim will be billed less than without the technique. The author concludes that the technique is able to detect intelligent/smart attackers with a good degree of accuracy, preventing higher bills to the victim.

The experiment carried out by the author followed good science principles as it in a controlled environment and was repeated 10 times, then the results averaged. From the results, it is clear that the mitigation technique did reduce the cost to the victim, however there is still an increase and therefore we still need further work on this mitigation technique.

Additionally, VivinSandar, S et. al. (2012) propose an algorithm which works in a similar way to Wang H et. al. (2016), where a request is made by a user and is first intercepted by the firewall. This is sent to an on demand puzzle server; the user then must solve the puzzle and if the server verifies the result is correct they will be added to the firewall ‘white list’.

The authors do this by conducting their experiment in the EC2 Cloud. Four EC2 instances are grouped together and to simulate an attack they sent repeated HTTP requests to the victim server and all packets are monitored through a packet capturing application, in this case they use Wireshark. The results of the authors experiment is as follows:

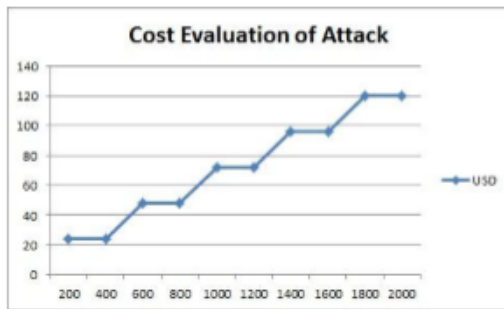


Figure 11: Number of attacks vs Cost
(VivinSandar, S et. al. 2012)

The results show that as more requests are sent to the server, the cost applied to the victim server again increases rapidly. Although the algorithm was in place, the results show that the cost factor still caused an issue and therefore they conclude that further search is needed to provide a better mechanism for detecting an EDOS attack sooner.

Overall, aside of the results from the test, the authors conducted this in a controlled environment to reduce bias. However, they only performed the experiment once, which should have been done more to check for any discrepancies. The conclusion however did match the experiments and the author made it clear that further research into EDOS prevention was needed.

Masood M et. al. (2013) have also proposed another similar EDOS mitigation technique called EDOS Armor. This technique has three components: challenge, admission, congestion control. The technique in effect only allows a certain number of users to access the server to avoid DDOS. Then, they check browsing behavioural patterns to allocate priority to users based on the user's priority level.

The experiment works when a client's access the server, they are passed to the challenge server which asks them to complete a puzzle, if they complete it correctly they can access the server. After this, the congestion control then filters out good clients from bad clients (bad clients being the ones sending tons of requests to the server) and allocates less resource to the bad client(s).

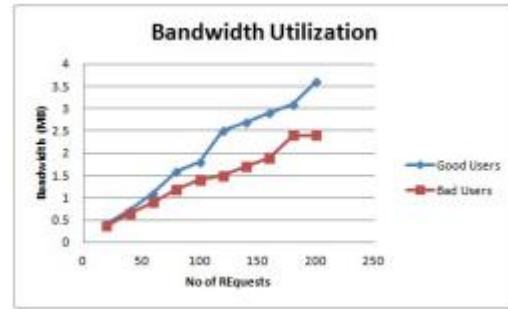


Figure 12: good client's vs Bad clients
(Masood M et. al. 2013)

The results show (Fig 11) that as the requests get higher, the mitigation technique reduces the resource allocated to what it deems as bad clients and therefore the bandwidth rate is less. They conclude that this is a good technique for filtering out good and bad clients, which will result in less impact of an EDOS attack.

Overall, the authors experiment is good as the results clearly show a difference in bandwidth allocation between good and potential bad clients. However, the experiment is not repeated and results averaged to get a better accuracy and they do not seem to have compared and tested this with similar mitigation techniques in place. The results could have been less biased had this been done.

3 Comparison of Current Measures

Upon evaluating each mitigation technique, the most favourable technique is proposed by Hoque N, et. al. (2017). The accuracy rate showed a 99% detection rate on the CAIDA dataset, 100% accuracy on the DARPA dataset and 100% detection using the TUIDS dataset. All of which are very promising results.

Wang C et. al. (2017)'s RDF-SVM algorithm has an 82.5% detection precision rate, which is lower than the experiment performed by N, et. al. (2017), and unlike Hoque N, et. al. (2017) they compared this with two datasets (KDD Train and KDD Test) which are very outdated and are irrelevant for modern intrusion detection systems.

A modern dataset would be that used by Hoque N, et. al. (2017), such as CAIDA. Therefore there is no surprise that their results were so much higher than other work evaluated in this paper.

4 Conclusions

This paper has critically evaluated 8 different mitigation techniques for detecting DDOS and EDOS attacks, which are both performed in very similar ways, quicker. It should be noted that this is a very large research area and these 8 techniques only represent a small portion of the techniques in place. Due to the importance of this field and the rise in DDOS and EDOS attacks it is anticipated that research will be constantly performed for the foreseeable future.

All mitigation techniques are similar in the way they have been created, however some have been performed in a way that has yielded better results for detecting such attacks.

Wang C et. al. (2017) concluded that their technique “can detect known and unknown attacks and distinguish random IP address attacks, real IP address attacks and Flash crowd more effectively” compared with other methods, whilst Hoque N, et. al. (2017) concluded that NahidVerc “is able to achieve an attack detection accuracy of 100% over benchmark datasets”.

Conclusively, it is evident that mitigation techniques for preventing DDOS and EDOS attacks are still not perfect, however they are becoming more and more accurate as these types of attack evolve. However, a lot of the experiments performed in this paper were performed with outdated software, datasets and not in a fully controlled lab environment. EDOS and DDOS attacks are constantly evolving and as a result we need to continuously build and use new datasets to test prevention of these – using outdated datasets will not prevent modern attacks. Therefore further research in this area is still needed and will be for some time.

References

- Alosaimi, W. Zak, M. Al-Begain, K (2015), ‘Denial of Service Attacks Mitigation in the Cloud’. *9th International Conference on Next Generation Mobile Applications, Services and Technologies*. pp47-53.
- Baig, Z. Sait, S. Binbeshr, F (2016), ‘Controlled access to Cloud Resources for mitigating Economic Denial of Sustainability (EDOS) attacks’. *Computer Networks* 97 pp31-47.
- Hoque, N. Kashyap, D and Bhattacharyya D.K. (2017), ‘Real-time DDOS Attack Detection using FPGA’, *Austin, Texas, August. AAAI*. pp 198-202
- Idziorek J, Tannian M and Jacobson D, 2012, ‘Attribution of Fraudulent Resource Consumption in the Cloud’, *IEEE Fifth International Conference on Cloud Computing*. Pages 99-100.
- Masood, M. Anwar, Z, Raza S. A. and Hur M. A, "EDoS Armor: A cost effective economic denial of sustainability attack mitigation framework for e-commerce applications in cloud environments," *INMIC, Lahore, 2013*, pp. 37-42.
- Saied, A. Overill, R. Radzik, T, (2015). ‘Neurocomputing’ 172. *Computer Networks*, 110, pp385-393.
- Somani, G. Gaur, M. Sanghi, D and Conti, M. (2016). ‘DDOS Attacks in Cloud Computing: Collateral Damage to non-targets’. *Computer Networks*, 110, pp48-58.
- VivinSander, S. Shenai. (2012), ‘Economic Denial of Sustainability (EDOS) in Cloud Services using HTTP and XML based DDOS Attacks’. *International Journal of Computer Applications*. 41 pp-11-16.
- Wang, B. Zheng, Y. Lou, W. Hou T. (2015), ‘DDOS attack protection in the era of cloud computing and software-defined networking’. *Computer Networks*. 81. Pp-308-319.
- Wang, C. Zheng, J. Li, X. (2017), ‘Research on DDOS Attacks Detection Based on RDF-SVM’. *International Conference on Intelligent Computation Technology and Automation*. Pp-161-165.
- Wang, H. Xi, Z, Li, F and Chen, S. (2016), ‘Abusing Public Third-Party Services for EDOS Attacks’. *Proceedings of the 10th USENIX Conference of Offensive Technologies*. Pp-155-167.

An Analytical Assessment of Modern Human Robot Interaction Systems

Dominic Button

Abstract

Human robot interactions are becoming ever prominent in the workplace. This paper analyses current human-robot interaction systems, evaluating the method in terms of robot learning and user interactions, various research papers on current topics such as subjective computing, data-driven and adaptive incremental learning. The methods are evaluated with a comparison of data-driven methods being provided. Finally, conclusions are reached demonstrating data-driven being the most prominent however combination of techniques is also suggested including further research for more progress.

1 Introduction

Human Robot interactions are becoming ever present with the number of environments which now house robots expanding Zhang and Wu (2015) articulate “Social robots have been deployed for different applications, such as supporting children in hospitals, supporting elderly living on their own.” due to this the way in which humans and robots interact must be considered.

Research has been conducted in human-robot interactions such improving GUI for users however the research covered in this paper will be specifically addressing social learning techniques to improve human-robot interactions.

De Greef and Belpaeme (2015) explores the possibility of social learning to improve human robot interactions stating, “Social learning has the potential to be an equally potent learning strategy for artificial systems and robots in specific.” While their research clarifies that social learning is an area that would be beneficial they also highlight the current limitations in this area “However, given the complexity and unstructured nature of social learning, implementing social machine learning proves to be a challenging problem”.

Wiltshire, et. al. (2016) reconnoiters the possibility that human perceptions of robots must be altered in a way in which they are teammates,

collaborators and partners through the advancement of social cognition for HRI.

Furthermore, Biswas and Murray (2016) elaborated on this by researching cognitive personality traits stating, “cognitive personality trait attributes in robots can make them more acceptable to humans” by creating an emotional bond between the humans and robots this would allow for an improved way in which humans and robots interacted with one another.

This survey paper will analytically assess current research that is being partaken in human robot interaction systems concentrating on social learning for the robot such as subjective computing, data driven and adaptive incremental learning.

2 Social Learning Research

This section will look at research aimed at the social learning of robots, to provide improved human-robot interaction’s through the robot being able to learn from their human counterparts.

2.1 Subjective Computing

Grüneberg and Suzuki (2014) propose subjective computing be used, allowing robots to exhibit more adaptive and flexible behaviors. The method explores the possibility of a robot to have autonomous self-referentiality and direct world-coupling, this was done using the coaching of a

reinforcement learning agent through binary feedback.

The first experiment tested a 6 DOF robotic arm in both a simulated and real environment. The task aimed at having the human trainers instruct the learning agent in balancing an inverted pendulum through binary feedback.

Experiment two utilized Nao robots, “coached Nao” is the subjective agent whereas “single Nao” is an adaptive agent. Coached Nao must sort colored balls depending on user preference. Whereas single Nao must re-unite a green ball and a red ball at a yellow spot by itself.

Finally, a questionnaire on the robots were given, participants ranged in nationality, age, and gender in addition to having a non-engineering background. Two videos were shown of coached Nao and single Nao performing the previous task, questions were based on the videos in specific domains such as autonomy and individuality with all questions being gathered using the Likert method.

Grüneberg and Suzuki (2014) explained that the results of individuality and situatedness showed no significant differences were noted between “coached Nao” and “single Nao” however despite this results indicate participants enjoyed the social and interactive behavior of the coached Nao when in comparison to single Nao demonstrated in the chart below.

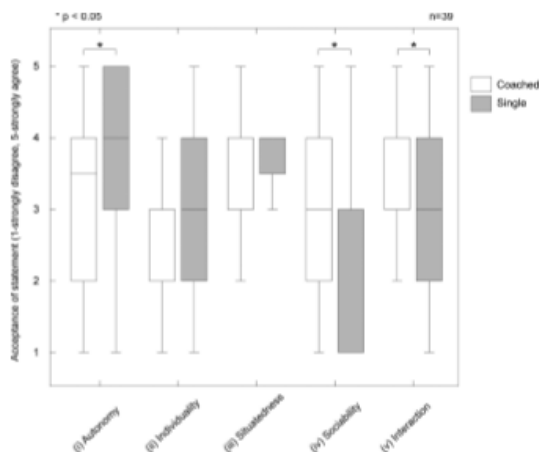


Chart 1: Differences Between Single and Coached Nao (Grüneberg and Suzuki, 2014)

Grüneberg and Suzuki (2014) concludes their research improved human robot interactions through subjective computing, demonstrated through the first experiment which saw the pendulum being balanced for 1 second. Furthermore, the coached Nao received positive feedback through the social and interactive behavior it demonstrated while being able to effectively perform its task.

Each robot had a different task when they should have been the same to draw comparison of results. The users of the robots during the tasks should also should have been questioned, instead external participants who did not interact with the robots were requested to view a video of the robots performing their tasks, Grüneberg and Suzuki (2014) did not state their reasoning for this.

Additionally, length times of the videos were in favor of the coached Nao, leading to the perception this method was faster. Participants of the questionnaire were a majority female also potentially producing bias into the robot’s interaction perception. Research by De Greef and Belpaeme (2015) highlighted that female participants were more responsive to the robot whereas males were less receptive.

Incorrect data tables and methods were placed in the paper by Grüneberg and Suzuki (2014) and were addressed in a corrections paper. Despite this the results were similar to the incorrectly released data. Therefore, while the results of the experiments do show robotic learning through human interaction has been achieved the experimental process reduced the validity of the results gathered. Unbalanced tasks and lack of human-interaction feedback, therefore concludes the research presented by Grüneberg and Suzuki (2014) cannot be taken as solid evidence of advancements in this area.

2.2 Data Driven

Liu et. al. (2016) explores data-driven HRI by having the robot learn social behaviors from human-human interactions. The robot was placed in a mock shop scenario tasked with interacting to customers, after having observed the way in which the humans had interacted with one another.

For comparison a second robot and method were created labelled the “*without abstraction*” system. The “*without abstraction*” method does not use

clustering techniques for speech, motion, feature vector and prediction unlike the first method. A similar method was utilized by Admoni and Scassellati (2014) in their research however it has been adapted in this experiment to take verbal communication.

17 paid participants 11 male and 6 female were used for the experiment. Eight trials each were partaken between two methods, after the eight trials in one condition was completed a questionnaire was then given to participants followed by the testing of the next condition along with another questionnaire finally concluding on an interview.

From the concluded results Liu et. al. (2016) stated that the participants enjoyed the interactions, with the robot being able to communicate and move with the participant with very little errors. The evaluation of the robot's behaviors between conditions effectively supported the hypothesis that the behavior in the proposed system was better than the comparative system. The figure below demonstrates the participant's results.

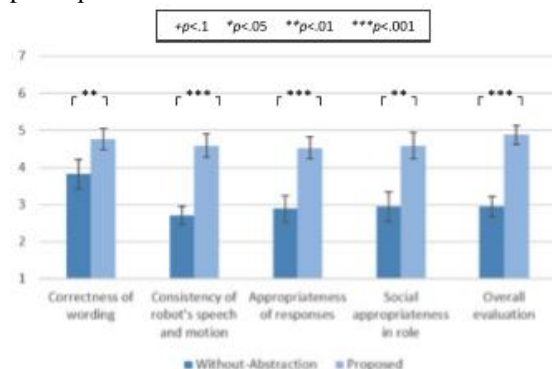


Figure 1: Evaluation Results of Robot Behaviors Between Conditions (Liu et. al. 2016)

The experiments allow for an even evaluation of both models with the participants having no bias towards either proposed condition, however the quick succession of the tests and questionnaires provides a potential misjudgment of each method. Following on from the tests and questionnaires is an in-depth interview for the participant. Extra time should have been allocated allowing for the testing to be spread over multiple days.

Despite minor flaws in the experimental testing the results gathered from participants effectively

demonstrate that the proposed data-driven model boosts human-robot interactions when compared to other data driven methods such as that of Admoni and Scassellati (2014).

Similarly, research by Keizer, et. al. (2014) utilizes the data driven approach to interact with multiple customers at once. The JAMES robot created through research by Foster et. al. (2012) was adapted for this research using Social State Recognizer (SSR) and a Social Skills Executor (SSE), essentially allowing for the robot to determine specific situations. Such as the situation shown below.



A customer attracts the bartender's attention
 ROBOT: [Looks at Customer 1] How can I help you?
 CUSTOMER 1: A pint of cider, please.
Another customer attracts the bartender's attention
 ROBOT: [Looks at Customer 2] One moment, please.
 ROBOT: [Serves Customer 1]
 ROBOT: [Looks at Customer 2]
 Thanks for waiting. How can I help you?
 CUSTOMER 2: I'd like a pint of beer.
 ROBOT: [Serves Customer 2]

Fig. 1. A socially aware robot bartender.

Figure 2: A Socially Aware Robot Bartender (Keizer et. al. 2014)

Through the proposed method JAMES could group customers into singles or groups and whether they wish to be served, then performing multiple transactions at once. Two types of SSR and SSE were tested one hand-crafted the other utilizing supervised learning.

A similar experiment as that of Foster et. al. (2012) was used to test JAMES. In their experiments the results concluded no customer that was seeking engagement were engaged in addition 104 of the 109 customers received a drink after a waiting time for the robot to pick up their position.

Building from that experiment using JAMES Keizer et. al. (2014) collected 37 subjects, resulting in 58 drink ordering interactions. 29 utilized the hand-coded SSE while 29 others used the trained strategy. 26 interactions utilized the rule-based classifier while 32 used the trained strategy.

