Selected Computing Research Papers

Volume 5

June 2016

Dr. S. Kendal (editor)

Published by
the
University of Sunderland

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

An Analysis of Current Computer Assisted Learning Techniques Aimed at Boosting Pass Rate Level and Interactivity of Students

Gilbert Bosilong

Abstract

Computer assisted learning is currently the best substitute of the traditional teaching methods. In this paper, we closely evaluate and analyze the current computer assisted learning techniques aimed at improving the effectivity and pass rate level of students. In the end, conclusions and recommendations are stated based on evidence provided throughout the paper.

1 Introduction

Technological advancements have played a humongous role in the development of societies. Nowadays almost everything is technology based. For instance, one area that has enjoyed the fruits of these technologies is the education sector. One of the technologies aimed at improving education include Computerized Assisted Learning (CAL) which has been pervasively recognized worldwide as a way of boosting interactive learning amongst students.

Wang and Young (2014) insists that students sometimes have a hard time learning openly in front of instructors and peers due to anxiety and as result computer based learning models can help improve this. Arslan and Sahin-Kızıl (2011) believe that a large class size and a limited learning time constitutes to teachers not providing equal learning opportunities to students and as result, the students may learn for a while but not grasp the concepts as a result CAL may be a good mechanism to fill this gap.

According to Johnson and Dickinson (2012), CAL can play a significant role in improving and standardizing content delivery, reducing costs and time of traditional instructional learning, improving retention and accommodating learners in diverse geographic locations. As a result, CAL is of the essence these days.
Traditional instructed learning has its own barriers ranging from inadequate directed experiential learning, poor standardized curriculum and insufficient numbers of appropriate teachers thus effective teaching of critical skills remains amiss but CAL has an immensely scalable approach to address these issues (Kalet, et al., 2012).

Throughout this survey paper we aim to analyze and evaluate different computer assisted learning approaches proposed by other authors in order to reach a scientific conclusion.

2 Using A Computer Assisted Learning System To Carry Out Word Based Mathematical Problems

To help low achievers in second and third grade in Taiwan, Huang, et al. (2012) a word based mathematical problem solving system was developed by the authors to help students through parts of problem solving processes that they often ignore. They devised this system because of their belief that it is not merely enough to evaluate students based on their writing abilities when it comes to solving arithmetic procedures as a result their system incorporates the element of thorough evaluation.

To test their method Huang, et al. (2012) focused on second and third graders who had average marks of 25% in their Mathematics monthly exams and had the need for remedial instruction. Class advisors according to this criterion

identified these students. Their method was conducted in order to figure out if the system would have any significant effect on the low achieving students. Through performing these tests, they wanted to observe if they could reach mutual conclusions pertaining other methods that other authors had implemented (Huang, et al., 2012).

The experimental research that Huang, et al. (2012) conducted involved a total of 28 elementary students from two elementary schools with seventeen second and third grades as an experimental group and eleven second and third grades as a control group. Huang, et al. (2012) implies that when testing for difficulty and discrimination the test questions, a pilot study was held between two exams from nine classes. The test results were divided into two examination papers and both papers were designed respectively for the third and second grades putting in mind the level of difficulty that is, for third grade they used three figure questions and for second grade they restricted the questions to two figures. Upon the completion of the tests, they combined the papers for third and second grade to analyze split-half reliability and difficulty. In their research, they employed independent-sample and dual-factor covariance analysis. They used pretest score as the dependent variable and the posttest score as their control variable. This intended to figure out if there was any discrepancy in learning achievement before and after the experiment in order to clarify if the system could serve its purpose. The figure below shows the results attained from this tests (Huang, et al., 2012).

Number of Pretest and Posttest Participants in Each Group and Pretest and Posttest Average Scores

Group	Grade	Total	Pretest	Posttest
Experimental Group	2	17	11.27	15.00
	3		11.33	13.83
Control Group	2	11	12.00	11.67
	3		15.20	14.80

Figure 1 Pretest and Posttest Number in Both Groups with Pre and Post Average Scores (Huang, et al., 2012)

Huang, et al. (2012) indicates that there is high score of experimental group students compared control group students wich stipulates that the method that they used can help students with basic word based arithmetic questions and this can help low achievers students to improve their marks.

From the methodology that Huang, et al. (2012) used we it is accepted that they really ensured that they reach a general conclusion by selecting samples from two schools. Two schools means that the aspect of generalizability is met because the results does not only affect one school but two, as result the aspect of biasness will be very limited. However, taking only 28 students as sample surely raises an eyebrow because a solid conclusion cannot be reached from such a small sample. Have they selected over at least 200 students then a sound conclusion could have been made. As such the conclusion here is that the authors' conclusions are unexceptable because, room for bias is massive due to a limited sample size. Besides the limited sample size, this is good research and it can be repeated by other researchers but only increasing the sample in order to reach a conducive conclusion out of this study area.

3 Using Computer Assisted Learning To Improve Retension On Science Education

Kara, (2012) proclaims that technological advancements have led to a rapid change of lifestyle and as a result, traditional teaching methods are not that much effective anymore and as a result computer assisted instruction plays a vital role in improving achievements of students. He further explains that instructional materials

are more effective when supported by images, sound and animations.

To conduct his research Kara, (2012) used cluster sampling by choosing a population study of elementary school students in 7th grade in Denzil city. The cluster sampling ensured that the criteria for choosing the population was choosing a group that is naturally together. In the end Kara, (2012) chose one elementary school with a sample of 136 students and six teachers who will be aiding the control group and to prevent any prejudice, Kara, (2012) ensured that all the chosen group study science.

As a way of trying to measure students' achievement, Kara, (2012) developed a test about force and presure units consisting of 25-questions. Both groups were given the test as a pre, post and retension test. Kara, (2012) explains that he used three CAI programs that are already in existence to conduct his experimental tests. These programs include "*Mobides CAI systems, Vitamin Educational Program and Educational program*".