Each SSR and SSE hard-coded and trained were compared, the results of the SSR showed the trained SSR had higher engagement changes number of 17.6 when compared to 12.0 therefore being more responsive. Additionally, preference for the trained SSE was shown. Keizer et. al. (2014) concluded their paper by stating their experiments confirm that data-driven techniques are suitable for human-robot interactions and further work into user behavior must be partaken.

Measure	Rule-Based SSR (sd)	Trained SSR (sd)
Detection rate	0.98 (0.10)	0.98 (0.09)
Time to first detection	5.4 (7.9)	4.0 (9.7)
Time to system response	7.0 (7.9)	6.4 (10.4)
Time to drink served	62.2 (22.2)	53.7 (14.0)
* Num. engagement changes	12.0 (10.2)	17.6 (7.6)

Table 1: Objective Results for The SSR Comparison (Keizer et.al. 2014)

Keizer et. al. (2014) stated their study was hindered by two aspects “all of the customers were explicitly instructed to seek engagement with the bartender,” and ground truth data on customers actual engagement-seeking behaviour was not available. Therefore, the results while demonstrating the trained method was greater than the rule based cannot be taken as a valid representation and this was noted by Keizer et. al. (2014) stating they are performing other evaluation classifiers to address these limitations.

2.3 Comparison of Data Driven Methods

Both methods proposed by Keizer et. al. (2014) and Liu et. al. (2016) have been applied to real-world scenarios. Either method present positive results however Keizer et. al. (2014) method allows for multiple customers to be served at once whereas Liu et. al. (2016) is one at a time. The robots themselves are very similar in modes of interaction and by incorporating the trained SSR and SSE from Keizer et. al. (2014) when combined with the method of and Liu et. al. (2016) could allow for a perfect method for data driven human-robot interactions.

The robots have been tested working solo and not cooperatively as a human-robot partnership. Additionally, the robots must watch other humans to learn how to interact, therefore they cannot be placed in a workplace without viewing other staff members. JAMES utilises visual ques such as body language and position to identify a potential interaction, grouping customers into singles or groups whereas Liu et. al. (2016) uses auditory ques to differ approaches of interactions, JAMES while serving multiple customers lacks diverse communication techniques.

Combining the two methods would allow for grouping and multiple conversations between humans and robots. Each varying on the personality types noted through Liu et. al’s. (2016) method, additionally the mobility provided through the robot model of Liu et. al (2016) would further enhance the work environments the robot could be placed in. Therefore, the data driven methods proposed both compliment the missing features of either method and once combined could result in a robust human-robot interaction method.

2.4 Adaptive Incremental Learning

Zhang et. al. (2015) focuses upon adaptive incremental learning through image recognition. The method will allow the robot to learn and categories images based upon human-robot interactions from a zero-knowledge beginning.

The method utilises an adaptive learning algorithm, Nadine (the robot) has zero-knowledge therefore when unlabeled images are shown to the Nadine vector-based visuals vectors will be used to detect the underlying semantics within the image. From this the Nadine can then create classes labelling the images to compare with new images when they are presented.

To test the method 2000 images were selected over 10 semantic categories, Average Precision (AP) is employed as a performance metric enabling for the evaluation of recognition results. Images from different categories will be shown typically 20 at a time. The user will then provide binary feedback learning Nadine how to categories images. The AP will be increased after each round, a figure below demonstrates the results.

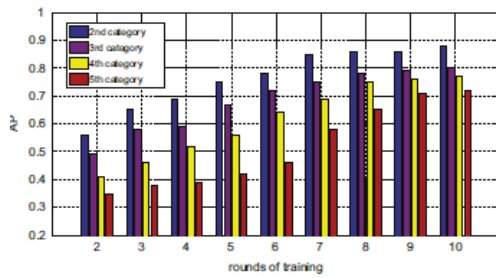


Figure 3: Adaptive Incremental Learning Results (Zhang et. al. 2015)

Results gathered in the first experiment demonstrate that the method works with Zhang et. al. (2015) stating “The first 6 rounds can be considered as the period of knowledge accumulation. Compared with the other three cases,”. Testing with other methods such as K-means, SVM, LDA and semi-supervised nonlinear learning method (SSNL) were carried out. The new method outperformed the others as demonstrated below.

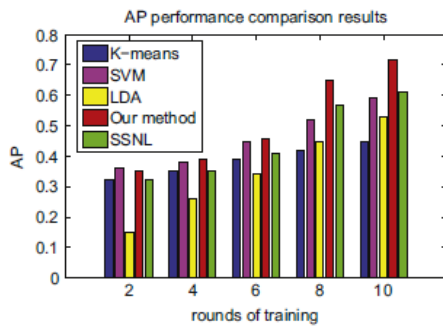


Figure 4: Comparison Results (Zhang et. al. 2015)

Finally, the system was evaluated with real users, 9 participants all from the Nanyang Technical University 6 male 3 females ranging from the ages of 23 – 32. The user’s interactions were gathered through a Godspeed questionnaire assessing anthropomorphism, animacy, likeability, perceived intelligence and safety along with a question on whether the robot could learn assessed through a Likert method.

Results gathered by Zhang et. al. (2015) demonstrate that the new method allowed for humans to teach robots as well as improving interactions between the two stating, “Overall the results of the questionnaire indicate that participants had a positive interaction.”. The

results gathered however through participant feedback and in comparison, tests of current methods reflect phenomenal improvements.

Zhang et. al. (2015) in conclusion state “Experimental results on the Nadine robot verify the feasibility and power of our algorithm.” they follow this with the research they have conducted on incremental learning and unlabeled images is significant.

Despite the positive results, the way in which the data was gathered during comparison of methods may be viewed as biased as there is no indication of the comparative tests used. All participants were of the university therefore being a convenience sample. Further limitations were to the small scale of the experiments and image sets; however, the research is easily replicable and with larger data sets could be improved upon. Therefore, the research despite a small scale is viable and will only be improved by a larger dataset of participants and images.

Further research by Gutiérrez et. al. (2017) implements Passive Learning Sensor Architecture (PLSA) allowing the robot to be able to learn an object through images, verbal communication and word semantics.

The experiment saw 5 tables inside an apartment have varied objects on them, such as table A which had hardware tools. The robot utilises a RGB-D camera, in the initial phase the robot taking photos of the tables and items. the robot was then tasked with using multimodal information to select 20 objects among the tables.



Picture 1: Demonstration of Experiment Layout Gutiérrez et. al. (2017)

The results were then compared with image segmentation and CNN image recognition systems. If the first object chosen in the query was correct it would be classed as success otherwise a failure. The test effectively demonstrates how PLSA outperforms leading CNN architectures. The semantic processing step which was also created not only improves PLSA but also the other CNN's that have been tested.

Method	Direct Match	Semantic Processing
PLSA	0.65	0.75
TH + GoogleNet	0.35	0.45
TH + GoogleNet2	0.35	0.45
TH + AlexNet	0.35	0.1
TH + VGG16	0.6	0.6
MCG + GoogleNet	0.5	0.5
MCG + AlexNet	0.15	0.25
MCG + ResNet	0.55	0.55
MCG + VGG16	0.55	0.45
CS + GoogleNet	0.45	0.5
CS + GoogleNet2	0.6	0.65
CS + AlexNet	0.2	0.25
CS + VGG16	0.45	0.45
R-CNN	0.4	0.0

Figure 5: Comparison of Methods Results (Gutiérrez et. al. 2017)

Gutiérrez et. al. (2017) when concluding their research states “It was demonstrated that it outperforms state-of-the-art algorithms” following from this Gutiérrez et. al. (2017) believes their devised method should be used as a firm candidate when allowing social robots to guess object locations.

While the results of the PLSA were significantly higher a variety of aspects during the experimental stage does hinder the results. Lack of external user participation in addition to timeframes of the experiments were omitted furthermore, the method in which the objects were queried were also precluded, such as verbal or hard coded. Additionally, no timeframe of task completion was provided, therefore no evidence as to how quick the method completed the query was provided. Despite this PLSA does significantly outperform other methods and with further experiments including external user testing and placement in a real-world environment could further solidify that PLSA is the best method for robotic object-location. The method also demonstrates that a robot can learn itself outperforming other learning methods currently available.

3 Conclusions

In this paper current research on human-robot interaction have been analyzed. The evaluation of the methods were derived through viability of results, performance through tasks and comparison of methods finally resting on usability of the proposed methods. From all the methods that have been analyzed data-driven research by Liu et. al. (2016) stands as a method which met all of the analyzed criteria.

Subjective computing research by Grüneberg and Suzuki (2014) while reflecting positive results was hindered through poor experiments. The results however were not as significant as other methods proposed. Research by Keizer et. al. (2014) in data-driven methods demonstrated that trained methods through human-human gazing far outperformed hardcoded interactions.

Research by Zhang et. al. (2015) and Gutiérrez et. al (2017) utilized adaptive incremental learning to allow the robot to learn and categories objects and images. Both of which reflected that robots can learn from a zero-base knowledge through human feedback.

The research covered in this paper presents the theory that robots learn better through interacting with humans both visually and auditory. Hard coded methods reflected lower results when in comparison to those utilizing human-robot interactions. JAMES, when gazing outperformed the hard-coded method, additionally Zhang et. al. (2015) reflected robots with no prior knowledge can quickly learn through feedback from humans. Therefore, no previous hard-coded knowledge would be needed for a specific environment as the robot can learn through their human partner how to perform their required role saving time and money whilst creating a human-robot partnership.

References

- Admoni, H. and Scassellati, B., 2016, “Nonverbal communication in socially assistive human-robot interaction.” *AI Matters*, 2(4), pp.9-10.
- Biswas, M., Murray, J., 2017, “The effects of cognitive biases and imperfectness in long-term robot-human interactions: Case studies using five cognitive biases on three robots”, *Cognitive Systems Research*, June Volume 43, pages 266-290.

- De Greef, J & Belpaeme, T n.d., , 2015, "Why Robots Should Be Social: Enhancing Machine Learning through Social Human-Robot Interaction", *Plos One*, 10, 9, Social Sciences Citation Index.
- Foster, M, E., Gaschler, A.,, Guilian, M., Isard, A., Pateraki, M., Petrick, R, P.A.,2012, "Two people walk into a bar: Dynamic multi-party social interaction with a robot agent", *In Proceedings of the 14th ACM International Conference On Multimodal Interaction (ICMT 12)*.
- Grüneberg, P and Suzuki, K. ,2014, "Corrections to "An Approach to Subjective Computing: A Robot That Learns From Interaction With Humans", " in *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 2, pp. 168-168, June 2014.
- Gruneberg, P. and Suzuki, K. ,2014. "An Approach to Subjective Computing: A Robot That Learns From Interaction With Humans". *IEEE Transactions on Autonomous Mental Development*, 6(1), pp.5-18.
- Gutiérrez, M., Manso, L., Pandya, H. and Núñez, P. , 2017, "A Passive Learning Sensor Architecture for Multimodal Image Labeling: An Application for Social Robots." *Sensors*, 17(2), p.353.
- H. Zhang and P. Wu, 2015, "Semi-supervised human-robot interactive image recognition algorithm," , *8th International Congress on Image and Signal Processing (CISP)*, Shenyang, 2015, pp. 995-999.
- Keizer, S., Ellen Foster, M., Wang, Z. and Lemon, O., 2014, "Machine Learning for Social Multiparty Human--Robot Interaction." *ACM Transactions on Interactive Intelligent Systems*, 4(3), pp.1-32.
- Liu, P., Glas, D., Kanda, T. and Ishiguro, H., 2016, "Data-Driven HRI: Learning Social Behaviors by Example From Human–Human Interaction." *IEEE Transactions on Robotics*, 32(4), pp.988-1008.
- Wiltshire, T, Warta, S, Barber, D, & Fiore, S., 2017, 'Enabling robotic social intelligence by engineering human social-cognitive mechanisms', *Cognitive Systems Research*, 43, Pages 190-207.
- Zhang, H., Wu, P., Beck, A., Zhang, Z. and Gao, X., 2016, "Adaptive incremental learning of image semantics with application to social robot". *Neurocomputing*, 173, pp.93-101.

Critical Evaluation of Current Power Management Methods Used in Mobile Devices

One Lekula

Abstract

The emergence of advanced mobile devices such as smartphones comes along with different applications that demands high power to work efficiently therefore this paper compares, analyze and evaluates different power management methods such as polling and pushing approach, hardware measurement and WANDA-CVD System architecture to help reduce energy consumed by mobile devices. Conclusions show that a combination of two of the methods would provide most valid and efficient technique that will help increase battery lifespan in mobile devices.

1 Introduction

Nowadays smartphones are mostly used as means of communication between friends and family while some are used in health facilities and in workplaces. For these devices to perform their designated task, certain applications, processors and other resources exists within these devices and they require enough power to run. (Cui et al. 2017).

Because mobile devices have become an important aspect in people's life, Damaševičius et.al. (2013) states that battery lifespan of these mobile devices becomes a constraint as users sometimes fail to complete their tasks due to low batteries on their devices. An increase in the services and communication capabilities that the mobile devices provide means an increase in the battery energy density.

According to Salehan and Negahban (2013) mobile devices can be used for Social Networking Services (SNS), Short Message Service (SMS) and connections like Wi-Fi and Bluetooth hotspots which demands different power rate to work. However, battery capacity grows at a slower rate which prevents mobile devices to support advanced mobile applications apart from the above mention ones. (Cui et. al. 2017).

Due to the above issues, researchers are motivated to develop efficient power management techniques that will help to manage power

consumed by mobile devices hence enabling smartphones battery power the ability to keep up with advance in technologies. (Trestian et. al. 2012).

This paper evaluates current research aimed at reducing power consumed by mobile devices using different power management methods. Methods such as Polling and pushing will be evaluated based on the experiments undertaken, outcomes of the experiments, critical evaluation of claims made by different researchers and conclusions reached.

2 Current Power Management Techniques

This section reviews three power management techniques that have been proposed by different researchers. It discusses how the method works, how valid the experiments are and implication of the results.

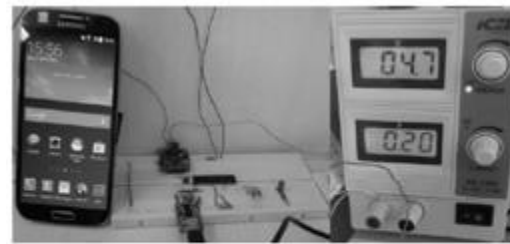
2.1 Pushing and Polling method

Increase in number of mobile devices has helped in identifying that batteries are vital in the use of these devices says (Abdelmotalib and Wu, 2012). Therefore, users are being frustrated by the lifespan of their devices as they discharge quickly and stop the determined use of features within the devices. In this research paper, Carvalho et al. (2014) proposes an analysis of energy consumption by comparing two main techniques being pushing and pulling method. These methods are used during data synchronization among

mobile apps and servers in the cloud to decrease energy consumed by mobile devices. Carvalho goes on to say that pushing technique occurs when the device directly requests and stay connected until the server sends data automatically or if any update is needed whereas polling has a device that frequently request a server update then disconnect.

The work conducted by Dihn and Boonkrong (2013) on comparing the two was done and the test were performed using Android OS using Powertutor software which estimates energy consumed by each method. The researcher confirms that pushing techniques is more efficient as compared to pulling. The claims made by Dihn and Boonkrong were confirmed by Carvalho et al. (2014) emphasizing that the pushing method can efficiently increase lifespan of battery on mobile devices. However, Dihn and Boonkrong (2013) works do not mention when it is appropriate to use polling and to what extend is pushing better than pulling method.

Carvalho et al. (2014) conducted an experiment based on the assumption that pushing method is more efficient as compared to pulling, A Samsung Galaxy IV smartphone with Android 4.3 Jelly Bean that uses 3G network from Claro provider. The experiment was to track world cup games by showing the game score and change in scores. Four components being the GCM server, game server and the two applications were used. It was done to show energy consumption measurement of the two applications.



The experimental environment



The application running

Figure 1 The experimental environment and the application running (Carvalho et al. 2014).

Applications Flow: the experiment was repeated 55 times for consistency and accuracy in statistical information with measuring of each application for an hour. However, the experiment does not provide justification for doing this for an hour and all those that consume battery were disabled during test for non-alteration measurement of apps. Once the application runs, games are loaded from the database to the view of the device on both methods. Then the polling application use one thread to make request for score update which update the database and what is seen on the device while pushing application there is connection to GCM server supported by Android service which receives data when there are any updates.

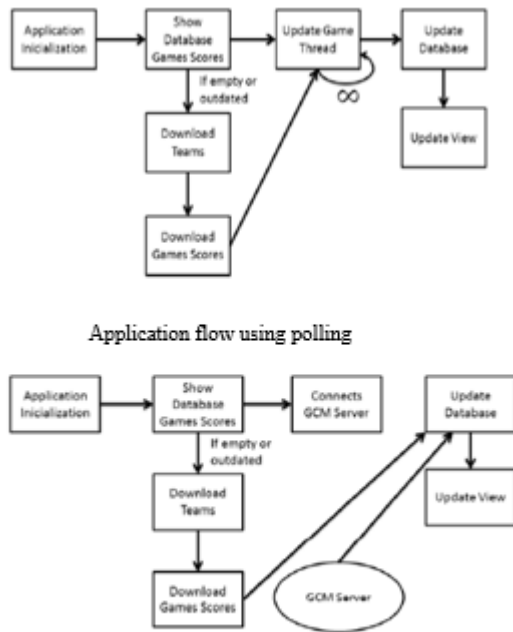
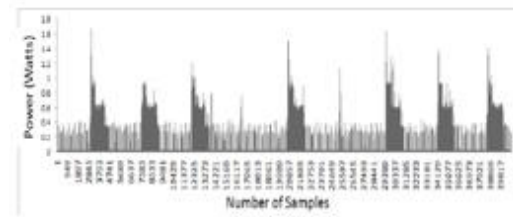
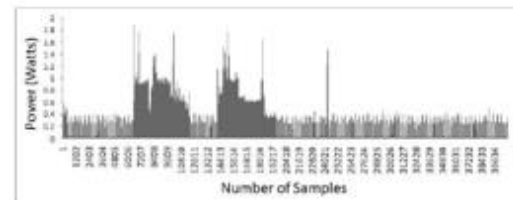


Figure 2 Application flow for polling and pushing (Carvalho et al. 2014)

Power consumption: Based on the results below, the research shows that when using the application in an interval of 5 minutes, polling method sent 7 requests as indicated by the peaks on the graph. Carvalho et al. (2014) emphasize that in an interval of 5 min the application can execute 1 request per minute and only have 5 peaks. Repetition of request is caused by network congestion leading to imprecise thread count time hence higher power consumption. Pushing method sent 3 peaks in an interval of 5 minutes therefore less power consumption.