Kara, (2012) states that 85 students were used as a control group and 47 students were used as an experiment group. The pre-test was administered to test the current knowledge base of both groups and then a posttest followed after five months to observe the difference as well as a retention test to further find if there is any significant difference on performance both groups and the results were as follows.

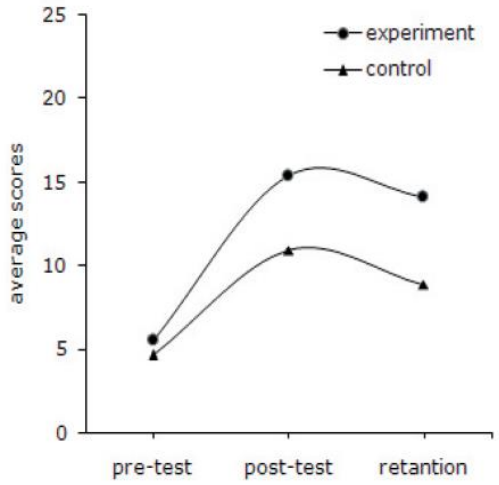


Figure 2 Comparision of achievement tests (Kara, 2012)

From the results in figure 2, it is observed that experiment group holds an upper hand in terms of pre-test, post-test and retention test. Kara, (2012) concludes that, from this this findings CAI does not only improve learning but also has a significant impact on improving retention. He further concludes that CAI is more effective than traditional teaching instruction teaching method.

The conclusions reached by Kara, (2012) seem valid because the methodology carried out is consistent with the results found and the conclusions reached. This experiment could easily be repeated since the author described all the inputs and outputs thereof, which means that even if the experiment is repeated the same results can be reached. The sample size is ideal since the population was chosen using cluster sampling and this makes it relatively easy to gather unbiased results since the population is chosen randomly in their natural environment. As result the conclusion is that this is a very sound research due to its state of the art execution.

4 7E Learning Model Based Computer Assisted Teaching Materials On Chemistiry Education

There are several pieces of work that investigates the effectiveness of computer assisted learning methods such as (Chang, et al., 2015), (Erbas, et al., 2015) and (Kalet, et al., 2012) and they all come to a common conclusion that computer assisted instruction is able to produce better effectiveness as compared to traditional teaching methods. They solidify their claim by explaining that simulations, animations, voice and pictures play a vital role in improving learning.

As a way of trying to validate this claim Kunduz and Seçken, (2013) developed a computer assisted instructional material which involves virtual lab, animations and educational computer applications based on a 7E learning model. The model focused on precipitations and tittrations on a chemistry topic trying to find out if there is any significant differences between computer assisted learning and traditional teaching methods. (Kunduz & Seçken, 2013)

To carry out their investigation Kunduz and Seçken, (2013) used the quasi-experimental design. They assigned control and experimental group to the already formed classes. Kunduz and Seçken, (2013) states that they collected data through quantification data collection tools and the data collected was quantitative data, which was collected by prepared achievement tests.

Kunduz and Seçken, (2013) points out that their sample study consisted of 89 students in the 11th grade. Control group consisted of fourty-five students and the remaining students formed an experimental group. They further explain that the control group was instructed using traditional teaching methods and they instructed the experimental group using the 7E computer assisted instruction system. Achievement test on precipitations and titrations was prepared and the data was collected through the administration of similar pre- and post-tests to both groups. Kunduz and Seçken, (2013) explains that the pre-test was given as a control to test the equality of both groups regarding their prior knowledge and understanding. For comparison of results of the two methods, they also conducted a post-test.

Figure 3 below shows the results of the pre-tests performed by Kunduz and Seçken, (2013).

Measurement	N	$\bar{x}$	SS	SD	t	p
Pre-test of experimental group	44	4.93	2.389	87	0.580	0.564
Pre-test of control group	45	5.20	1.946			

Figure 3 Comparison of Pre-Tests For Control and Experimental Groups (Kunduz & Seçken, 2013)

From Figure 3 above, Kunduz and Seçken, (2013) explain that the initial tests shows that there is no big difference pertaining to prior knowledge of the precipitation and titration subject between the two groups.

Kunduz and Seçken, (2013) stresses that they later administered the post-tests after teaching both the control and the experiment group with the two methods and the results were as follows.

Groups	N	$\bar{x}$	SS	SD	t	p
Experimental group	44	11.20	4.663	87	-2.756	0.008
Control group	45	9.16	1.623			

Figure 4 Comparison of Pre-Tests For Control and Experimental Groups (Kunduz & Seçken, 2013)

Kunduz and Seçken, (2013) explains that the difference in standard deviation of experimental group is higher than that of the control group when comparing the pre-test results with the post-tests. From this significant difference, they conclude that the use of computer-assisted instruction is much more effective than traditional teaching methods. Kunduz and Seçken, (2013) further explain that they cannot deny the fact that traditional teaching methods do yield results as we can observe from the two tables but CAI holds an upper hand in terms of statistics.

The laying out of this experimental design of this study was phenomenal and it is evident that the same experiment could be repeatable since the authors stated out all the inputs and outputs of their experiment. However, contrary to the conclusions made by the authors, there are some points where they are really not right. The reason being that their sample size of eighty-nine is not enough to make a valid conclusion and it could mislead the results of the study. The quasi-experimental design was used in this study, this experimental design lacks the aspect of randomization, and randomizing samples plays a vital role in getting rid of bad science. Hence, the use of this experimental design could bias the results. In conclusion, further research needs to be done in this area and better experimental designs need to be implemented with a better sample size as a results the conclusions of this study are not accepted.

5 Interactive White Board and Computer Based Graphing Utility

Erbas, et al., (2015) conducted a study aimed at investigating how IWB technology can affect the achievement of students in quadratic equations by comparing it to traditional instruction. The IWB was used with other graphing software known as "NuCalc graphing software".