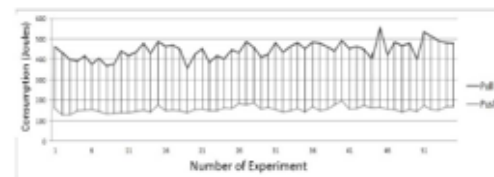
Consumed power using polling



Consumed power using pushing

Figure 3 Power consumed when for polling and pushing application (Carvalho et al. 2014)

Total Energy: Based on the results below, energy consumed by polling application is higher than energy consumed by pushing application.



Energy consumption of polling and pushing approaches

Figure 4 Energy consumed by Polling and Pushing application (Carvalho et al. 2014)

Gain Percentage: Pushing application has average gain of 187% against polling application.

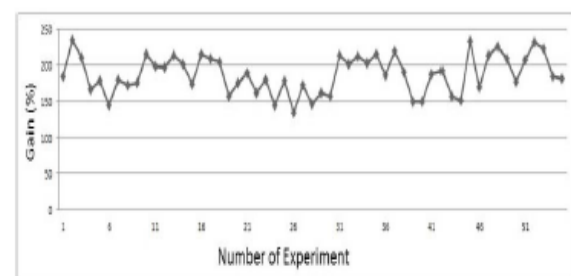


Figure 5 Gain percentage in pushing approach (Carvalho et al. 2014)

Request time: Energy analysis was done in request with different times. The test were done at an

interval of 5 min and results below show that polling approach has variations in time requests which becomes safe to use it if the application does not make more than one request in a range of 40 min or more for energy efficiency. While the pushing approach has a long term connection and its energy consumption is stable.

Request Time / Approach	Polling (arithmetic mean)	Standard Deviation
1 minute	447.86 J	41.18 J
5 minutes	298.94 J	13.12 J
10 minutes	280.63 J	14.19 J
15 minutes	245.76 J	13.94 J
20 minutes	236.35 J	37.77 J
30 minutes	183.38 J	15.76 J
40 minutes	159.38 J	11.02 J

Table 1 Energy consumption for different time request (Carvalho et al. 2014)

The claims about pushing techniques being efficient has been proved by Dihn and Boonkrong (2013) as per the test carried out using Powertutor software which estimates energy consumed by the pushing and polling approaches. Carvalho et al. (2014) did a positive review by stating the reason for analysing power consumption using the two approaches even though the polling approach was not said to be efficient due to the variations in the request time. Carvalho et al. (2014) was able to give a reason when to use the polling approach to avoid power consumption.

The approach used by Dihn and Boonkrong do indeed show that pushing approach is much efficient at any interval of 5 min as compared to polling approach which can only be used when application sends a request in 40min or less. However, the experiments done by Carvalho were only conducted on a Samsung device with an Android version connected using 3G network therefore the claims that pushing approach are not valid since the results do not show if the approach would perform the same thing on different device with different version such as IOS and it only work in 3G networks, it does not work over Wi-fi therefore not valid.

Results from Figure 2 shows that polling has a higher energy consumption as compared to pushing because the pushing approach sent 3 request in an interval of 5min while polling sent 7 request in an interval of 5 min but consumed more power therefore the claims made by Carvalho et al.

(2014) were proven to be consistent and accurate as displayed in the results of the experiment done.

2.2 WANDA-CVD System Architecture

According to Alshurafa et al. (2014), smartphones are used as a means of data collection, for measuring physical activity and giving feedback to the users. However, battery lifespan becomes a constraint and the researchers present WANDA-CVD architecture as a new optimization method which increase battery lifespan of smartphones used for monitoring physical activity. It also suspend the processing power until the nurses want information, or when the smartphone has been charged at night for enhancement of battery lifespan. Below is the diagram showing the WANDA_CVD system architecture which has a smartphone hub containing measuring, communicating and data collection from sensors smartphone application where data is then collected and analysed. However, this paper focuses at ways to optimize battery consumption for improved adherence.

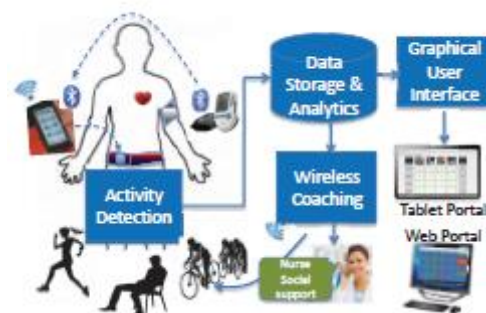


Figure 6 WANDA-CVD System Architecture (Alshurafa et al. 2014)

Based on the above description, the smartphone battery life time will manage to last longer because the method ensure that when the phone is not in use it will turn to sleep mode to reduce the accelerometer's sampling rate and for the phone to enter an initial state where accelerometer can be switched off if the phone is plugged to the charger. The formula below were used to calculate the adherence rate.

$$K_m = \sqrt{\frac{1}{n-1} \left[Q - \frac{1}{n} (P) \right]}, \text{ where}$$

$$Q = \sum_{i=0}^n x_i^2 + \sum_{i=0}^n y_i^2 + \sum_{i=0}^n z_i^2, \text{ and}$$

$$P = \left(\sum_{i=0}^n x_i \right)^2 + \left(\sum_{i=0}^n y_i \right)^2 + \left(\sum_{i=0}^n z_i \right)^2.$$

$$Com(i)_s^c = \begin{cases} 1 & \text{if } S \text{ compliant in week } i \text{ in category} \\ 0, & \text{otherwise} \end{cases}$$

$$Com_s^c = \frac{\sum_{i=1}^n Com(i)_s^c}{n},$$

$$TC_c = \frac{\sum_{j=1}^m Com_s^c}{m}.$$

Figure 7 Adherence Rate Formula (Alshurafa et al. 2014)

Based on the above formula a battery optimization was applied using the formula as shown below.

Procedure 1: Battery Management

We set a fixed window size W (5 seconds), a buffer rate B (seconds), and a K_m threshold value T . We calculate the current state $State_C \in \{Charge, Active, Inactive\}$, while adjusting accelerometer sampling rate f . Experimentally: $W = 5, B = 60$.

```

1: for  $i = 1$  to  $i = B$  do
2:   Calculate  $K_m$  from  $i$  to  $i + W$ 
3:    $i = i + W$ .
4: end for
5:  $K_m^A = \text{mean}(K_m)$ 
6: if  $K_m^A < T$  then
7:    $State_C = Inactive$ ;
8:    $f = 1$ ;
9: else if  $K_m^A \geq T$  then
10:   $State_C = Active$ ;
11:   $f = 10$ ;
12: end if
13: if  $Phone = Charging$  then
14:   $State_C = Charge$ ;
15:  Upload  $K_m$  values for previous day
16:   $f = 0$ ; Turn off accelerometer
17: end if

```

Figure 8 Battery optimization procedure (Alshurafa et al. 2014)

According to Alshurafa et al. (2014) they carried out an in-lab pilot experiment to test the smartphone applications with and without battery optimization. 7 participants to test the system without optimization for two months and with battery optimizations for the remaining 4 months. This system transfers users measured data when using different networks being the Wi-Fi and 3G/4G. A Motorola Droid RAZr Maxx with 3330 mAh Li Ion battery was used and participants went through a lesson of how to manage the smartphone throughout. Transferred batter usage events were

recorded when the phone was in use, not in use, charged or battery empty.

Alshurafa et al. (2014) state that the WANDA-CVD application was tested under four different conditions being the Wi-Fi mode only, Airplane mode, NG only and Wi-fi and NG enabled.

During this experiment participants had their smartphone on their pouch all day, doing daily activities and subject to irregular Wi-Fi and NG communication. The participants did not use the device features such as gaming and browsing through the internet. The author did not provide justification to this.

Based on the results of using this techniques for optimization of battery, there is an improvement in the lifespan of the battery. The results compares with optimization against those without optimization for the test carried out

The results below shows that when the device is on airplane mode without battery optimization it lasted for 35.2 hrs while it lasted for 71.6 hrs when on optimization state hence showing that the Architecture helped in reducing power consumed by smartphones at different modes. With optimization, users were able to achieve 160%, 400%, and 355% improvement in use of different modes. Most users managed to complete their day with the ease of charging at night.

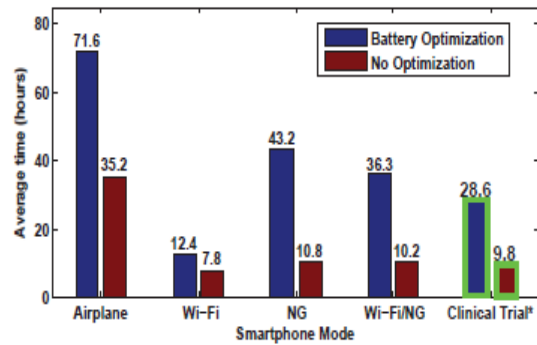


Figure 9 Battery lifespan improvements with optimization and without optimization (Alshurafa et al. 2014)

The main claims made on the WANDA-CVD system proves that optimization of mobile devices at different modes can help increase battery lifespan of mobile devices. Experiments carried out by Alshurafa et al. (2014) do prove that

optimization of mobile devices at different modes would help increase battery lifespan of devices.

The results from Figure 9 shows that there was a significant improvement in battery lifespan with optimization of battery in different modes. Comparison of how much energy was consumed at different modes was excellent as the result displayed at which mode is battery consumed more at what average time when there is no optimization and when there is optimization.

Alshurafa et al. (2014) experiment was consistent and valid since it was done using different modes over a period of 6 months using 7 participants. Therefore their method could be used as the results prove that optimization of mobile devices can help manage the battery lifespan on devices.

2.3 Hardware Measurement

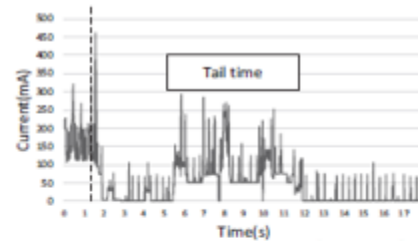
According to Wang et al. (2016), smartphones come along with great number of hardware mechanisms within them, but their battery lifespan decreases with time and forces users to recharge their phones every now and then. Researcher introduces number of methods that can help in saving energy on smartphones.

Hardware measurement is used to measure the limitations of a smartphone through runtime and external hardware. (Wang et al. 2016). According to Deng and Balakrishnana (2012) it consists of power meter which uses Monsoon Power monitor for measuring current value in various platforms that supplies a stable voltage to the smartphone and uses current value as representation of power consumption. It also consists Wi-Fi Traffic Monitor where the smartphone is used without any SIM card but rather connected to Wi-Fi to identify the cause of power consumption by Wi-Fi traffic in the current traces.

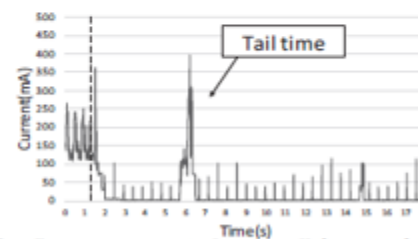
Wang et al. (2016) conducted an experiment based on identification of several features of energy consumed in different setting when on standby modes. The researchers used Google Nexus S smartphone to perform this experiment with certain applications installed for measuring power in different settings.

Based on the experiment two approaches were used for hardware measurement. The tail energy during screen switch off which displays the

measured current when the system switches off the screen automatically within a specified time without any operations and result show that power consumption of the device could not drop fast after the screen goes off.



Power measurement of turning off the screen by the system (Wi-Fi on)



Power measurement of turning off the screen by the system (Wi-Fi off).

Figure 10 Power measured from turning off the screen by the system when Wi-Fi is on and when Wi-Fi is off (Wang et al. 2016)

Wang et al. (2016)'s method was used to measure the energy consumed and failed to show how the energy consumed can be managed to increase battery lifespan of their devices. Their claims only shows the results of one mode being the Wi-Fi. The author failed to provide a solid justification of why the system would perform some optimization if the screen is turned off using the power button.

It also failed to give prove from the experiment why hardware component would consume substantial power during the standby mode even if they are not in use.

3 Recommendations

Use of pushing approach can be efficient during data synchronization on a 3G network with consideration of data size, network speed, and signal quality as it can send many requests in a short period of time hence conserving battery power.

On the other hand, WANDA-CVD method can be used for optimization of battery consumption for improved adherence. This method is much better than other as it yielded positive results and experiment was well explained. It was done at all modes of connection. Therefore, to a higher extent this method can be highly recommended as compared to the other two methods.

Hardware measurement can be best used for measuring the energy consumed at different network connections instead of reducing energy consumed by mobile devices.

4 Conclusions

To efficiently manage battery lifespan of mobile devices, different techniques can be used to help overcome this issue. Upon completion of evaluation the above methods, it has shown that the pushing approach by Carvalho et al. (2014) can be used for data synchronization in the cloud using 3G network since the application server can update database even if the foreground that sustains the application is disabled. It can also send desired request in a short period of time hence consuming less energy. This was supported by (Dihn and Boonkrong 2013).

WANDA-CVD method and Hardware measurement also proved to be efficient as they can be used to measure and optimize the battery lifespan of devices at different context.

However, a combination of WANDA-CVD and Hardware measurement could be more efficient for managing power consumption by mobile devices at different modes such as the Airplane, Wi-Fi, 3G and 3G/Wi-Fi combined.

References

Abdelmotalib A. and Wu Z., 2012, 'Power Management Techniques in Smartphones Operating Systems', *International Journal of Computer Science Issues*, Vol.9 (3), Pages. 157-160.

Alshurafa N., Eastwood J., Nyamathi S., Xu W., Liu J.J. and Sarrafzadeh M., 2014, 'Battery optimization for remote health monitoring system to enhance user adherence,' *In Proceedings of the 7th international conference of Pervasive Technologies Related to Assistive Environments*, Pages.8.

Carvalho S.A.L., de Lima R.N. and da Silva-Filho A.G., 2014, 'A pushing approach for data synchronization in cloud to reduce energy consumption in mobile devices', *In Computing Systems Engineering (SBESC)*, Brazilian Symposium, Pages 31-36.

Cui Y., Xiao S., Wang X., Lai Z., Li M. and Wang H., 2017, 'Performance-aware energy optimization on mobile devices in cellular network', *IEEE Transactions on mobile Computing*, Vol. 16, No. 3, pages 1073-1089.

Damaševičius R., Štuikys V. and Toldinas J., 2013, 'Methods for measurement of energy consumption in mobile devices', *Metrology and measurement systems*, Vol. 12, (3), Pages 419-430.

Deng S. and Balakrishnan H., 2012, 'Traffic aware techniques to reduce 3G/LTE wireless energy consumption,' *in proceedings of the 8th international conference on emerging networking experiments and technologies (CoNEXT '12)* ACM, New York, USA, Pages. 181-192.

Dinh P.C. and Boonkrong S., 2013, 'The Comparison of Impacts to Android Phone battery between Polling Data and Pushing Data', *International Conference on Computer Networks and Information Technology ICCNIT*, Bangkok, Thailand.

Salehan M. and Negahban A., 2013, 'Social networking on smartphones: when mobile phones become addictive', *Computers in Human Behavior*, Vol.29 (6), Pages 2632-2639.

Trestian R., Moldovan A.N., Ormond O. and Muntean G.M., 2012, 'Energy consumption analysis of video streaming to android mobile devices' *in Network Operations and Management Symposium (NOMS)*, Pages 444-452.

Wang C., Guo Y., Xu Y., Shen P. and Chen X., 2016, 'Standby Energy Analysis and Optimization for Smartphones', *In Mobile Cloud Computing Services and Engineering (MobileCloud)*, Pages 11-20.

A Critical Evaluation of Current Face Recognition Systems Research Aimed at Improving Accuracy for Class Attendance

Gladys B. Mogotsi

Abstract

In these few years, face recognition technology has come to be mature. In this paper, we will have a comparative study of three most recently methods for face recognition accuracy in class attendance. The approaches studied are Eigen Face, Fisher Face and Local binary pattern histogram. After the employment of the above three approaches, we learn and analyze the experiments and results of each algorithm then see the difficulties for the implementation. The contributions of this research paper are 1) FR approaches, 2) Comparisons 3) Conclusions.

1 Introduction

Face recognition (FR) is one most significant presentations of image analysis. For quite a long time, face recognition system (FRS) have attempted to defeat the deterrents in accomplishing higher recognition accuracy (Akhtar & Rattani 2017). FR has caused gigantic changes in areas where it has been employed. Accuracy in FR has been a problem from ancient times to this era. There are many factors that have affected face recognition accuracy (Shi, K et al. 2012), which are environmental features, algorithms and quality of image databases.

Additional factors include face shape, face texture, glasses, age, hair, and the elements that are unsteady, for example, lighting, and so on. Subsequently, any controllable elements ought to be controlled to consume a slight effect on the recognition system (Ling, H et al. 2007). Despite that, literature has shown that a lot of research has been carried out on face recognition. Phankokkruad, M et al. (2016), carried out research on class attendance system implementation. According to their findings it is often difficult to control student's facial expressions and some environmental elements. These mentioned factors had a high contribution on affecting the FR accuracy. The research studied few algorithms being Eigen faces, Fisher faces and LBPH.

Moreover, (Wagh, P et al. 2015) mentioned that, the previous face recognition based attendance system had few issues: intensity of light problem and head pose problem. In that attendance research, various techniques such as PCA, illumination invariant and Viola and Jones algorithms were brought in to overcome these problems.

Gross & Brajovic (2003) proposed Illumination Invariant algorithm for enhancing the light intensity and head pose problem. The thought in forming an illumination invariant is to post-process input picture information by forming a logarithm of an arrangement of chromaticity coordinates.

The research will have a good impact on the society as well as stakeholders since it aims at reducing the imprecise of FRS. This study paper will fundamentally assess the exploration as of now being done, concentrating on zones such as, factor variations and the proposed algorithms that have been probed.

The purpose of this paper is to evaluate, analyze and compare various researches on what is being prepared to solve the problem of FR accuracy. The researcher will evaluate existing research that has been carried on predicting accuracy in FR. Experiments done from past and current researches will be based on different FR areas, utilizing unique algorithms. The whole document includes three sessions being FR approaches, comparisons and conclusions.

2 FR Approaches

This section presents and evaluates FR techniques and the theoretical aspects of FR are discussed. The algorithms used in the experiment process are Fisher faces, Eigen Faces, and LBPH.

2.1 Eigen Faces Theories

According to Zhang and Turk (2008), Eigen Faces is generally based on the principle components analysis (PCA) of a distribution of faces. This method is a machine learning technique primarily utilized for reducing dimensionality of the feature vector space whilst retaining the main properties of data. The method of FR has s-dimensional vector faces within the training set of each PCA might be a T-dimensional vector space. This technique is another way to find set of weights from well-known face images. When the image is the vector of random variables it is defined as PCA eigenvectors of the diffusion matrix ST that is defined as:

$$S_T = \sum_{i=1}^M (x_i - \mu) (x_i - \mu)^T$$

Equation 1 Diffusion Matrix (Jain & Li 2011).

Where U are images within the training set. The matrix is encompassed of T eigenvectors and creating T-dimensional space face.

During face detection faces are cropped from image, therefore various factors such as distance between eyes, nose, outline of face etc. are then removed. With these faces as Eigen Features, students are recognized and through matching them with the face of database their attendance are marked. The image was captured at the same place in the light controlled environments (Phankokkruad & Jaturawat 2017). Below is a sample of database that was used for experiments. The student faces in the test set are for students that exist in the reference database, yet not a similar picture.





















Face Position	Expressions			
	Normal	Close eyelids	Smile	Grin
Frontal				
45 degree left				
45 degree right				
45 degree lower				
45 degree upper				

Table 1 Images of the face in controlled database (Phankokkruad & Jaturawat 2017).

The testing conducted utilized a closed test set of thirty students which have ten images for every student. There were twenty characteristics of images that combined four types of facial expression and five facial viewpoints. Within the frontal face position, the images were then collected with four unique expressions; a normal face, closed eyelids, smiling and grinning. The dataset chosen comprises of images that have unique variation in pose illumination, facial expressions and face position. Therefore, it would be a flawless dataset to give different situations for each subject. Moreover, Phankokkruad & Jaturawat (2017) says that this method takes time to collect images from each student and it is inconvenient for students to come at an exact time to take photos. Consequently, this method is unsuitable for a classroom that has numerous students.



Figure 1 Face expression variations (Phankokkruad & Jaturawat 2017).

Factors that were tested in the experiments were facial expressions, facial viewpoints and light exposure. Face expression is one of the factors that are difficult to control in the automatic FR system since whilst students pass by the camera their gesture will always vary. The results proved four types of face expressions: normal faces, close eyelids, smile, and grin as shown in Fig 1.



Figure 2 Face viewpoints variations (Phankokkruad & Jaturawat 2017).

The face viewpoints are the factors that are associated to student gesture. Unbalanced of face viewpoints is normally triggered by the movement of the student body. This viewpoint may influence the details of vectors that characterize the faces, thus that might trigger an error in FR. Outcomes demonstrates only five possible occurrences of face viewpoints that are frontal faces, tilted left, tilted right, looked up, and looked down. Fig 2 shows that.

Below is a table that shows how Eigen Faces tackled the issue of accuracy.

Algorithms	Total faces	Correct	Accuracy (%)
Eigenfaces		140	46.67

Table 2 Result of accuracy without confounding factors (Phankokkruad & Jaturawat 2017).

As depicted from table 2, Eigen Faces do not perform well with confounding factors. Therefore, the accuracy percentage is low by 46.67%. The test was conducted by making use of a closed test sample of 300 total faces of students. This is the testing with the non-adjustment image of the students in the test set.

Algorithms	Accuracy(%)		
	Normal	Smile	Grin
Eigenfaces	46.88	51.52	38.10

Table 3 Result of accuracy with variation of facial expression (Phankokkruad & Jaturawat 2017).

The above table shows that Eigen Faces does better with “smile” viewpoints at 51.52%. The lowest came with “grin” face viewpoint at 38.10%. This is because Eigen Faces recognition rates decreases under varying poses and illumination.

The experiment could have obtained higher accuracy but because face viewpoints (the looking down faces) has the greatest impact on FR accuracy, it was very difficult to obtain good results. Eigen Faces need unchanging background that may not be satisfied in most natural scenes of class attendance. Hence that is one reason the method did not give good results. This technique requires some preprocessing for scale normalization of which in this experiment it never happened. The most direct problem of utilizing this method is that it does not consider any face’s detailed aspects like face parts (eyes, nose, lips etc.).

During these experiments, there was never repeatability of experiments hence leading to bias of the experiments. Therefore, there is no credibility to this research paper since results might not be too genuine.

2.2 Fisher Faces Theories

A Fisher face is an algorithm with an argument in favor of employing linear methods for dimensionality reduction within FR issues (Phankokkruad & Jaturawat 2017). The learning set is labeled, it is sensible to use this information to build a more reliable method for decreasing the dimensionality of the feature space. Using linear methods for dimensionality discount may get improved recognition rates. However, results of several researches show that both algorithms have an effective processing time and storage usage (Phankokkruad & Jaturawat 2017). An example of a class specific method is fisher linear discriminant (FLD), since it attempts to “shape”

the scatter in mandate to make it more reliable for a classification. This technique picks w in such a way that the ratio of the between-class scatter and the within class scatter is maximized. The matrix for the between-class scatter is defined as:

$$S_w = \sum_{j=1}^C \sum_{i=1}^{n_j} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$$

Equation 2 (Shi, K et al. 2012).

$$S_b = \sum_{j=1}^C (\mu_j - \mu)(\mu_j - \mu)^T,$$

Equation 3 (Ling et al. 2007).

Fisher face is alike to Eigen face but with enhancement in better classification of different classes image (Jaiswal, S 2011).

Table 1 above under Eigen faces theories shows the database of faces for students. Within the frontal face position, the images were then collected with four unique expressions; a normal face, closed eyelids, smiling and grinning. Fisher Face was experimented using the same database for Eigen Face experiment with factors such as facial experiments, face viewpoints and light exposure.

Algorithms	Total faces	Correct	Accuracy (%)
Fisherfaces	300	208	69.33

Table 4 Result of accuracy without confounding factors (Phankokkrud & Jaturawat 2017).

The sample size of total faces used was still 300. As demonstrated in Table 4 above, Fisher Faces performs much better than Eigen Faces algorithm because Fisher Faces uses linear methods for dimensionality reduction within FR issues. Accuracy level is at 69.33% for this algorithm.

Algorithms	Accuracy(%)		
	Normal	Smile	Grin
Fisherfaces	71.88	66.67	57.14

Table 5 Result of FR accuracy with variation of facial expression (Phankokkrud & Jaturawat 2017).

From Table 5 it shows that Fisher Faces works well with normal facial expression and very poor with “grin” facial expressions. As for “smile” it gave a moderate percentage of 66.67%.

From the experiments results above, Fisher Faces has better accuracy percentages than Eigen Faces. This method instantly eliminates the initial three principle components accountable for light intensity changes. Fisherfaces method attempts to maximize the ratio of the between-class scatter versus the within-class scatter.

The experiment could have obtained higher percentages if some of the confounding factors and facial expressions were considered. Unbalanced of face viewpoints is normally triggered by the movement of student body. Moreover, this viewpoint may affect the details of the vectors that characterize the faces, thus that might trigger an error in FR.

2.3 Local Binary Pattern Histogram (LBPH) Theories

LBPH is the local feature based for face representation proposed by Ahonen et al. (2006). This technique is centered on local binary patterns (LBP). During the approach for texture classification, all existing codes of the LBP within an image are composed into a histogram. Classification is then implemented by computing simple histogram similarities. However, considering a similar approach for facial image representation results in a loss of spatial information, and so the texture information should be codified while holding their locations (Phankokkrud & Jaturawat 2017). LBPH has the benefit of invariant to light intensity yet it takes more time for processing rather than holistic approach. A histogram of the labelled image $f_i(x, y)$ can be defined as:

$$H_i = \sum_{x,y} I \{f_i(x, y)\} \quad i = 0, \dots, n - 1$$

Equation 4 (Phankokkruad & Jaturawat 2017).

Therefore, it is clarified that n is the number of unique labels produced by the LBP operator.

$$I \{A\} = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases}$$

Equation 5 (Phankokkruad & Jaturawat 2017).

This equation signify a histogram that was attained from the images holding information about local facial micro patterns together with face's edges, eyes, and location.

In the situation of class attendance checking system, usually the face expression and face viewpoints of the students are the factors variant and difficult to control. This operator is determined by comparing all pixels' values around the center pixel along with the center pixel value.

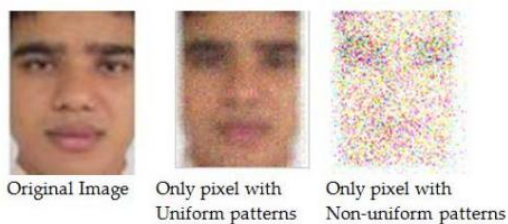


Figure 3 Face image split (Rahim, M.A., 2013).

From figure 3 is an example of how a student face would look like being taken through LBPH process. This image demonstrates an image which is divided in an image with only pixels with uniform patterns and in an image with only non-uniform patterns. It therefore shows that an image with pixels with uniform patterns contains better amount of pixels i.e. 99% of the original space.

In class attendance system using the 300 controlled database LBPH was proposed in the

experiments to determine its percentage accuracy with varying FR factors.

Algorithms	Total faces	Correct	Accuracy (%)
LBPH		245	81.67

Table 6 Result of accuracy without confounding factors (Phankokkruad & Jaturawat 2017).

LBPH appears to be the one having good results so far. Its accuracy is 81.67% with confounding factors. From the experiment, this method showed higher percentages. This is because the method has an advantage of invariant to light intensity, though it may take more time processing than the holistic approach.

Algorithms	Accuracy(%)		
	Normal	Smile	Grin
LBPH	79.69	80.30	80.95

Table 7 Result of FR accuracy with variation of facial expression (Phankokkruad & Jaturawat 2017).

As depicted from Table 7, the experiment was done based on three elements of variation of facial expression being Normal, Smile and Grin. Grin gave an output of 80.95% as the highest from them all. Followed by smile at 80.30% then lastly normal at 79.69%. LBPH is able to deal with variation of face expression with stable and high accuracy. In LBP, histograms are removed and concatenated into one feature vector. This feature vector is used to measure comparisons between images.

LBPH method gives great outcomes, both as far as speed and discrimination performance. The method appears to be robust against face images with dissimilar facial expressions, different lightening conditions, image rotation and grin. However, this method has its own limitations that make it not achieve 100% accuracy rate.

3 Evaluation and Comparisons of FR Algorithms

This section provides comparisons between the three algorithms that were experimented with a dataset of thirty student faces.

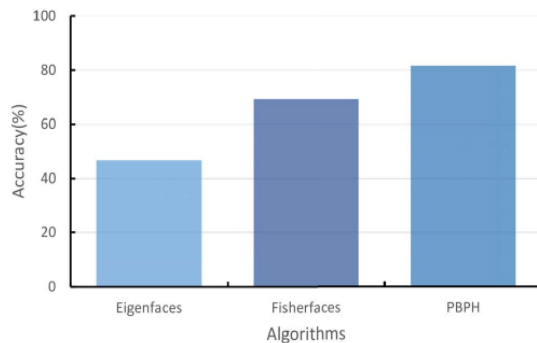


Figure 4 A comparative of still image FR accuracy without confounding factors (Phankokkrud & Jaturawat 2017).

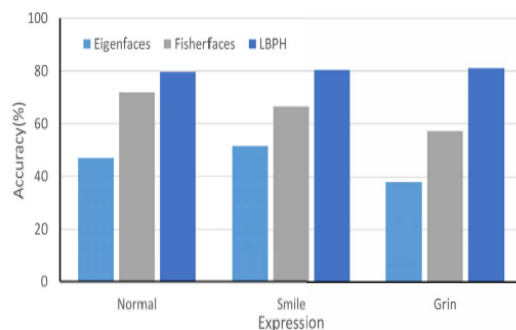


Figure 5 A comparative of still image FR accuracy with face expressions variations (Phankokkrud & Jaturawat 2017).

The researcher would like to make comparisons on the results of three algorithms (see Fig 4 & 5). Eigen Faces and Fisher Faces find space based on the common face features of the training set images. Both methods are quite similar as Fisher Face is a modified version of Eigen Face (Jaiswal, S 2011). In contrast to the previous algorithms, FR using LBP methods provides very good results both in terms of speed and discrimination performance (Rahim, M.A., 2013). The method turns to be vigorous against face images with unique facial expressions, different lightening conditions, image rotation and aging of persons.

The results shows that the performance varies significantly and LBPH has the best performance

in all areas experimented on. The trends of the accuracy from Fig 4 and 5 shows that LBPH method is followed by Fisher Face then Eigen Faces in the case of a small dataset.

4 Conclusions

FR is a personal identification technique that utilizes biometrics. In that case, FR has been chosen to be applied in class attendance checking system. Implementation of these FR systems is usually done at unique places in unconstrained environments, and so the work has studied the main factors that affect the FR accuracy. The researcher figured out from prior work that facial expression and face viewpoints are factors that affect the accuracy of the system. Furthermore, this study intends on comparing the facial recognition accuracy of the three chosen algorithms; Eigen faces, Fisher faces, and LBPH. Experiments that were conducted in respect of the facial expressions and face viewpoints variations were done in an actual classroom. Results of the experiments illustrated that LBPH got the highest accuracy of 81.67% in still-image-based testing and achieved 80.95% with variation of facial expression. A face expression that has the most impact on the accuracy is the “grin”, and face viewpoints that affect the accuracy are “looked down”, tilted left and right respectively. LBPH is considered the most appropriate algorithm for class attendance checking system after being picked among other algorithms.

Generally, the current research that was looked into was of a good standard, but unfortunately some of the factors affected different methods in each experiment. Hence, this lowered accuracy of some methods that were experimented. Especially Eigen Faces and Fisher Faces. Factors such as varying poses, illumination and face viewpoints had a bad impact on Eigen faces. Whereas, unbalanced viewpoints affected Fisher Faces.

5 Future Works

The approaches described in this paper are initially positive and promising in face recognition of class attendance.

It is obvious that the results of this face recognition system are perfect with LBPH method only. There is still a room for improvement for the future especially with Eigen Face and Fisher Face approaches.

Due to time constraints, the researcher was not able to look into more approaches of face recognition that might have better results than what was found.

Increment of database with illumination variation, pose variation, expression variation etc. conditions must be considered.

The current research study reports witnessed factors that affect FRS. The exploration did not attempt to explain cause of the effect in detail. Answering the cause will somehow assist in designing more algorithms that are robust.

Many problems have been faced with recognized face images from database. In the future to improve these issues, techniques can be combined to build a unified system for video-based face recognition (Rahim, M.A., 2013).