Their methodology consisted of sixty five high school graduates divided into control and experiment group. According to Erbas, et al., (2015) 34 students made up the control group and 31 students made up the experiment group with age ranging between 18 and 20. They further state that the students in both groups attended two different bridging courses during evening classes.

Erbas, et al., (2015) explains that the instrument they used to test for achievement is the graphing achievement test which they developed to measure the achievement of students in graphing quadratic functions. The test consisted of multiple choice questions, and essay type questions which suited the curriculum. This was used to assess how students interprets questions. This instrument was crosschecked by experienced teachers to see if its level of difficulty is reasonable. This test ensured that there is a scaling concept involved. Erbas, et al., (2015) graphing achievement test was done in the form of pre and posttest and both tests were separated by two months for both groups. For procedure Erbas, et al., (2015) explains that two learning environments were used, one for traditional teaching and the other for incorporating IWB. The experimental group students

were taught IWB in a computer lab and control students were taught in classroom traditionally with blackboards. For effectivity both the groups were taught by the same people who used similar content and lesson execution explains Erbas, et al., (2015). Data was collected over two weeks and a series of interviews was conducted over the course of the investigation in order to find out if students are benefiting in the study or not.

Table 1. Descriptive statistics regarding experimental and control groups' pre-test, post-test, and delayed post-test scores for the GAT

	Pre-test		Post-test		Delayed post-test	
	M	*SD*	*M*	*SD*	*M*	*SD*
EG (n = 31)	10.68	9.24	45.58	5.75	43.26	5.02
CG (n = 34)	10	8.59	35.21	8.64	32.62	7.39

Figure 5 Pre Post and Delayed Post Scores for GAT (Erbas, et al., 2015)

According to Erbas, et al., (2015) the mean shows that both groups being control and experiment have gained knowledge content. Erbas, et al., (2015) also explains that the mean shows an increase from pretest to posttest then it shows a decrease from posttest to delayed posttest while in overall the experiment group shows greater increase. As a result the findings is in favor of the experimental group. The author concludes that when technology is implemented effectively it can positively boost students' performance, morale, engagement and interest in learning resulting in improved knowledge.

The authors excellently executed their investigation by stating out their purpose of investigation and the literature thereof. Erbas, et al., (2015) then marvelously states out the methodology that they used by giving out the number of sample population. The only problem that arises here is that the author did not stipulate how they went about to select their population sample but they only state the sample size and their geographic properties. Failure to do this may raise a few issues because now we will not be certain about the outcome of the results since the author may select the population sample to their benefit which may have led to the cherry picking effect. It is not easy to take sampling seriously when they are not clearly stipulated. However the instrument for measuring success was clearly laid out and the criteria for experiments carried out was on point. The conclusions reached matched the experiments used and the methodologies used. This was indeed a good piece of research.

6 Conclusions

In this research paper some of the most recent computer assisted learning techniques have been analysed. The focus was put more into finding out whether they can surpass the traditional teaching methods in terms of effectiveness.

Most of the papers analysed reveal that CAL is more effective as compared to traditional teaching methods. However, some researchers may have produced biased results due to poor experimentation. For example Huang, et al., (2012) and Kunduz & Seçken, (2013) used a very small smple size which could have lead to an invalid conclusion.

It should be put in mind that these computer assisted learning systems need to be created to cater for majority of students worldwide which means that a single researcher may develop a program for oneself but due to essentric development, the system may lack important properties hence may produce biased results when experimented with. Due to this drawback it can be suggested that when researchers want to develop an internationally recognised CAL system, there should be a pool of researchers involved and a very large sample size catering for most of the ethnic groups worldwide. This way not only mean the system will be robust but the research implemented will be worthwhile.

One of the critical findings of this survey is that CAL involves a lot of multimedia aspects such as sound, animations, simulations, graphics etc which cannot offered by traditional teaching methods and this has

shown tremendous effect in retaining students' attention and focus thus improving effectiveness on students perfomance and interaction. Even though different research focuses on different subjects in education, this has proven that CAL can be a beneficial technology in the education sector worldwide.

All in all, more research has to be done in order to produce a state of the art CAL system that facilitates for ethinic background, age, gender, race etc of students. This alone could be the next generation of technological education. This survey has shown that CAL is more effective than the traditional teaching methods and it is highly recommended to various schools more especially those suffering from low academic perfomance of students.

References

Arslan, S. R. and Sahin-Kızıl, A., 2011. 'How can the use of blog software facilitate the writing process of English language?' *Computer Assisted Language Learning*, 3(23), pages. 183-197.

Chang, S., Ku, A., Yu, L., Wu, T., Kuo, B., 2015. 'A Science, Technology, Engeneering and Mathematics Course with Computer-Assisted Remedial Learning System Surpport for Vocational High School Students.' *Journal of Baltic Science Education,* 14(5), pages. 641-654.

Erbas, A. K., Ince, M. and Kaya, S., 2015. 'Learning Mathematics with Interactive Whiteboards and Computer-Based Graphing Utility.' *Educational Technology & Society,* 18(2), pages. 299-312.

Huang, T.-H., Liu, Y. and Chang, H., 2012. 'Learning Achievement in Solving Word-Based Mathematical Questions through a Computer-Assisted Learning System.' *Educational Technology & Society*, 1(15), pages. 248–259.

Johnson, D. A. and Dickinson, A. M., 2012. 'Using Post-Feedback to Improve Retension of Computer Based Instruction.' *The Psychological Record,* Issue 62, pages. 485–496.

Kalet, AL., Song, H.S., Sarpel, U., Schwartz, R., Brenner, J., Ark, T.K., Plass, J., 2012. 'Just enough, but not too much interactivity leads to better clinical skills performance after a computer assisted learning module.' *Medical Teacher,* Issue 34, pages. 833–839.