References

- Ahonen, T., Hadid, A. and Pietikainen, M., 2006. 'Face description with local binary patterns: Application to face recognition'. *IEEE transactions on pattern analysis and machine intelligence*, Vol28, pages. 2037-2041.
- Gross, R. and Brajovic, V., 2003, June. 'An image preprocessing algorithm for illumination invariant faces recognition'. *International Conference on Audio and Video Based Biometric Person Authentication*. In *AVBPA*, Vol. 3, pages. 10-18.
- Jain, A.K. and Li, S.Z., 2011. *Handbook of face recognition*. New York: springer.
- Jaiswal, S., 2011. 'Comparison between face recognition Algorithm-Eigen faces, fisher faces and elastic bunch graph matching'. *Journal of Global Research in Computer Science*, Vol2, pages.187-193.
- Ling, H., Soatto, S., Ramanathan, N. and Jacobs, D.W., 2007. 'A study of face recognition as people age'. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* pages. 1-8. IEEE.
- Phankokkrud, M., Jaturawat, P. and Pongmanawut, P., 2016. 'A real-time face recognition for class participation enrollment system over WebRTC'. In *Eighth International Conference on Digital Image Processing (ICDIP 2016)*, Vol10033, pages. 100330V. International Society for Optics and Photonics.
- Phankokkrud, M. and Jaturawat, P., 2017. 'Influence of facial expression and viewpoint variations on face recognition accuracy by different face recognition algorithms'. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017 18th IEEE/ACIS International Conference on* pages. 231-237. IEEE.
- Rahim, M.A., Azam, M.S., Hossain, N. and Islam, M.R., 2013. 'Face recognition using local binary patterns (LBP)'. *Global Journal of Computer Science and Technology*.
- Shi, K., Pang, S., and Yu, F., 2012. 'A real-time face detection and recognition system', *2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, April 2012, pages. 3074– 3077.
- Wagh, P., Thakare, R., Chaudhari, J. and Patil, S., 2015. 'Attendance system based on face recognition using Eigen face and PCA algorithms'. In *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on* pages. 303-308. IEEE.
- Zhang, S. and Turk, M., 2008. 'Eigen faces'. *Scholarpedia*, Vol3, pages. 4244.
- Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfeld, A., 2003. 'Face recognition: A literature survey'. *ACM computing surveys (CSUR)*, Vol35, pages.399-458.

Usability of E-commerce Website Based on Perceived Homepage Visual Aesthetics

Mercy Ochiel

Abstract

Homepage aesthetic appeal now plays a significant role in influencing user's first impression of website quality and subsequent user satisfaction. This paper critically evaluated web aesthetic literature to determine important visual design elements crucial to aesthetics design and effects of the elements on aesthetic perception. Some of the methods analysed are card sorting approach, aesthetic computational theory where design elements are extracted converted into vector features and eventually evaluated. In conclusion, recommendations are made on practical approaches and design factors that strongly influence webpage aesthetic appreciation.

1 Introduction

With the advancement of web technology and the impact of e-commerce, most businesses are now using a website not only as a marketing tool but to offer online services, such as e-retailing. Beside the importance of functionality, performance and information delivery, homepage visual design is now considered a significant factor in enhancing website usability.

Yang-Cheng Lin et. al. (2013) states that users' perception of aesthetic appeal is strongly influenced by user's first impression of the webpage. Therefore a homepage should represent a captivating visual design. In a study focusing on how effective manipulation of graphic and text influence webpage aesthetic.

In a study to determine web aesthetic patterns Shu-Hao Chang et. al. (2014) found that webpages perceived to be visually appealing influence positive behaviour in users, that ultimately lead to sales, and that user satisfaction is hugely affected by the perceived webpage aesthetic.

Djamasbi S et. al. (2014) Proposed a hypothesis that implementing main image on a homepage can contribute to improving visual appearance of the page. The study found that use of image to create visual hierarchy has strong correlation of how users evaluate aesthetic design.

Tanya Singh et. al. (2016) states that attractiveness is one of the contributing factors of usability, in their empirical study investigating key factors that determine website usability.

This paper evaluate existing web aesthetic studies focusing on design elements that are essential to aesthetic design and how these key elements influence aesthetic appreciation.

2 Web Aesthetic

Users' evaluation of aesthetic can be comparatively diverse due to the subjective nature of beauty. However, based various web aesthetic literature there appears to be common web aesthetic evaluation factors.

This section is divided into two parts (2.1) investigate design element considered essential to webpage aesthetic design (2.2) discuss the effects the elements have on aesthetic perception.

2.1 Elements that Determine Aesthetic Appreciation

Jiang Zhenhui et. al. (2016) proposed a hypothesis that users initial perception of quality of five design elements (i.e. unity, novelty, complexity, intensity and interactivity) subsequently influence their perception of quality of web aesthetic, web usability and positive attitude towards the website.

Conducting two studies, using qualitative approach to collect data from literature reviews, online source, web design forums, and website for web design competitions, design guideline websites, books and professional web designers. 41 participants with web design knowledge were used in the categorising, refinement and sorting of the data. In study-two 300 students evaluate design elements of ten websites.

Results shows (Figure1) that unity, novelty, complexity, intensity and interactivity are essential design elements in evaluation of web aesthetics. Novelty design leads with 0.34, intensity design 0.31, interactivity 0.16, unity 0.15 and complexity 0.13.

outlined ensuring data accuracy. There is evidence showing how validity of the five design elements were determined. Based on the evidence provide their claim is valid and the experiment is scientifically justified.

Weilin Liu et. al. (2016) used a similar approach to establish design elements considered crucial to aesthetic design and the elements absolute level.

14 users participated in one rounds of focus group discussion, determining elements they considered important to homepage aesthetic, and absolute level of the elements. In study two, 214 user tested effects of the elements on aesthetic perception.

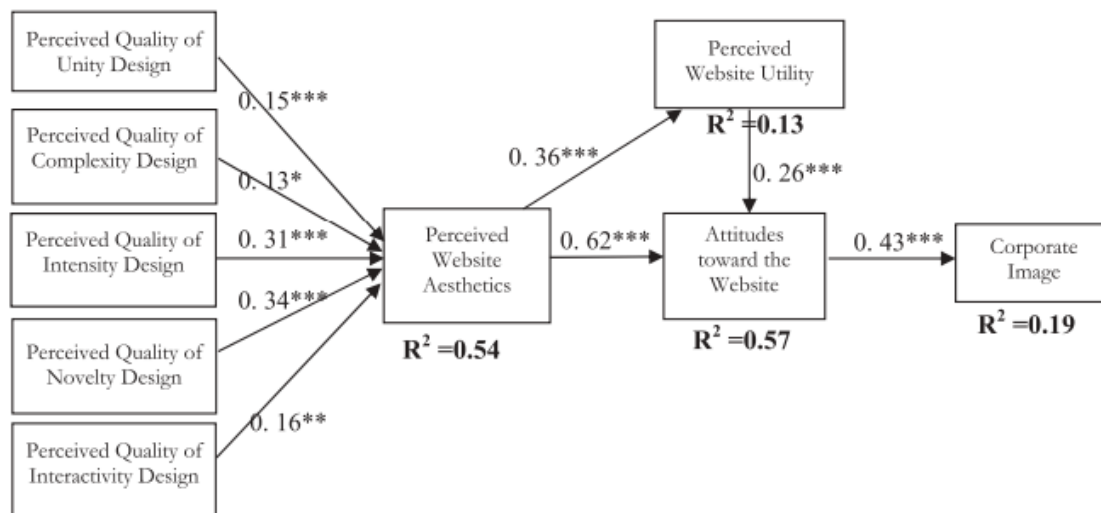


Figure 1 Research Model Testing Using PLS (Jiang Zhenhui et. al. 2016)

In conclusion it states that to enhance web aesthetic, novelty, interactivity, unity, complexity, and intensity should be jointly improved. And that user perception of aesthetic has stronger influence on user attitude towards a website than it has on website usability.

The study's data sample is comprehensively collected making the dataset diverse and of a wide range. Participants had no prior knowledge of the purpose of the experiment therefore eliminating bias. Although there is no mention of how study one participants were recruited, there appears to be relative equal gender ratio at every stage of the experiments. This approach is valid as other studies (Weilin Liu et. al. 2016) have used it. Data collection and processing procedure are clearly

Study-one mentions, layout style, body colour and presentation form to be top three design elements that influence homepage aesthetics. Results show colour to have no significant effect on aesthetic appreciation.

The study claims that the (Table1) design elements are most important design factors and that homepage aesthetic influence user satisfaction.

Independent variable	Dependent variable	F	p-Value
Body color	AF	1.240	0.291
	AA	1.241	0.291
Layout style	AF	7.151	0.001
	AA	0.996	0.371
Presentation form of advertisements	AA	0.028	0.973
	AF	8.834	0.000

Table 1 Test of Subject Effect (Weilin Liu et. al. 2016)

In contrast Jiang Zhenhui et. al. (2016) found variables of colour to highly influence users' perception of aesthetics while this study (Weilin Liu et. al. 2016) found body colour to have no significant influence on aesthetic appreciation.

Even though study-two results shows user's evaluation of the three design elements, study-one dataset was not comprehensively sourced with just one round of focus group discussion to determine the elements. Accuracy of the dataset cannot be verified with no evaluation criteria mentioned. Also there is no evidence to show how validity of the three elements and their various levels were determined arguably these make the study unrepeatable. There is not enough evidence to validate the claim and to scientifically justify the study.

Ou Wu et. al. (2016) proposed a new visual aesthetic assessment model where design elements were extracted, converted into a feature vector and evaluated. The proposed methodology implements multimodal features used in existing computation aesthetics.

They conducted two experiment, dataset-one consist of 1000 screenshot of homepages. Visual features are extracted using image processing technique, structural features are extracted based on structure mining of the page and functional feature are extracted from HTML source code. 10 students participated in the evaluation. The pages were categorised according to functionality using a soft-MT-fusion learning algorithm. Probability equation was used to check classification accuracy. Dataset-two consisted of 430 screenshot of webpages and was rated randomly by online user.

CLASSIFICATION ACCURACIES Acc (%) ACHIEVED BY THE COLOR HARMONY, TEXTUAL, AND GLOBAL VISUAL FEATURES INDEPENDENTLY FOR DS1 UNDER DIFFERENT CATEGORY GAPS WHEN BLOCK SIZE IS 24×24

Gap	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6
Color harmony features	70	71.00	74.61	75.43	74.66	78.42	80.95	80.90	79.01
Textual features	64.2	65.46	67.19	69.83	71.39	73.56	74.83	77.15	78.60
Global visual features	69.7	73.26	72.75	73.90	76.30	77.20	77.14	77.62	78.42

Table 2 Classification Accuracy of 24×24 Block Size (Ou Wu et. al. 2016)

Results rates colour harmony 0.70, textual features 0.65 and global visual features 0.73 (Table 2).

was perceived beauty. 270 homepage screenshot were rated by 267 participant in an online survey.

CLASSIFICATION ACCURACIES *Acc* (%) ACHIEVED BY THE COLOR HARMONY, TEXTUAL, AND GLOBAL VISUAL FEATURES INDEPENDENTLY ON DS2 UNDER DIFFERENT CATEGORY GAPS WHEN BLOCK SIZE IS 16×16

Gap	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6
Color harmony features	55.35	58.99	60.56	63.24	67.93	69.15	72.94	75.40	76.27
Textual features	51.90	52.17	55.14	59.51	63.22	65.60	69.90	70.87	71.74
Global visual features	55.74	57.59	60.85	62.84	65.43	68.01	72.45	73.90	75.94

Table 3 Classification Accuracy of 16 x 16 Block Size (Ou Wu et. al. 2016)

Dataset-2 average colour; harmony 0.79, textual features 0.75, and global visual features 0.82 (Table 3).

The study concluded that structural features, visual features and functional feature jointly influence user perception of aesthetics. And the new model effectively extract aesthetic design elements.

In comparison colour harmony's higher ratings can be considered similar to Jiang Zhenhui et. al. (2016) study which show intensity to strongly influence aesthetic perception.

In spite of the first survey following sound scientific procedure, it is highly possible that the results are flawed with cognitive biases by having 10 users test 1000 webpage, results might vary with reduced workload. With possibility of bias the reliability and accuracy of the dataset cannot be verified, hence validity of the claim is questionable.

Alexandre N Tuch et. al. (2012) proposed that visual complexity and prototypicality are design elements which influence user perception of webpage aesthetic.

Conducting two experiments, independent variable were visual complexity, prototypicality and presentation time, while dependant variable

59 undergraduate students tested the screenshots under a controlled experiment.

Study one results show that complexity high influence user perception of aesthetics. Highly complex websites were perceived to be less appealing. Websites with high prototypical were perceived to be more appealing.

In study two using similar procedure 80 page were evaluated by 82 participants

Results shows that even at 17ms webpage complexity high influence user perception of aesthetic (Table 4). While user perception for prototypicality is developed with longer exposure time (Table 5).

The study concludes that, visual complexity and prototypicality are important design factors that highly influence user perception of aesthetics on first impression. Users perceive websites with low complexity and high prototypical to be more attractive.

Controlled experiment participants had no visual or web design education limiting bias. The workload was sparsely divided. The results are based on user first impression as familiar webpages were omitted from data analysis. Every

Prototypicality	Visual complexity		
	Low	Medium	High
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
	50 ms		
Low	4.34 (.60)	4.37 (.66)	4.38 (.65)
High	5.52 (.79)	5.37 (.96)	4.59 (.83)
	500 ms		
Low	4.18 (.79)	3.92 (.82)	4.16 (.79)
High	5.57 (.69)	5.21 (.71)	4.16 (.85)
	1000 ms		
Low	3.91 (.62)	3.76 (.55)	3.95 (.60)
High	5.58 (.86)	5.13 (.70)	4.14 (.97)

Table 4 Visual Complexity and Prototypicality (Alexandre N Tuch et. al. 2012)

Main effects and interactions	<i>df</i>	<i>F</i>	η_p^2
Within-subject			
VC	1	85.60	.52*
PT	1	97.15	.55*
VC × PT	1	30.85	.28*
VC × time	2	1.46	.04
PT × time	2	8.83	.18*
VC × PT × time	2	2.45	.06
Error	79		
Between-subject			
Time	2	1.32	.03

Table 5 Effect of Complexity and Prototypicality (Alexandre N Tuch et. al. 2012)

stage of this study followed good science practices with several controls taken to limit bias. However the final claim that users find websites of low complexity and high prototypicality to be more visually appealing might vary with user familiarity with a web page.

2.2 Design Elements Effect on Aesthetic Perception.

Using subjective questionnaire approach Seckler Mirjam et. al. (2015) examined how structural and colour elements interrelates with various aspects of subjective aesthetic perception factors.

Conducting five online tests with various stimulus, using 25 homepage screenshot. Having variables of structural features symmetry and complexity and variables of colour hue, saturation and

brightness independently measured. 217 students participated in the survey. Using a version of Visual Aesthetic of Website Inventory to measure simplicity, colourfulness, diversity and craftsmanship.

The study result shows that

- Symmetry and complexity strongly influence simplicity and variety
- Both structural and colour factors influence complexity
- symmetrical interface were preferred by most user
- Less complex web pages received higher aesthetic appreciation
- Blue hue version of the webpages received higher rating while purple received the least

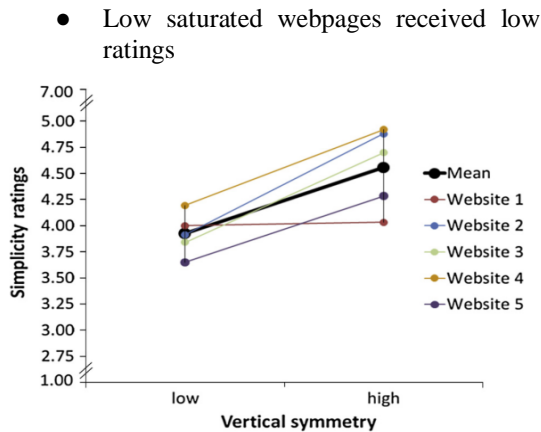


Figure 2 Simplicity Rating Based on Webpage Vertical Symmetry (Seckler Mir-

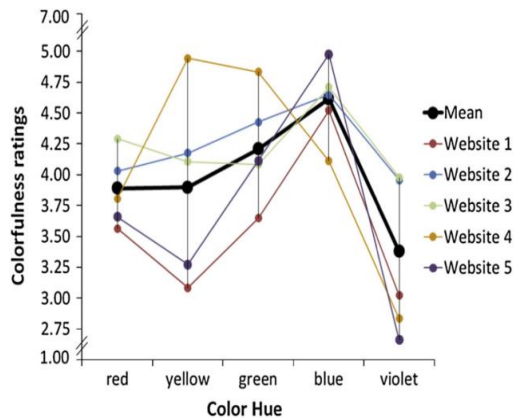


Figure 3 Colourfulness Rating Based on Webpage Colour Hue (Seckler Miriam et al

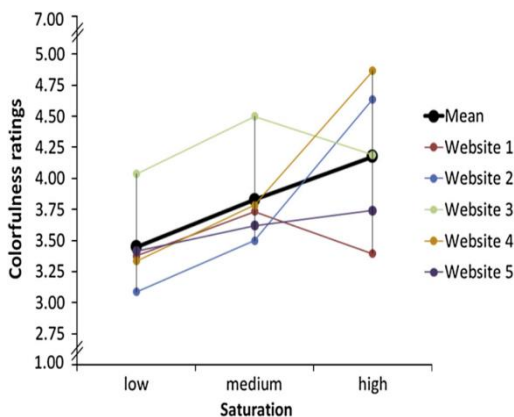


Figure 4 Colourfulness Rating Based on Saturation (Seckler Mirjam et. al. 2015)

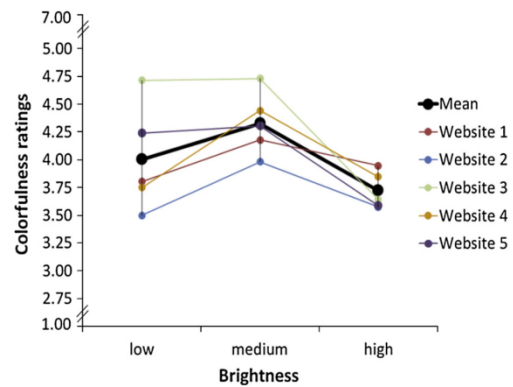


Figure 5 Colourfulness Rating Based Brightness (Seckler Mirjam et. al. 2015)

The study concludes that all the variables of structural and colour elements had a significant effect on subjective user perception of aesthetics. With different elements having different effects on aesthetic perception factors.