Kara, İ., 2012. 'The Effect on Retention of Computer Assisted Instruction in Science Education.' *Journal of Instructional Psychology,* 35(4), pages. 357-364.

Kunduz, N. and Seçken, N., 2013. 'Development and Applica tion of 7E Learning Model Based Computer-Assisted Teaching Materials on Precipitation Titrations.' *Journal of Baltic Science Education,* 12(6), pages. 784-792.

Wang, Y.-H. and Young, S. S.-C. Y., 2014. 'A Study of the Design and Implementation of the ASR-based iCASL System with Corrective Feedback to Facilitate English Learning.' *Educational Technology & Society,* 2(17), pages. 219–233.

Evaluating the Ability of Anti-Malware to Overcome Code Obfuscation

Matthew Carson

Abstract

Obfuscation has become a common technique used to overcome the protection provided by anti-malware systems. With computers and technology becoming significantly more integrated into modern society, protection of these integrated devices becomes more and more critical. Through the analysis of three proposed methodologies for improving anti-malware systems accuracy rate in comparison to modern malware, this paper identifies the strongest aspects of these methodologies and ascertains the trustworthiness of the results aligned with these methods. Utilizations of these strengths are then suggested for future development of anti-malware technologies.

1 Introduction

With current technologies advancing as they have, malicious use of this technology has advanced as well. Malware was once easily dealt with via anti-malware techniques, Alazab (2015) suggests the recent failures of these techniques to the iteration-over-innovation approach that malware coders take regarding their malware. This constant refinement of viruses has led to the corrupting code far out-classing the effectiveness of the counter-measure.

Code obfuscation is a very simple technique that has led to the high amount of different malware being active. Chen. et.al (2011), states that code obfuscation and defense lowering techniques are used by malicious software to ensure its own safety. Overcoming this defensive measure is something that current techniques are incapable of doing.

Malware is ever developing ensuring that a 100% accuracy rate in detection is nigh impossible with current methods, however we must continue to strive to find "clues to the precise analysis of malware"(Zhao. et.al, 2014) and improve our detection methods with severe innovation.

Current research has brought forward a range of solutions that could potentially result in a more effective anti-malware solution. Shih. et.al (2013) believed that further preventative action was needed but most researchers seem to prefer reactive measures. Following the works of these researchers, we will identify which of their techniques can overcome code obfuscation and could be used to avoid losing control of the devices that have become integrated in many integral parts of society.

2 Current Anti-malware Techniques

Proposed solutions to the malware problem seem to run along the same line of thinking. Keeping in line with the reactionary approach of current anti-malware techniques, many solutions suggest identifying the actions of a program and evaluating it based on said actions, this is to counter the obfuscation methods employed by malware coders. Through analyzing their results we can identify which solutions yields the most benefit.

2.1 Binary File Checking

POPA (2014) suggests in their research that infected object files will allow malicious applications to easily activate. The method suggested by the researcher implies recording details about the object's binary files so that infection can be identified. Their method claims to operate based upon a method of comparing scanned binary files to stored binary files of normal and malicious code to identify what may or may not be infected. Continuing forward, POPA (2014) states that the layout of object files can change once they have been infected to hide malicious behavior of applications.

The researcher claims that recording 5 key factors of the object file's state will result in the detection of malicious applications attempting to use altered files. "*TimeDateStamp, NumberOfSections, PointerToSymbolTable, NumberOfSymbols* and *Characteristics"*(POPA. 2014) are suggested to be the key factors in identifying infected binary files, the research then describes the necessity of each factor. The work done within the research in question is theoretical in regards to improving static-call based malware detection systems and as such provides no testing of the theories.

The main claim made by this paper is that, when using this method, "lack of malicious code is also assured when the program source code is reviewed and accomplishes the minimum coding standards" (POPA 2014). Though no testing has been provided, the author is willing to make the claim that this method ensures that no false negatives can occur, the main problem with this is that it is stated matter of fact instead of adding to the point that this is theoretical work, this can lead to many people following these methods without having ensured they are actually an improvement over current techniques.

Modifying the mobile anti-malware work of Wang and Wang (2014) for use with a home network could prove more beneficial than testing and applying the work of POPA (2014) as the work of testing has already been performed for this technique. Though the application of this mobile-based technology could yield positive results, there is as little certainty that the results will be significant as there is for the technique of POPA (2014).

Without testing, this theory cannot be trusted, however through performing testing we can ascertain the ability of this system. Should the claims made by the researcher then be true after application to a system, we can expect an improved anti-malware system.

2.2 Classifying Packed and Polymorphic Malware

Cesare. et.al.(2013) bring their work forward by first describing the weaknesses of current techniques, then swiftly continues to state the basics of their own work. Through the analysis of binary information, the system is able to detect if the file has undergone packing leading into the next step in which static analysis then identifies characteristics of malicious code to be used for control flowgraphs as comparative data. Once this process has been completed it moves onto its second phase and analyses the unpacked binary for proper classification of the code and final analysis of malicious intent. Testing this method, the researchers used 14 different packing service tool on code that had been modified by themselves from standard Windows executables that could be found on any computer. Finally, the researchers concluded the paper with the claim of where their technique would be best applied to.

Name	Revealed code and data	Number of stages to real OEP	Stages unpacked	% of instr. to real OEP unpacked
upx	13107	1	1	100.00
rlpack	6947	1	1	100.00
mew	4808	1	1	100.00
fsg	12348	1	1	100.00
npack	10890	1	1	100.00
expressor	59212	1	1	100.00
packman	10313	2	1	99.99
pe compact	18039	4	3	99.98
acprotect	99900	46	39	98.81
winupack	41250	2	1	98.80
telock	3177	19	15	93.45
yoda's protector	3492	6	2	85.81
aspack	2453	6	1	43.41
pepsin	err	23	err	err

hostname.exe

Name	Revealed code and data	Number of stages to real OEP	Stages unpacked	% of instr. to real OEP unpacked
upx	125308	1	1	100.00
rlpack	114395	1	1	100.00
mew	152822	2	2	100.00
fsg	122936	1	1	100.00
npack	169581	1	1	100.00
expressor	fail	fail	fail	fail
packman	188657	2	1	99.99
pe compact	145239	4	3	99.99
acprotect	251152	209	159	96.51
winupack	143477	2	1	95.84
telock	fail	fail	fail	fail
yoda's protector	112673	6	3	95.82
aspack	227751	4	2	99.90
pespin	err	23	err	err

calc.exe

Figure 1(Cesare. et.al 2013)

The results, as shown in Fig.1 above, reveal that many of the standard packing services used did not hide the synthetic malicious code from the detection method, however several errors and failures have been shown in the results, as well as some very low detection rates in regards to one of the altered files.

Fail results in the figure are marked this way from failed packing of the code not failed detection, as such nothing about them was recorded. Error in detection was explained as "possibly due to unused encrypted data remaining" (Cesare. et.al, 2013) and the low result was potentially due to heap usage, however nothing is stated outright and these appear to be assumptions of the problem not definitive answers.

'Synthetic samples' was the key alarming factor of the experiments, having never properly defined the term, it is only possible to assume what this means, reading further the researcher states the number of malware used and where it originates from, leaving 2 available options for assumption. Firstly, the definition could be simple malware designed purely for the purpose of testing, if this is the case then no issue can be brought up about this. Secondly, it could be defined as malware altered to have no harmful behaviours but simply read as malicious. Assuming the second definition is true, the problem arises that fully active malware can interfere with the operating of an application, in particular anti-malware applications, thus results in the

malware avoiding detection. Without properly defining this, the results taken for this section of the testing cannot be properly trusted.

The biggest success brought forward by the researchers was that their system showed a high resistance to the 'false positive' problem that many other techniques can face, to ensure that was the case, they performed rigorous testing to ensure this was the case, most commonly achieving a similarity result of 0.25. Results for the method developed by Elhadi. et.al (2014) however had a much more significantly low false positive rate, nearing 0 at peak true positive rates, even comparing it to many other systems with similar results. Taking this information, the results provided regarding these rates become significantly less impressive and boastful for Cesare. et.al (2013)

The researchers conclude their work by reiterating the problem they were working against, and claiming that Desktops, Internet gateways and anti-virus systems were all suitable areas of application for their system. Having previously stated that the intention of this system is intended for use on the application level, this seems to be a fair conclusion to make. Though the detection of this system seems excellent on paper, without a proper definition of 'synthetic samples' and with false positive rates being inferior to that of other techniques, this method should not be considered as detection of active malware the safety of benign software cannot be ensured.

Though the claims made by the researchers are sound in theory, when compared to other researcher's proposals the application of this work would be inferior to that of more recent work, even though it was produced within the previous 2 years, a great development seems to have been made since then, rendering the points made here moot.

2.3 Static System Calls

Yuxin. et.al (2011) suggests many of the same issues as the previous paper, their approach to the issue is also reminiscent, at least in part, of the previous work leaving little reason to summarize the specifics of the technique. The testing was focused around a collection of 300 benign all of which were taken from the system files of a copy of windows XP and 400 different malwares including "200 virus, 100 worms, and 100 Trojans" (Yuxin. et.al, 2011), taken to ensure high testing accuracy. Ensuring a lack of bias, 100 each of the benign and malware executables were used for training of the new method while the rest were used for testing. This training data was then used to produce 2 graphs, seen in figures 2 and 3 below, displaying the frequencies of recurring malicious and benign traits.

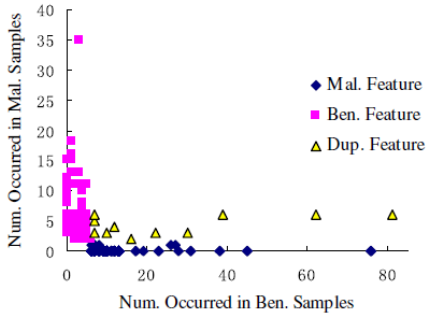


Figure 2 (Yuxin. et. al, 2011)

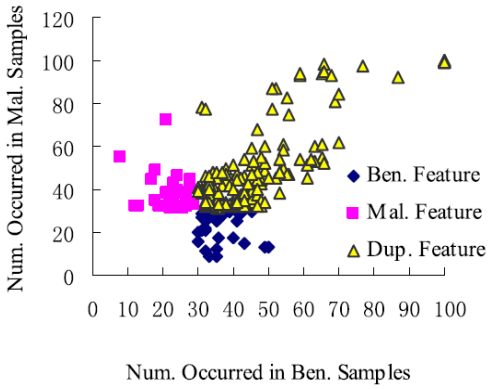


Figure 3 (Yuxin. et.al. 2011)

The figures show that which is associated with benign and malicious executables as well as traits shared between the 2. This data was used by the researchers to give a classification of characteristics for the system to identify between malicious and benign software. Through identifying traits both unique and similar across both types of executable, the system is then able to ensure lower false positive and higher true positive rates.

The executables used for testing provide an incredibly large range of malicious content as well as benign software, ensuring that the data presented represents the potential threats faced by computers in a real world scenario. The main alternative to this controlled experiment would have been to have performed a real-world experiment. Ignoring the ethical dilemmas that this proposes, there is no guarantee that the system would come into contact with such a wide range of unique viruses, yielding a narrower data set for evaluation or ensuring that the project would take substantially longer until the system does encounter these threats. Most, if not all, researchers seem to take this approach to testing their solution as Zhang. et.al (2014) and Jiang. et.al (2010) as well as the other researchers mentioned previously all opted to use controlled experiments.