To ensure data reliability, the study omitted incomplete questionnaires from data analysis, questionnaires with colour impaired vision checked were also omitted from data analysis. The results show how the elements differently influence aesthetics so the study claim is justified, this study followed sound scientific procedure and provide enough evidence that scientifically validate the experiments conducted.

Ruben Post et. al. (2017) study proposed two hypotheses that unity and variety of a webpage has strong influence on user perception of aesthetics. And that manipulating and combining unity and variety creates the highest level of webpage aesthetic.

In the experiment a website designer developed 36 webpages with varying stimulus. Variables manipulated were Symmetry and colourfulness, contrast and dissimilarity. A total of 206 participants rated the designs at various stages of the study.

Result shows that:

- contrast influence both unity and variety
- Symmetry strongly influence unity
- pages with high contrast rated higher
- Unity rating increased with increase in colour and symmetry level
- Increased colour and symmetry had no influence on variety

Following the same procedure as study-one. Study-two and study-three were conducted to test validity of the results of study-one with new sets of webpages.

Study-two and study three results reaffirm study one results and the proposed hypothesis.

In conclusion the study states, both variety and unity significantly influence aesthetic perception. Effective manipulation of unity and variety rates high. Manipulating colour and symmetry can independently influence aesthetic appreciation.

To ensure dataset reliability the study omitted consecutive rating, where users rated all the sample equally. Even workload with every user evaluating nine webpages. The experiment was repeated three times with same results for study-

Mirjam et. al. (2015) shows that colour significantly influence complexity but not variety.

Liqiong Deng et. al. (2012) proposed complexity and order as two main important factors in aesthetic design.

They conducted a controlled experiment, 24 homepages with varying stimulus were designed and coloured prints used in testing. In study one 47 participants rated the level of aesthetic similarity of the homepages. In study two, 55 participants rated level complexity, order and preference of the homepages.

Results show, complexity with 0.933 and order 0.903. There is significant influence on aesthetic perception when order is manipulated at medium complexity and at low levels of complexity.

Multiple regression		1	2
	<i>F</i> value	88.248	52.150
	DF	2/21	2/21
	<i>R</i> square	0.894**	0.832**
Beta			
Dependent variables		Perceived Complexity	Perceived Order
Independent variables	Dimension 1 coordinates	-0.933**	0.311*
	Dimension 2 coordinates	-0.332**	0.903**

Table 6 Perceived Complexity and Order (Liqiong Deng et. al. 2012)

one and study-two, study three results validates the proposed hypothesis. This study has provided sufficient evidence to validate the claims, with limited possibility of bias in data collection and analysis process, with enough evidence and clearly outline methodology this study is repeatable and is scientifically justified.

In contrast Seckler Mirjam et. al. (2015) found symmetry to highly influence variety while this study found symmetry doesn't significantly influence variety. Both studies found colour to significantly influence aesthetic appreciation but in different aspects, this study shows that colour significantly influence variety while Seckler

Combining low level of order with high level of complexity received high preference. (Table 6)

The study claims order and complexity are important design factors in achieving web aesthetics. Webpage visual complexity positively correlates with aesthetic appreciation. Manipulation of high order and medium complexity strongly influence aesthetic appreciation.

The webpages had neutral content to limit confounding and preferential experience biases. Although the dataset is a precise reflection of regular e-commerce users the participants were

mainly students, the results might vary with different age range. With rigor at every stage of the study and adequate evidence provided the claim is valid and the experiment scientifically justified.

In comparison both Liqiong Deng et. al. (2012) and Ruben Post et. al. (2017) results shows

1) Both complexity and unity positively influence aesthetic elevation 2) manipulation of high level of complexity and order rated higher for aesthetic appreciation. 3) Ruben Post et. al. (2017) result show complexity influenced aesthetic appreciation more than order while only two participants in Liqiong Deng et. al. (2012) study had similar view.

Johanna M et. al. (2016) used appraisal theory of emotion to investigate correlation between visual elements and user emotional experience.

Data was collected using expressing, experience and emotion template. 50 Users expressed their perception through writing and drawing. Each student evaluated 2 webpages giving 100 data samples, the two webpages had same textual content with varying visual appearance. The image drawing were interpreted into words.

Result shows, balance was assessed by most users through symmetry, use of space, colour scheme, guiding gaze and grouping of elements were significant factors assessed.

In conclusion the study states unity, visual appearance perception and intelligibility of the design significantly contributes usability of the UI

Incomplete result and element with low frequency of mention were omitted from analysis. This study was conducted in guideline with good science practice, However, there could be one possible limitation, translation of the drawing could possibly confusing and inaccurate, we do not discredit the study based on this but would suggest use of computation theory for more accurate translation of the drawings.

3 Conclusions

With webpage aesthetic becoming a vital factor in website usability evaluation, it's important for web designer to know design elements that affect webpage aesthetics.

Ou Wu et. al. (2016) new visual aesthetic assessment model produced profound results, we recommend that for a comprehensive evaluation of the model, Jiang Zhenhui et. al. (2016) robust card sorting method to be implemented and integration with the method to replace the online user evaluation and to reduce the possible cognitive work load in the method used.

Ruben Post et. al. (2017) study used live websites and was robustly repeated three times with same results each time, in the aspect of realism we recommend Ruben Post et. al. (2017) pragmatic approach over Seckler Mirjam et. al. (2015) approach.

Based on the evidence of Jiang Zhenhui et. al. (2016) and Seckler Mirjam et. al. (2015) we recommend that colour is also a significant aesthetic design factor.

Based on the contrasting result when live websites and screenshots or printed images are used we recommend use of live websites for future studies, with exception of Ou Wu et. al. (2016) computation approach.

References

- Alexandre N Tuch, Eva E Presslauer, Markus Stöcklin, Klaus Opwis, Javier A Bargas-Avila, 2012, 'The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments', *International Journal of Human - Computer Studies*, Vol 70, Page 794-811
- Djamasbi S, Siegel M, Tullis T, 2014, 'Can Fixation on Main Images Predict Visual Appeal of Homepage', 2014 *47th Hawaii International Conference on System Sciences, System Sciences (HICSS)*, Page 371-375
- Jiang Zhenhui (Jack), Wang Wei-quan, Tan Bernard C.Y, Yu Jie, 2016, 'The Determinants and Impacts of Aesthetics in Users' First Interaction with Websites', *Journal of Management Information Systems*, Vol 33 Issue 1, Page 229-259
- Johanna M. Silvennoinen, Jussi P. P. Jokinen, 2016, 'Appraisals of Salient Visual Elements in

Web Page Design', *Advances in Human-Computer Interaction*, Vol 2016, Page 1-14

Liqiong Deng, Marshall Scott Poole, 2012, 'Aesthetic design of e-commerce web pages – Webpage Complexity, Order and preference', *Electronic Commerce Research and Applications*, Vol 11, Pages 420-440

Ou Wu, Haiqiang Zuo, Weiming Hu, Bing Li, 2016, 'Multimodal Web Aesthetics Assessment Based on Structural SVM and Multitask Fusion Learning', *Transaction on Multimedia*, Vol 18, Page 1062-1076

Ruben Post, Nguyen Tran, Hekkert Paul, 2017, 'Unity in Variety in website aesthetics: A systematic inquiry', *International Journal of Human - Computer Studies*, Vol 103 Page 48-62

Seckler Mirjam, Opwis Klaus, Alexandre N Tuch, 2015, 'Linking objective design factors with subjective aesthetics: An experimental study on how structure and color of websites affect the facets of

users' visual aesthetic perception', *Computers in Human Behavior*, Vol 49, Page 375-389

Shu-Hao Chang, Wen-Hai Chih, Dah-Kwei Liou, Lih-RuHwang, 2014, 'The influence of web aesthetics on customers' PAD', *Computers in Human Behavior*, Vol 36 Page 168-178

Tanya Singh, Sachin Malik, Darothi Sarkar, 2016, 'E-Commerce Website Quality Assessment based on Usability', *International Conference on Computing, Communication and Automation (ICCCA) 2016*, Page 101-105

Weilin Liu, Fu Guo, Guoquan Yea, Xiaoning Liang, 2016, 'How homepage aesthetic design influences users' satisfaction: Evidence from China', *Displays*, April 2016, Vol.42, Page 25- 35

Yang-Cheng Lin, Chung-Hsing Yeh, Chun-Chun Wei, 2013, 'How will the use of graphics affect visual aesthetics? A user-centered approach for web page design', *International Journal of Human - Computer Studies*, Vol 71(3) Pages 217-227

An Overview Investigation of Reducing the Impact of DDOS Attacks on Cloud Computing within Organisations

Jabed Rahman

Abstract

Due to the rising popularity of cloud computing it is a prime target for the hackers. This research paper is focused on the many DDoS detection systems that are available today and many that have been proposed on detecting DDoS attacks ranging from TCP flood, botnet attacks and also focusing on minimising downtime. In this research paper we will evaluate and analyse the DDoS detections system and provide an evaluation of the methods that were done so it can provide understanding of how reliable the detection methods are.

1 Introduction

Cloud computing is now more popular than ever. Despite having so many advantages it comes with the risk of many malicious attacks as it's a big market for the hackers. DDoS attack can shut down the server, so it is very important to detect the attack as soon as possible. "Security experts have been devoting great efforts for decades to address this issue, DDoS attacks continue to grow in frequency and have more impact recently" (B Wang et. al. 2015). The attacks on cloud computing are on the rise and its growing rapidly each year so it is very important to try to combat these security issues. "Over 33 percent of reported DDoS attacks in 2015 targeted cloud services, which makes the cloud a major attack target" (G Somani et. al. 2017).

From the example mentioned above it is clear that DDoS attacks are rapidly rising in cloud computing, however there is a lot of research that has been done regarding mitigation of DDoS attacks. These consists of using different methods of detecting DDoS attacks. An experiment carried out by P Shamsolmoali and M Zareapoor (2014) using NaiveBayes has detection accuracy above 96% with 0.5% false alarm rate. An experiment carried out by V Matta et. al. (2017) concludes that using botbuster for a network with 100 normal user and 100 bots the result is 90% of the bots are accurately identified. In this research paper we will be critically analyzing various methods of DDoS detection such as botnet attacks and TCP/IP flood attacks. The main focus is going to be the problems we are currently facing regarding DDoS

attacks and what kind of methods have been proposed to reduce the impact. Firstly, we will be analyzing current DDoS detection methods then we will be comparing against each other in order to find the best methods and have the conclusions of the most effective solutions to tackle against this problem.

2 Evaluation of the Current DDoS Detection Methods

This section of the paper will be used to evaluate current DDoS detection methods in place to stop attacks from happening.

2.1 TCP Flood Attacks

A Shahi et. al. (2017) proposed a new approach in regarding defends of DDoS tcp flood attacks called CS_DDoS. Classification based system insure security and availability of stored data. The incoming packets are classified to verify the behavior of the packet within a time frame. This is done to determine whether the source associated with the packets are attacker or are they actual clients.

The figure 1 is the architecture of the network testing. The testing was done by sending TCP ping then measuring respond time on average and recording results. Then they monitored filtered packet to see if it was genuine and attack was performed to test their method.

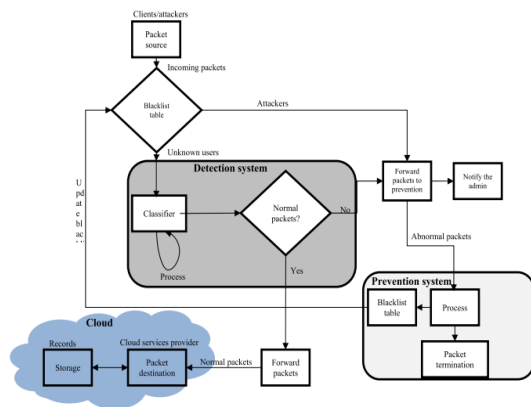


Figure 1 Architecture (A Shahi et. al. 2017)

The result of the experiment done by A Shahi et. al. (2017) CS_DDoS using LS-SVM. 97% of the time it can identify the attacker accurately during single attack with kappa coefficient of 0.89. The accuracy goes down to 94% when attacked by multiple sources and the kappa coefficient 0.9 (figure2).

	Classifiers	Accuracy	Detection results		
			Sensitivity	Specificity	Kappa coefficient
1	LS-SVM	99.5%	95.3%	96%	0.91
2	Naive Bayes	80%	92.3%	93%	0.82
3	K-nearest	75%	93.5%	95%	0.74
4	Multilayer perceptron	88.3%	95.3%	97%	0.78

Figure 2 Results (A Shahi et. al. 2017)

A Shahi et. al. (2017) concluded that DDoS attacks will always be an open research problem. In the future they would like to improve the CS_DDoS to overcome problems of spoof ID DDoS attacks.

The architecture of the method is well justified and it is very efficient as it has black lists of threats from previous attacks. None black listed attack get passed through the classifier. If the packet is considered to be abnormal it will be sent to the prevention system and the administrator will be alerted, then the attacking source will be black listed and terminated.

The experiment was done well and follows principles of good science as they measured time and accuracy of multiple methods and compared them against each other and found the most effective method to mitigate against DDoS attacks. The result they have backed up by experiments was done without any bias and in controlled environment. The experiment can be done again and it was consistent. They

acknowledge DDoS will always be an issue and considered future research on spoof ID DDoS attacks. (A Shahi et. al. 2017)

Another research carried out by Al-Hawawreh, M Sulieman on TCP SYN flood attack (2017) states detecting TCP SYN flood attacks are based on arrival of the packets which causes it to have many setbacks as delay in detection and high computational cost. Their work focuses on detecting SYN flood attack using anomaly detector to statistically characterise TCP/IP headers.

The experiment consisted of two scenarios: normal and attacking. In the normal scenario they used I macro script as bots in virtual machine two and three. Virtual machine four was used to browse and fill the form of the webserver presented in virtual machine one, and capturing the traffic for an hour using TCPDUMP tool. The attacking scenario virtual machine two and three were used to launch TCP SYN flood attacks. Same as the first scenario the traffic will be captured. This command was used to launch the attack "Hpin3 -S --Flood -V -p 80 10.0.2.4". (Al-Hawawreh, M Sulieman 2017)

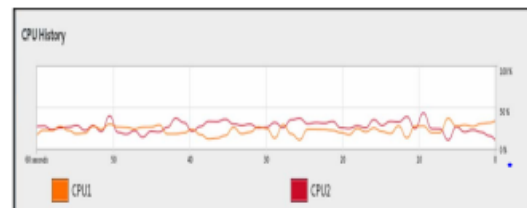


Figure 3 Before the Attack (Al-Hawawreh and M Sulieman 2017)

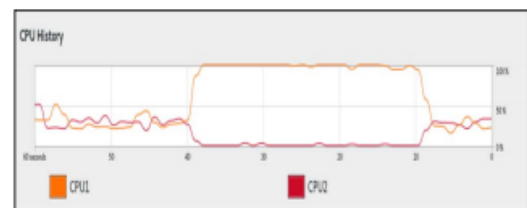


Figure 4 After the Attack (Al-Hawawreh and M Sulieman 2017)

The conclusion states that all the algorithms are proven highly effective as all four of the detection algorithms had accuracy of over 98% but further

modification should be considered using different machines and applying them in real life cloud environment with necessary modification.

The experiment is justified as it was performed in a controlled environment using many scenarios without bias towards the algorithms, and the results were measured to prove why their algorithms are effective. The experiment was explained well and they also made it clear more work has to be done before applying it in real cloud environment.

W Dou et. al. (2013) also looked at TCP attacks. Using CBF the method was divided into two periods; attack and non-attack. Non attack period will generate existing profile for packets then the number of appearance will be counted with confidence value being calculated to update nominal profile. Same was done in the attack period but stopped generating nominal profile and looks for flows in confidence values. TCP SYN flag was set to 40 for the length of the packet with other attribute being randomly selected. The result was 7.7% false positive and negative rate while the intensity of the attack was 5x, with doubling the intensity the false negative and positive rate stayed very similar. The author considers this as an effective filtering practice. The author concluded that using CBF can calculate incoming packets score during the attack period to conduct filtering and in the future more flexible discarding strategy is required. Better algorithms needs adapting to increase the speed and accuracy of CBF.

The claims made by the researcher is well justified as experiments were performed without any bias being involved and different types of attacks were used. The performance of the CBF method was evaluated to justify the effectiveness and stated to improve speed and accuracy they will need a better algorithm.

2.2 Botnet Attacks

R Kaur et. al. (2017) states that DDoS attacks are inspired by botnets. It's even more concerning that the attackers don't need to build the botnets themselves as it can be rented.

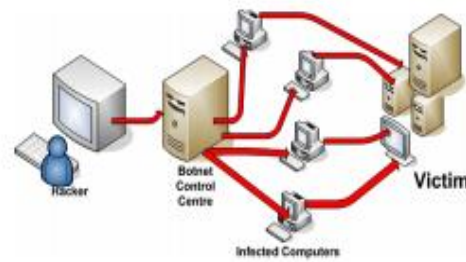


Figure 5 (R Kaur et. al. 2017)

In the figure five it shows how the hacker using botnets to launch an attack using high bandwidth using botnet instead of own machines.