The researcher arrived at the claim that Static Detection methods are more accurate than dynamic methods, supporting this claim they performed a selection of tests with the results of the static method being 5%-20% higher than the dynamic. This information was used as a primary reason for justifying their choice of static detection even though dynamic can fetch a result much quicker. Suggesting that static detection is superior to dynamic detection, without other context, comes across as a simple attempt to narrow the competition against the researcher's work, though with the information brought forward by the testing of the methods this claim is thoroughly supported and valid. Overall, it is shown that the proposed static method does have a higher accuracy rate than the dynamic method used as well as a lower false positive rate.

Through ensuring a lack of bias as well as a solid methodology for testing, the research makes claims that their system overall would ensure the safety of computer systems. The techniques proposed by the researcher could be a valid aspect of application to solving the malware crisis, though due to the age of the work and the rapid improvement of work when comparing that done by Cesare. et.al (2013) and Elhadi. et.al (2014), it would be a reasonable assumption that the work done by this researcher is also out of date and may not account for many of the newer threats that can so easily over-come anti-malware systems today.

3 Comparison of Methods

Through the analysis of these methods as well as the data proposed by the testing of the latter two methods, we can derive that the proposal of Yuxin. et.al (2011) shows great promise in overcoming modern threats to computer devices, though the main concern is that the method is relatively old. Compared to the work of Cesare et.al (2013), the former's work is more trustworthy and yields greater results. Overcoming obfuscation will undoubtedly be a herculean task, though using the work of Yuxin. et.al (2011) as a base would be a simple and effective way to start.

POPA (2014) proposed a characteristic based technique to overcome obfuscation, though as the method was untested the information provided cannot be substantiated. Assuming the claims made by the author prove to be true, this method could then be applied to methods of Yuxin. et.al (2011) to yield a more modern technique with the ability to analyse obfuscated and potentially malicious code.

4 Conclusions

Three main methods have been brought forward and analysed, through doing this several applications of this data have been acknowledged. Firstly, the static analysis method brought forward by POPA (2014) does show promise though the method has not been tested previously, once testing has been performed for this technique, the benefits can be confirmed and then the method can be applied to another methodology.

When analyzing the work of Cesare. et.al (2013) it once again shows promise of solving the malware-problem we are currently facing, however due to lack of definition of certain terms, the work performed by this researcher cannot be strictly trusted as the testing results of the work greatly depend on this. Furthermore, assuming that the results could have been trusted, the results of the testing are lacking when compared to results of another's work simply a year later, because of this the work can neither be trusted but is also outclassed.

Finally the work of Yuxin. et.al (2011), the testing performed by the researchers ensured that their results were more than trustworthy and accurate. However due to the age of the research it is a fair assumption that the methods proposed may be out of date in regards to modern malware and would not be a viable solution.

Though the work may not solve the obfuscation problem as individual pieces, using the work of two different authors in conjunction with each other could yield positive results. Through further research and development of techniques we can continue to better defend against the ever developing malware of the current technological age.

References

Alazab, M., 2015, 'Profiling and classifying the behavior of malicious codes', *Journal of Systems and Software*, 100, pp.91.102

Cesare. S., Xiang. Y., and Zhou. W., 2013, 'Malwise—An Effective and Efficient Classification System for Packed and Polymorphic Malware'. *IEEE Transactions on Computers*, 62.6, pp. 1193-1206.

Chen, Z., Liang, Z., Zhang, Y. and Chen, Z., 2011, 'Evaluating Grayware Characteristics and Risks', *Journal of Computer Networks and Communications*, 2011, pp.1-28.

Elhadi. A., Maarof. M., Barry. B., Hamza. H., 2014, 'Enhancing The Detection Of Metamorphic Malware Using Call Graphs'. *Computers & Security*, 46, pp 62-78.

Jiang, X., Wang, X. and Xu, D., 2010, 'Stealthy malware detection and monitoring through VMM-based "out-of-the-box" semantic view reconstruction', *ACM Transactions on Information and System Security*, 13(2), pp.1-28.

POPA. M., 2014 'Detecting Malicious Code By Binary File Checking'. *Informatica Economica 18*, 1, pp.111-119.

Shih, D., Chiang, H., Yen, D. and Huang, S., 2013, 'An intelligent embedded system for malicious email filtering', *Computer Standards & Interfaces*, 35(5), pp.557-565.

Wang, P. and Wang, Y., 2014, 'Malware behavioral detection and vaccine development by using a support vector model classifier', *Journal of Computer and System Sciences*, 81.6, pp.1012-1026.

Yuxin. D., Xuebing. Y., Di. Z., Li. D., Zhanchao. A, 2011 'Feature representation and selection in malicious code detection methods based on static system calls' *Computers & Security* 30.6-7, pp. 514-524.

Zhang, B., Yin, J., Wang, S. and Yan, X., 2014, 'Research on virus detection technique based on ensemble neural network and SVM', *Neurocomputing*, 137, pp.24-33.

Zhao, Z., Bai, J. and Wang, J., 2014, 'Malware detection method based on the control-flow construct feature of software', *IET Information Security*, 8(1), pp.18-24.

Evaluation of Current Research in Machine Learning Techniques Used in Anomaly-Based Network Intrusion Detection

Masego Chibaya

Abstract

Internet and network-based technologies work hand-in-hand and have shown rapid growth over the years because different sectors in the industry depend on them to carry out their day to day businesses activities. This has led to network intrusions or attacks to increase exponentially. Several Machine learning techniques are being introduced that can be used to counter against these attacks. This paper focuses on the analysis, comparison and evaluation of different Supervised and Unsupervised Machine learning techniques used to detect any intrusions in a network. Centering on the evaluation, one technique that outclassed the rest will be presented and recommended for further research.

1 Introduction

Different organisations use computer networks over the internet. This advancement has led to the number of network attacks under the internet environment to increase exponentially due to the fact that not all internet users have good intentions; some are perpetrators who a willing to create harm than good by finding and creating breaches that attack known vulnerabilities in our systems, even (Xiaxong and Ning 2014) indicated that the rapid expansion of network connection, especially the opening-up of internet and access of financial areas, has led to more intruding attacks in network systems. These perpetrators or hackers are quiet inventive people who are continuously designing new vulnerabilities, often referred to as "unknown attacks".

Anomaly-based intrusion detection is an effective measure that combats against such attacks. Liu et. al. (2014) says anomaly detection is based on the fact that anomaly activities are noticeably different from normal systems activities and thus detectable.

Even though anomaly-based intrusion detection proves to be an important part of a network defense mechanism, it is susceptible to generate false positive alarms due to the ever changing nature of applications and networks which affects accuracy. Several algorithms and techniques have been proposed to minimize false positive rates hence improving accuracy, (Gondal et. al. 2015) says over the past years various kinds of anomaly-based intrusion detection systems have been proposed, some are statistical-based, while others use machine learning, data mining and evolutionary techniques but the application of Machine Learning techniques to anomaly-based intrusion detection is promising. A breakthrough in this field will benefit different organisations in more ways than one because reducing false positive alarms gives IDS's the opportunity to tackle alerts that are harmful hence securing networks.

This paper presents an overview of current research being undertaken concerning Machine Learning techniques used to manage the false positive problem. The paper is organized as follows, in Section 2 detailed discussions on machine learning techniques used like Bayesian Network, Genetic Algorithm and Artificial Neural Networks are described and evaluated. Finally, the conclusion of this paper will be presented in section 3.

2 Machine Learning Techniques

Machine learning techniques are divided into two types, namely; supervised learning and unsupervised learning. These techniques are implemented in intrusion detection because of their ability to learn from and make predictions on data without being programmed where to look.

2.1 Bayesian Network (BN) Approach

Xu and Srihari (2014) define a Bayesian Network (BN) as a probabilistic graphical model that represents joint distributions of a set of variables proficiently in a factorised way. This supervised learning technique has been broadly used in the intrusion detection area due to its tremendous ability to reason under uncertainty and its robustness in modeling joint variables (Xiao et. al. 2014).

Xiao et. al. (2014) implemented the Bayesian concept and built a Bayesian Network Model Averaging (BNMA) classifier that calculated the average of K-best BN classifies to detect intrusion's such as Remote to Local, Denial of Service (Dos), User to Root attacks, and so on. Their main aim was to improve accuracy and the false-positive alarm rate triggered in a network. The authors of this paper say they formulated the metric below to compute the K-best average.

$$p(x|D) \approx \frac{\sum_{i=1}^{k} p(x|G^i, D)p(G^i|D)}{\sum_{i=1}^{k} p(G^i|D)}.$$

Figure 4 Formulated metric (Xiao et. al. 2014).

To arrive to the detection accuracy conclusion Xiao et. al. (2014) compared the proposed approach with two other classifiers, namely Naïve Bayes (NB) and Bayes Network built using greedy (BN {Greedy}). The authors used the same dataset consisting of 12 features presented in Table 1 across all methods.

Table 1 depicts that some selected features take continues values, but existing BN classifiers can only handle discrete values. For the authors to build their classifier, they adopted a discretization algorithm to discretize continues values.

FEATURE NAME	DESCRIPTION	TYPE
service	network service on the destination, e.g., http, telnet, etc.	Discrete
src_bytes	number of data bytes from source to destination	Continuous
dst_bytes	number of data bytes from destination to source	Continuous
logged_in	1 if successfully logged in; 0 otherwise	Discrete
count	number of connections to the same host as the current connection in the past two seconds	Continuous
srv_count	number of connections to the same service as the current connection in the past two seconds	Continuous
serror_rate	% of connections with errors (refer to the same-host connection)	Continuous
srv_serror_rate	% of connections with errors (refer to the same-service connection)	Continuous
srv_diff_host_rate	% of connections to different hosts	Continuous
dst_host_count	sum of connections to the same destination IP address	Continuous
dst_host_srv_count	sum of connections to the same destination port number	Continuous
dst_host_diff_srv_rate	the percentage of connections to different services, among the connections aggregated in dst_host_count (32)	Continuous

Table 1 Feature list extracted (Xiao et. al. 2014).

To conduct their experiment they used the NSL-KDD dataset (consisting of 41 features) simulated environment entailing of selected records from KDD Cup 99 dataset where redundant records were removed. Xiao et. al. (2014) articulates that this was a binary classification scenario that is, if a record has a probability higher than 0.5 it is classified as an attack otherwise normal.

A graph was produced (figure 2 below) which shows the accuracy of all the methods compared against each other.

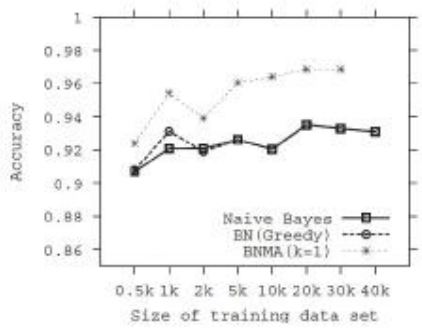


Figure 2 Comparison of methods used (Xiao et. al. 2014).

Figure 2 depicts the detection accuracy results where (Xiao et. al. 2014) says the BNMA classifier is much better than the other two. The authors continue to say that even when their proposed classifier trained a small set of 20000 it still outperformed the other two using a large training set of 40000. Experiments were repeated four times (Xiao et. al. 2014).

The authors of this paper clearly stated that their top priority was to limit the number of false positive alarms triggered by the IDS. However throughout the paper they do not articulate anything concerning that issue. Xiao et. al. (2014) only showed a mathematical metric that is to be used to build their method and concluded that indeed accuracy has improved. Had they formulated another metric that shows how to calculate the false positive side of things to back up the accuracy claims, their results would be reliable. Instead of using the KDD'99 dataset the authors of this paper used the NSL-KDD dataset because of the improvements made eliminating redundancy and guarding against biasness.