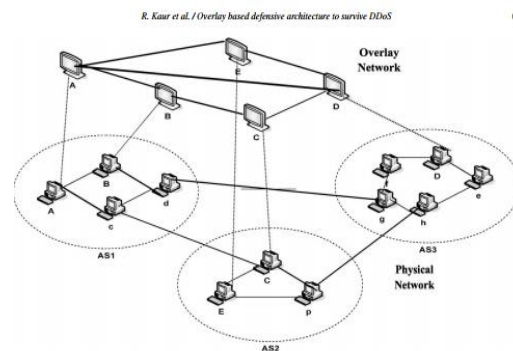


Figure 6 (R Kaur et. al. 2017)

R Kaur et. al. (2017) proposed overlay based defensive architecture to mitigate against DDoS (figure 6). Defensive perimeter around the end serves to be able to maintain sufficient amount of connectivity between the protected server and authorised clients. Proactive defence are used to defend against DDoS by having layers of security between the client and the victims to proactively defend against DDoS attacks. In reactive defence, soon as DDoS get detected it's vital to restore network connectivity.

The research done by R Kaur et. al. (2017) covered a lot of different types of attacks the architecture can do to help mitigate the attacks. Lists of advantages and disadvantages are also covered. It shows that no bias is involved. A drawback in this research is they haven't actually performed an experiment using their architecture; they just proposed it and stated how it will help mitigate DDoS attack, therefore the claims made by the researcher isn't justified by evidence. To make this

is more justified they would have to perform an attack and measure the results.

A Sadeghian and M Zamani (2014) have proposed a black hole filtering model to locate triggers within the ISP. This helps detect the packet inside and drop the malicious packet.

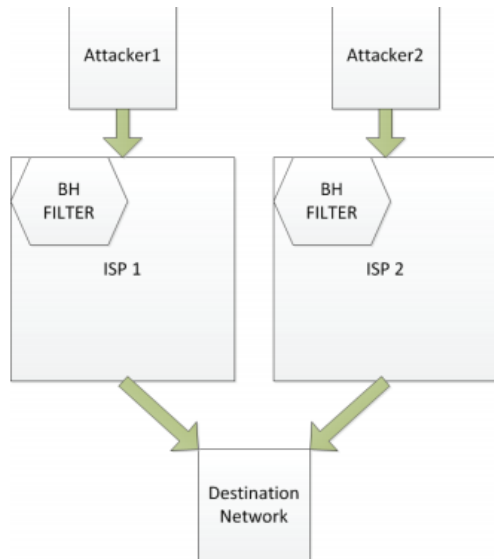


Figure 7 Self-triggered Black Hole Filtering Model (A Sadeghian and M Zamani 2014)

The BH filters each ISP and will load the traffic related to its owner therefore handling will be efficient. Some limitation applies using this model such as cost and coordination.

The author reached the conclusion that trigger might be ineffective with high amount of botnets attack therefore self-triggered black hole filtering was proposed as it is more closer to the attacker's computer and malicious packets will be automatically detected and dropped by being sent to null interface.

The proposed method could be effective in the fight against DDoS but the author hasn't done any experiment to prove how good their method could be. To improve their work they would need to actually imply the method in a controlled environment then measure the results. There were some positives as they went into detail about the method and also put in couple of disadvantages to remove bias. Also they stated it can be costly so other cheaper options can be used instead.

2.3 Minimising Down Time

G Somani et. al. (2016) states that it is highly important to detect the attack quick and mitigate with minimum down time and being aware of budget and sustainability.

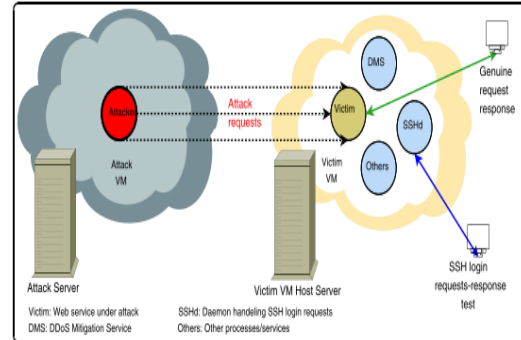


Figure 8 Experiment Set Up to Analyse DDoS Attack (G Somani et. al. 2016)

The experiment set up by G Somani et. al. (2016) had two virtual machine victims and attacker and having a co-located service on the same virtual machine operating system. They used connection count based attack filtering service called DDoS-Deflate. The major motivation for this attack is to measure the service down time, effect on other service and the detection time. The co-locate service was used for evaluating impact of the attack. The virtual machine sends SSH request to victim's server and if the session is granted immediately it logs out from session. The test is done for 500 SSH login-logout cycle during the periods of attack. To check if the target machine is available, genuine requests get sent for 100 times during attack period. Then attack was launched by sending 500 attack requests. The attack traffic, SSH traffic and genuine traffic was sent at the same time.

Table 2
Various Metrics: DDoS attack experimental study.

Attack reporting time	Victim service unavailability time	SSH unavailability time	Maximum response time (SSH)	Minimum response time (SSH)	Average response time (SSH)
39 s	945 s	517 s	186.930 s	0.129 s	1.226 s

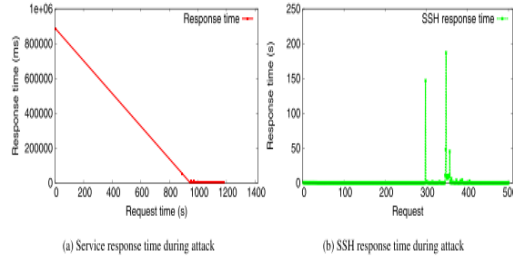


Figure 9 Result (G Somani et. al. 2016)

The conclusion reached by G Somani et. al. (2016) states that it can detect the attack source based on its policies after 39 s of attack being launched. The service becomes unavailable for 945 s. The graphs in figure 9 shows each request served one after the other. After the attack gets detected, mitigation adds rules to firewall and drops all the TCP connection involving the attacker.

The experiment was performed well in a controlled environment and no bias argument was used. Test was carried out on a large amount of SHS login-logout session which gives more accurate results. It shows that they have been successful and the down time is only 945 s.

N Hoque et. al. (2017) proposed a frame work for to detect DDoS attack in real time. The frame work consists of three major components: pre-processor, hardware module to detect attacks and security manager. The attack detection module receives traffic from pre-processor and also receives the threshold value from profile database. The detection system first calculates Nahidverc between input traffic instance and normal profile. Calculated value is compared with threshold to decide if it is classified as an attack. The detection is done based on deviation. It detects attack when deviation is larger than threshold value. The result are in the table below.

Table 4
Comparison with other methods on CAIDA dataset.

Method	DR	FPR	FNR
NFBoost [25]	97.2%	8.1%	NA
SNMP-based method [26]	99.4%	1.80%	0.60%
Proposed method	99.95%	0%	0.004%

Table 5
Comparison with other methods on DARPA 2000 dataset.

Method	DR	FPR	FNR
RBFBoost [27]	99.4%	3.7%	NA
Optimized traffic matrix [28]	98.7%	NA	NA
Chaos theory [29]	94%	0.45%	NA
Proposed method	100% (max)	0%	0.18%

Figure 10 (N Hoque et. al. 2017)

The conclusion reached by the author states that their system is able to achieve 100% accuracy over benchmark datasets. In the future they want to work on detecting crossfire attack in less time.

The work done by the author is done well as different methods of attacks and defence systems have been used and been compared against each other. The have also managed to achieve 100% accuracy on benchmark dataset which can be used to help improve accuracy of other systems in the future.

Additionally A. Saied et. al. (2016) states that their work is to detect and mitigate DDoS attack before it reaches the victim. Three types of attacks were selected due to their popularity: TCP, UDP and ICMP DDoS attacks. First they studied how the attacker build their approach. They reviewed related academic DDoS mechanism. Then physical environment was built to perform the experiment, then they launched three different attacks. They have launched 580 known and 580 unknown attacks. 100% known attacks were detected but 95% of unknown attacks were detected. The conclusion states that by using their method the result was 98% which is higher than other algorithms mentioned in the research but acknowledging some limitation in their work.

The work done by A. Saied et. al. (2016) was performed to a high standard. The experiments were conducted after analysing current work to give them more understanding about how to conduct their experiment. They used the same amount of known and unknown attacks to give them accurate results. Their conclusion admit some limitation and promoting to research on DDoS attacks.

3 Comparison of Current DDoS Detection Methods

The research done by A Shahi et. al. (2017) and R Kaur et. al. (2017) has some similarity; both methods have layers of security. A Shahi et. al. (2017) architecture terminates and black lists the attacking packet to prevent the same attack from the future whereas R Kaur et. al. (2017) layers of security will likely be preventing attacks from happening but it is more efficient to black list for future attack prevention. The major difference is A Shahi et. al. (2017) has actually done an experiment which backs up the claim, the same claims can't be done by R Kaur et. al. (2017) therefore evidence shows that A Shahi et. al. work is more proven and will work better.

For minimizing downtime N Hoque et. al. (2017) had proposed a framework to detect attack in real time. A. Saied et. al. (2016) reviewed academic journal then built their own physical system. Experiment done by N Hoque et. al. (2017) is more flexible as they have tested with many different types of attack and measured the accuracy for the attacks. Whereas, A. Saied et. al. (2016) only measured the defensive system on three popular attacks based on known and unknown attacks. Based on the results and the flexibility, the framework done by N Hoque et. al. (2017) is the better way to minimise downtime as it is a more flexible defensive system, but if it was for an organisation that are having issues with these three types of attacks it would be better for them to use A. Saied et. al. (2016) system as they have researched and identified problems before building the algorithm. The results were pretty similar. The results by A. Saied et. al. (2016) was 95% for unknown attacks detected and 100% of the known attacks detected. While N Hoque et. al. (2017) had results of over 94% accuracy and 100% accuracy for benchmark dataset.

4 Conclusions

N Hoque et. al. (2017) method of using framework to detect DDoS attack was proved to be very effective as the result of the experiment was very high. They even managed to get 100% accuracy on bench mark dataset. This can be used to improve other detection systems.

While the methods used in this paper are quite good, A Shahi et. al. (2017) method of CS_DDOS

shows promise of solving the DDoS problems we are facing. It also stands out the most as it is the most detailed and proven to be most effective against detecting DDoS attacks as it was done very well and against different methods, and they were compared against each other and it is backed up very well by experiments.

The research that has been analysed and evaluated in this paper have mostly good claims and the results do match with their claims. Although the methods conducted above do not necessarily solve the on-going issues with DDoS and some only have presented the theory but conducted no experiment, it can still be used by future researchers to perform experiments on them which could lead to having a very good detection method as it does show promising results. With help of the current system we can continue the on-going work of development of better defensive algorithms to fight against DDoS attacks. Hence further research is still needed in this area.

References

- Al-Hawawreh, M Sulieman 2017, 'Detecting TCP SYN Flood Attack in the Cloud', *8th International Conference on Information Technology*, page 236-243.
- A Sadeghian, M Zamani, 2014, 'Detecting and preventing DDoS attacks in botnets by the help of self-triggered black holes', *Asia-Pacific Conference on Computer Aided System Engineering*, page 38-42
- A Sahi, D Lai, Y Li, M Diykh 2017. 'An Efficient DDoS TCP Flood Attack Detection and Prevention System in a Cloud Environment', *IEEE Access, IEEE*. 5, page 6036-6048.
- A Saied, E Overill, T Radzik 2016. 'Detection of known and unknown DDoS attacks using Artificial Neural Networks', *Neurocomputing*, 172, page 385-393.
- B Wang, Y Zheng, W Lou, Y Hou 2015. 'DDoS attack protection in the era of cloud computing and Software-Defined Networking', *Computer Networks* 81, page 308-319
- G Somani, M S Gaur., D Sanghi, M Conti, M Rajarajan, R Buyya, 2017, 'Combating DDoS Attacks in the Cloud: Requirements, Trends, and

Future Directions.’ *IEEE Cloud Computing*, 4(1), page 22-32.

G Somani, M S Gaur., D Sanghi, M Conti, M Rajarajan, R Buyya 2016. ‘DDoS victim service containment to minimize the internal collateral damages in cloud computing’, *Computers and Electrical Engineering* 59, page 165-179

N Hoque, H Kashyap, D.K Bhattacharyya 2017 ‘Real-time DDoS attack detection using FPGA’, *in computing communications 110*, page 48-58

P Shamsolmoali, M Zareapoor, 2014, ‘Statistical-based filtering system against DDOS attacks in cloud computing’, *International Conference on Advances in Computing, Communications & Informatics 2014*, page 1234-1239,

R Kaur, AL Sangal, K Kumar, 2017, ‘Overlay based defensive architecture to survive DDoS: A comparative study’, *Journal of High Speed Networks*, 23(1), page 67-91

V Matta, M Di Mario, M Longo 2017, ‘DDoS Attacks with Randomized Traffic Innovation: Botnet Identification Challenges and Strategies’, *EEE Transactions on Information Forensics & Security Vol. 12 Issue 8*, page1844-1859,

W Dou, Q Chen, J Chen 2013 ‘confidence-based filtering method for DDoS attack defense in cloud’ *future generation computer systems* 29(7),page1838-1850

Critical Analysis of Online Verification Techniques in Internet Banking Transactions

Fredrick Tshane

Abstract

Lack of consumer trust has been an impediment to the use of online banking as user's fear losing money due to some fraudulent activities, and this affects their overall performance and daily operations of financial institutions. Nonetheless, diverse security measures have been developed to prevent this altering fraudulent technique. This research paper analyses and evaluates verification techniques that are used to prevent fraud in online banking transactions. It presents the use of one-time identity password, biometric finger print, online verification signature, and Kerberos authentication.

1 Introduction

“With the current expanding internet driven services, the investment in online channels represents a strategic choice for nowadays banks” (Yadav 2015). Online banking is one of highly embraced services as it allows easy fund transfer, e-commerce and continuous access to cash information. However, despite the adoption of online banking by financial institutions, users are still hesitant to use this service because of security issues like fraud which Philip and Bharadi (2016) stated that are because of compromised weak authentications and lack of internal controls.

Additionally, it has also been observed that the fraudulent on-line activities have not only been a source of concern to users but also to the financial institutions as they have led to massive losses due to practices like phishing. Therefore, failure to “effectively and efficiently detect Internet banking fraud is regarded as a major challenge to banks at large, and this is an increasing cause for concern. The use of biometric based authentication and identification can help in addressing these security and privacy issues.

Research has been done to address the above online banking security issues. Chadha et al. (2013) recommends the use of online signature verification technique in internet banking transactions. Gandhi et al. (2014) suggests a technique that prevents replay attacks and increase in security, Nwogu (2015) suggests a security measure that combines identity –based and mediated cryptography and

Tassabehji and Kamala (2012) suggests a biometric finger print technique. The revealed solutions above address the issues of privacy and security in internet banking transactions.

The scope of this research paper will be based on 4 online verification techniques which has been suggested by distinct researches and it will be organized as follows; Introduction in section 1, online verification techniques will be introduced in section 2 and section 3 which will be the last section, will be conclusion on the above-mentioned techniques.

2 Online verification Techniques

In this section online verification techniques used for online banking will be analyzed and evaluated. This evaluation will be based on the online banking methods, the tests and results obtained by different researchers.

2.1 One-time identity password

Gandhi et al. (2014) conducted a research on one-time password (OTP) that uses QR-code and authenticate with authorized certificates.

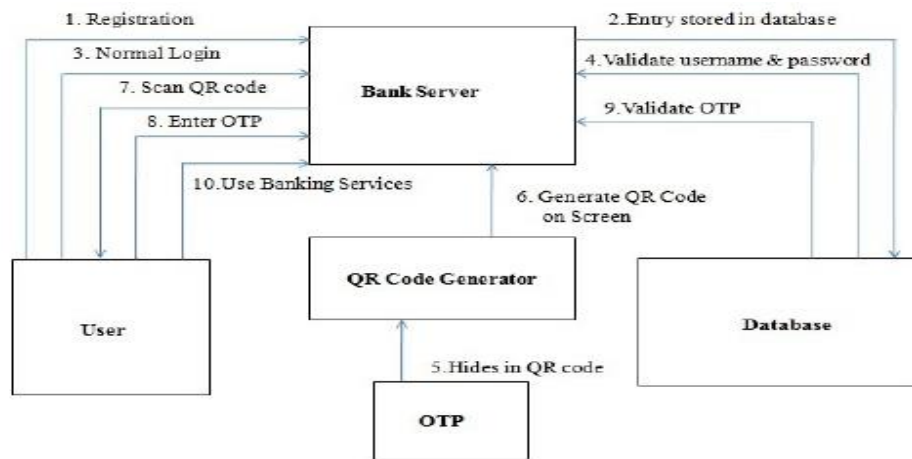


Figure 1 Working of Authentication System (Gandhi et al., 2014)

Gandhi et al. (2014) explained that the user must firstly register and create an account followed by a login session where the user will provide their authentication details, being the username and password. After providing this information, an OTP will be generated using the Customer ID, random number (RN) and the current system time and it will be hidden in the QR code image. Gandhi et al (2014) expressed that QR-code which is made of the OTP1, size and format will be displayed on the bank server. In the bank server, combination logic is applied on the OTP1 and the IMEI number of a customer mobile and the OTP2 will be generated and reserved in the bank database.

A customer then will have to scan the QR-code using the mobile QR-code scanner and in this process OTP1, extraction and permutations are done. Again, OTP2 will be generated which the customer has to enter to login. However, if the new OTP2 does not match the one in the database, the transaction is declined and if they match, the transaction will succeed.

The researcher mentioned that because a mobile is a gateway there are higher chances of intrusion or attacks and as such the QR –code scanning is decoded on a user’s mobile to prevent these intrusions. This is a good approach as attackers cannot easily have access to user’s mobile phone. Even though attackers can hardly have access to user’s mobile phone there are chances that the phone can be lost, Gandhi et al. (2014) did not provide an alternative means of authenticating after the loss. The researcher claims that using OTP and QR-code provides better security and convenience over other methods but when carrying this proposed authentication system, no tests were made hence no results/clear evidence to validate the accuracy and safety of this measure. Therefore, the reliability of his method is questionable. This method has

potential to be great if all the limitations are addressed.

2.2 Biometric fingerprint

“Biometric finger print authentication is an automated method verifying a match among different human finger prints” (Sana and Rana, 2014). It is preferred because of its uniqueness, accuracy, speed and it is easy to use. Figure 3 below shows the process of verifying a claimed user identity and enrolment of a person into the system. In the enrolment process the minutiae points are extracted and stored into the template database where upon the process of recognition the stored attributes will be retrieved and matched (Jani, 2015).

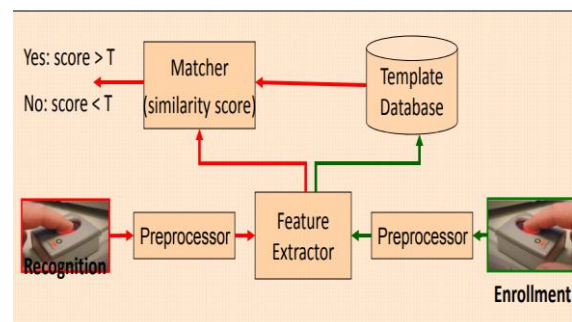


Figure 2 Physical Registration (Jani, 2015)

Tassabehji and Kamala (2012) illustrates the Schematic diagram of proposed biometric banking system where a user attempt to access online banking service. The user firstly must register the fingerprints so that the print information is captured securely. When accessing online banking services, a user must place a finger on the finger print reading device to authenticate through the help of attributes, which were captured in Figure 2 above being the minutiae points, ridges and furrows of the finger. Upon successful authentication, a web browser will be launched on the personal computer (PC) and a secure key will allow the user to login.

The personal computer used does not allow the use of uniform resource locaters (URL's) to prevent any man in the middle attack which may deflect the connection to other addresses. After launching a web browser on the personal computer, the key in the device will then establish a secure connection to the bank hence granting access to the bank services. However, in cases where a wrong fingerprint was entered for several times, key will lock as such a call for re-validation.

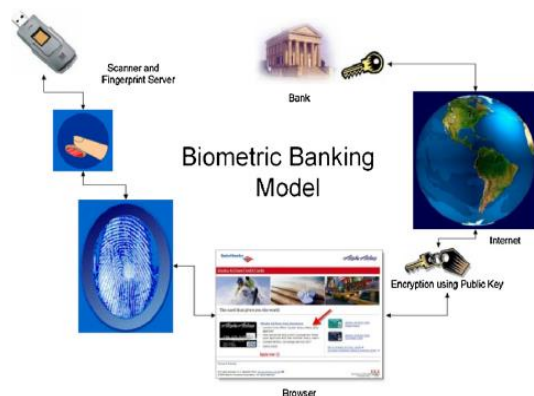


Figure 3 Biometric Banking System by Tassabehji and Kamala (2012)

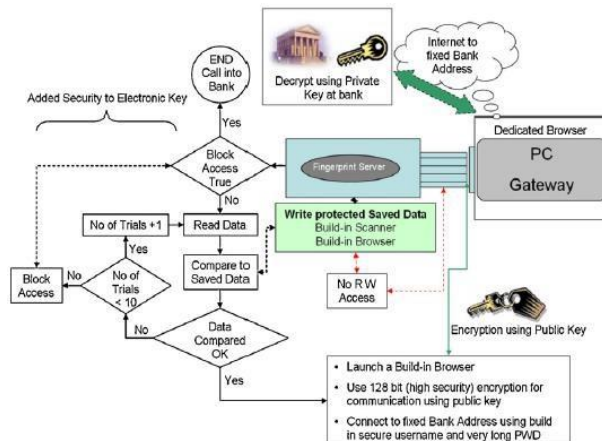


Figure 4 Method control proposed for the system by Tassabehji and Kamala (2012)

To evaluate the proposed biometric banking system a usability scale (SUS) was used .116 users were given questionnaires based on the Brooke and in this testing, users feared that the bank will keep copy of their biometric information. This was made evident as only 44% of the participants were willing to re-verify and re-authenticate.as for the biometric banking technology specifically, majority were not familiar with the system.

An experiment was carried by Tassabehji and Kamala (2012) and the results shown in figure 5

below expresses that biometric finger print is the mostly preferred technique over other biometrics techniques which are facial recognition, iris scanning and voice recognition.

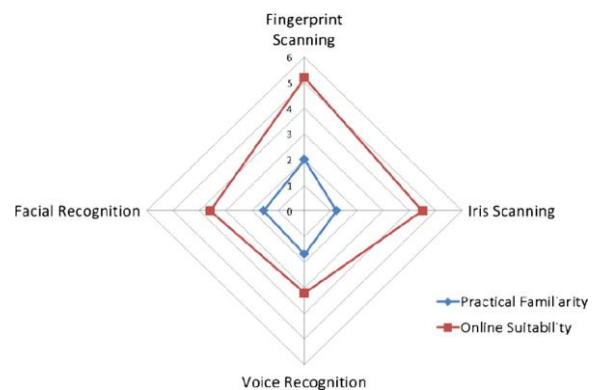


Figure 5 Experience of using biometrics by users (Tassabehji and Kamala, 2012)

Tassabehji and Kamala (2012) mentioned that the system uses corresponding minutiae to authenticate the users. However, other authors like Karthikeyan and Vijayalakshmi (2013) stated that finger cuts and marks can prevent the user from successfully authenticating, and as such recommends the use of correlation-based fingerprint verification system as it is able to verify the users print even when the minutiae cannot be extracted and it is also able to deal with finger prints that can suffer from non-uniform shape distraction.

The findings show that the SUS was efficient, but the author mentioned that the score is not absolute and as such in some of the occasions, it can be difficult to interpret qualitatively. On top of this, it was mentioned that the assessment is not that accurate and this express that the SUS must be improved and reviewed. The researcher also expressed that despite SUS being able to provide usability information it does not give information on how the system can be improved and as such, this called for a thorough research and investigation on how best it can be improved.

Moreover, Tassabehji and Kamala (2012) claim that finger print approach requires less input from the user, ease of access and security. However, different authors articulated that fingerprint is not as secure; Saini and Rana (2014) stated that fingers prints are not as private as finger scanners can be bypassed by 3D printed mold and stolen from selfie photos. They continue to state that it can take several swipes to authenticate which can take too much time than expected. Omogbhemhe and Bayo (2017) also stated that fingerprint could be cheated using artificial fingerprint and as such recommend that multifactor biometric technique be used as it provides strong security.

2.3 Online Signature Verification

Chadha et al. (2013) proposes efficient method to signature recognition using Radial Basis Function Network (RBFN). This method ensures there is a correlation between the newly entered signatures and the ones that exists in the database. Chadha et al. (2013) articulated that financial organizations need signatures to authorize confidential transactions.

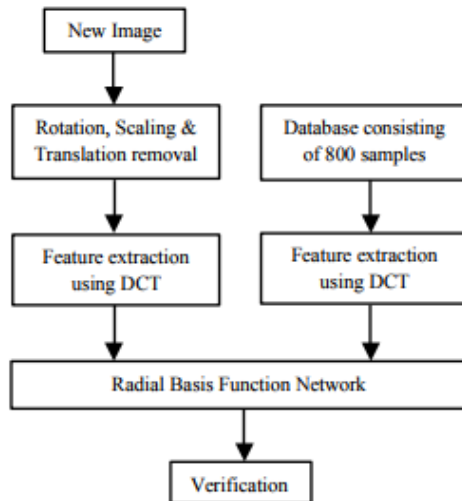


Figure 6 Proposed system (Chadha et al., 2013)

Chadha et al. (2013) carried an experiment to evaluate signature recognition using the Radical Basis Function Network (RBFN). A Wacom Bamboo digital pen tablet was used to capture the new signature image that will undergo rotation, scaling and translation combination. Chadha et al. (2013) mentioned that signature rotation, scaling and translation combination algorithm was used to process the signature image. This was done to be able to validate the signatures as there are dynamic characteristics in the process of signing.

The signature features will be extracted using the DCT and the image will be provided to RBFN that is trained using a database. An image database made of 700 signatures samples was invented and 10 signature samples each was collected from 70 people. To test the signature recognition system, MATLAB was used. This system is of advantage as it uses neural networks and as such requiring only few samples for the system to be trained. Chadha et al. (2013) articulated that the RBFN will match the new signature and the one that exists in the database, if the signature is recognized as the one that exists in the database, the user will be granted access and if it does not access will be denied. Figure 7 and 8 shows results for the validation of the combined rotation, scaling and translation algorithm respectively

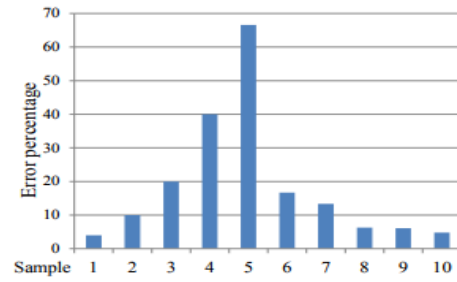


Figure 7 Graph representing errors experienced in angle rotation (Chadha et al., 2013)

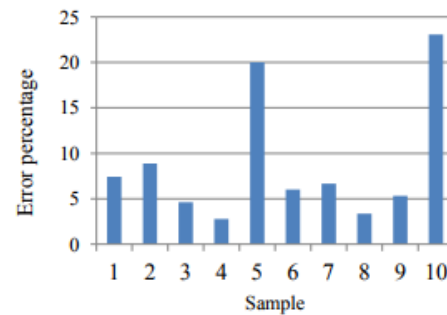


Figure 8 Graph representing errors experienced in scaling parameter (Chadha et al., 2013)

An experiment was carried and the results in the figure below shows the success rate of using online signature verification technique after using MATLAB as according to (Chadha et al. 2013). From the 700 samples, 500 were tested and 80% recognition rate was done from 200 samples which gives the success rate of the system.

RECOGNITION RATE FOR VARIOUS NUMBERS OF SAMPLES

Number of samples	Recognition rate
500	71.34%
400	76.8%
200	80%
100	79.2%
50	72.65%

Figure 9 Recognition rate of new signature as compared to those that exists in the database (Chadha et al., 2013)

The researcher claim that the method is efficient but from the results obtained there is inconsistency of results for example 50 samples have a recognition of 72.65% less than 80% of 200 samples and as such a pattern cannot be derived whether the more the signature or the less the signature the higher the recognition rate. This make it difficult to come to make a conclusion on the overall performance of the system therefore additional testing is required to validate the claim. Moreover, some of the people

cannot write consistently because of their fine motor skills combination and as such given how the system work, the relevant user can be deemed invalid based on signature inconsistency. To improve accuracy, (Jain and Gangrade 2014) method of using global features can be used. Both methods can be integrated to produce more accurate results. Provided all the limitations of the method are met, a greet method can emerge

2.4 Kerberos Authentication

Nwogu (2015) proposed a method that is known as Kerberos authentication that protects client login details as it uses symmetric key cryptography, data encryption standard and end-to-end security between a client and a distribution center. Additionally, it comprises of servers that manages the Key distribution Centre (KDC), the Ticket Granting services (TGS) functions and authentication services. Moreover, Kerberos provides timestamps that helps in reducing the message numbers that are required for authentication and allows cross-realm authentication.

Kerberos ticket provides a session key, verify and authenticate the client. Nwogu (2015) expressed that the ticket is encrypted and as such, the Kerberos server can only recognize it and the online banking server after the client has sent it.

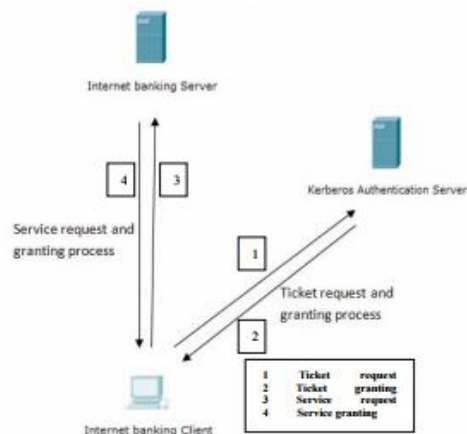


Figure 10 Kerberos authentication method (Nwogu, 2015)

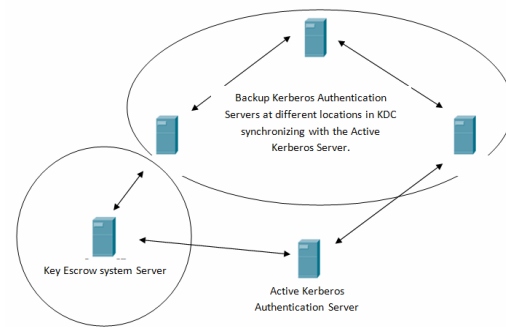


Figure 11 Key escrow system server and Kerberos server (Nwogu, 2015)

Nwogu (2015) explained that firstly the user enters their personal identification number (PIN) and biometric finger print into the client. In the client, entered credentials will be encrypted with DES for transmission to the KDS where they will be verified. In the KDS, a ticket granting ticket (TGT) will be generated and the users' credentials will be hashed. The TGT that the client installs will be encrypted with the DES. In the internet banking service request, Nwogu (2015) said that the client will send the TGT which it received from the KDC back with a request to be granted permission to access to the internet banking services. The KDC will then validate the request and if accepted the service ticket (ST) will be generated and sent to the client. Upon receiving of the ST, the client will send it to the internet-banking server and verify. When the ST is verified, the Kerberos will be complete, a session will be opened, and data transmission will start.

The KDC used by the researcher indeed expresses the higher security level of Kerberos as it is a two-factor authentication method and as such in cases where attackers get one factor right, they will still be requested to enter the second factor. Even though the researcher presented that both PIN and fingerprints are used in the KDC, there was no explanation given on how the factors are integrated together. Nwogu (2015) claim that the system is secure as it ensures there is confidentiality non-repudiation and data integrity, but no tests were made. It would have been ideal if there were results attained and presented to know how reliable the method can be.

3 Conclusions

In this document, current literature on online verification techniques in internet banking transactions has been fully evaluated and analyzed. Evaluation of these techniques was centered on security and performance.

For some techniques, it is difficult to make a conclusion on their level of security and reliability, as there are no experiments and test results

presented. Nevertheless, experiment carried by Tassabehji and Kamala (2012) addressed most of the factors and as such was satisfactory.

Looking at research done by Nwogu (2015) and Gandhi et al (2014), it is difficult to make a concrete conclusion and validate their claims on proposed techniques as no tests were done. Further experiments and tests are needed to validate these proposed methods. If more tests and research are done these techniques are promising. It would have ideal if a real online banking system was needed to thoroughly assess the techniques on their level of security and performance.

Nonetheless, considering the presented schemes, a combination of research done by Tassabehji and Kamala (2012) and (Nwogu,2015) can be done to come up with a better performing and more secure measure needed in online transactions.

References

- Chadha A., Satam N., Wali V., 2013, 'Biometric Signature Processing & Recognition Using Radical Basis Function Network', *International Journal of Digital Image Processing*, 3(1), pages 5-9.
- Gandhi A., Salunke B., Ithape S., Gawade V., Chaudhari S., 2014, 'Advanced Online Banking Authentication System Using One Time Passwords Embedded in Q-R Code', *International Journal of Computer Science and Information Technologies*, 5(2), pages 1328-1329.
- Jain K., 2015, 'Banking on Biometrics', *International Journal of Applied Information Systems*, 5(2), page 8.
- Jain P. and Gangrade J., 2014, 'Online Signature Verification Using Energy, Angle and Directional Gradient Feature with Neural Network', *International Journal of Computer Science and Information Technologies*, 5(1), page 216.
- Karthikeyan V. and Vijayalakshmi V.J., 2013, 'An Efficient Method for Recognizing the Low Quality Fingerprint Verification by Means of Cross Correlation', *International Journal on Cybernetics & Informatics*, 2(5), pages 4-6.
- Nwogu E. R., 2015, 'Improving the Security of the Internet Banking System Using Three-Level Security Implementation', *International Journal of Computer Science and Information Technology & Security*, 4(6), pages 173-175.
- Omogbhemhe M.I. and Bayo .I, 2017, 'A Multi-Factor Biometric Model for Securing E-Banking', *International Journal of Computer Applications*, 159(4), pages 21-23.
- Philip J and Bharadi V. A., 2016, 'Online Signature Verification in Banking Application: Biometrics SaaS Implementation', *International Journal of Applied Information Systems*, 3(5), pages 28-29.
- Saini R and Rana N., 2014, 'Comparison of Various Biometric Methods', *International Journal of Advances in Science and Technology (IJAST)*, 2(1), pages 26-27.
- Tassabehji R and Kamala M. A., 2012, 'Evaluating Biometrics for Online Banking:The Case For Usability', *International Journal Of Information Management*, 0(0), pages 490-493.
- Yadav G., 2015, 'Application of Biometrics in Secure Bank Transactions', *IJITKM*, 7(1), page 124.