However using a simulated environment remains a problem due to the fact that internet traffic is exceedingly difficult to simulate realistically, so the results presented may be biased. Another striking issue is that it was not stated whether the dataset used the 'same' features which is important for any classifier during training; this alone makes the results even more questionable. Claims that experiments were carried

out four times were made but there is no evidence of results to back-up the author's declarations.

Even though measures were implemented when faced with situations, like discretization of continues values which makes the features easier to understand and interpret, this method is not ready to be implemented in the real world due to its biasness, lack of realism and relevance.

2.2 Genetic Algorithm (GA) Approach

GA is an unsupervised learning method which begins with a set of potential chromosomes (solutions) which contains a population that is selected at the beginning and these chromosomes evolve during various generations using techniques like mutation and crossover producing new offspring's (Pal and Parashar 2014).

Pal and Parashar (2014) presented a GA for intrusion detection which they claim achieved a low false positive rate. The authors continue to articulate that the main aim of the algorithm is to generate rules that will differentiate between attacks and normal classes. Figure 3 shows the proposed algorithm.

Algorithm: Rule set generation using genetic algorithm.
Input: Network audit data, number of generations, and population size.
Output: A set of classification rules.
1. Pre-process data by converting the symbolic feature into numeric data
2. Select 15 features based on information gain
3. For each extracted features
4. Normalize and Fuzzify each selected attribute and divide into fuzzy classes
5. Initialize the population
6. $W1 = 0.2$, $W2 = 0.8$, $T = 0.5$
7. N = total number of records in the training set
8. For each chromosome in the population
9. $A = 0$, $AB = 0$
10. For each record in the training set
11. I f the record matches the chromosome
12. $AB = AB + 1$
13. End if
14. If the record matches only the "condition" part
15. $A = A + 1$
16. End if
17. End for
18. Fitness = W1 * AB / N + W2 * AB / A
19. If Fitness > T
20. Select the chromosome into new population
21. End if
22. End for
23. For each chromosome in the new population
24. Apply crossover operator to the chromosome
25. Apply mutation operator to the chromosome
26. End for
27. If number of generations is not reached. go to line 4

Figure 3 Proposed algorithm (Pal and Parashar 2014).

The dataset used to carry out experiments was provided by MIT launch labs and consists of 41 features, of which 38 are numeric and the rest are symbolic (Pal and Parashar 2014). First the dataset is preprocessed to convert the symbolic features to numeric using the Rapid Miner tool. Due to large amount of network traffic in terms of features, only top features presented in table 2 where generated from the preprocessed data based on information gain.

Attribute	Information Gain
Att 23	1
Att 34	0.885
Att 2	0.850
Att 38	0.795
Att 25	0.770
Att 39	0.762
Att 26	0.723
Att 1	0.711
Att 37	0.652

Table 2 Top attributes selected based on information gain (Pal and Parashar 2014).

All the extracted features are normalized so that all the values range between 0 and 1 (as depicted in table 2 above), from there fuzzification takes place using the triangular function formula presented in figure 4.

$$\text{triangle}(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right).$$

Figure 4 Triangular function formula (Pal and Parashar 2014).

Rules are initiated based on the attribute values ranging between 0 and 1. The rules take place using substrings (L, ML, M, MH and H). These rules are used to differentiate between normal and anomalous behavior.

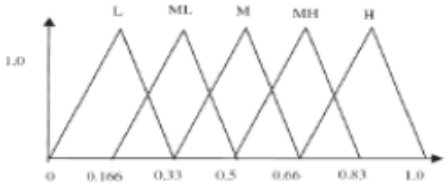


Figure 5 Division of class (Pal and Parashar 2014).

A Selection Operator is used to ensure that the best substrings are chosen individually. The Crossover Operator tool then randomly pairs the previously individuals to breed and exchange substrings. Then some of the substrings are altered using Mutation tool.

To test the model the authors say they used the KDD'99 data set and the system is feasible to apply in real-time. To attain the results (Pal and Parashar 2014) compared the proposed algorithm against the existing algorithm and expressed that the method yielded good Detection Rate (DR) which reduced the false positive rate, as shown in table 3 and figure 7. The Detection Rate (DR) is computed using the formula in figure 6.

$$DR = \frac{\text{TruePositive Rate}}{\text{FalseNegative Rate} + \text{TruePositive Rate}}$$

Figure 6 Detection Rate formula (Pal and Parashar 2014).

Type	Proposed Algorithm		Existing Algorithm	
	Detection Rate	False Positive	Detection Rate	False Positive
Normal	96.86	3.1	93.64	6.13
Attack	97.46	2.5	91.88	6.72

Table 3 Results of experiment (Pal and Parashar 2014).

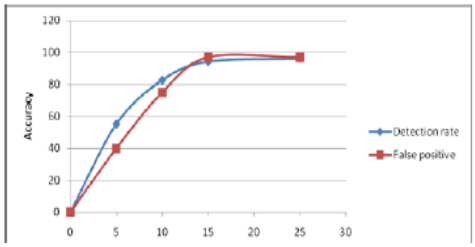


Figure 7 Graph showing Detection and False Positive rate (Pal and Parashar 2014).

The authors of this research have a clearer picture of what they wanted to achieve as it reflects in their investigation effort and the experiments conducted. However working with actual network traffic would have greatly strengthened their study because the dataset provided by MIT launch labs (Well known as KDD'99) no longer reflects current attacks and that may make the results seem bias.

The testing conducted seems to be well thought out because the authors actually compared and contrasted the proposed algorithm against an existing algorithm, which makes the results credible. In spite of this they only tested their method once. Experiments conducted several times would have ensured that mistakes and errors were accounted for to solidify their conclusion.

Pal and Parashar (2014) expressed that their system can be used in real time manner, but the authors not confirming that it has indeed been tested in real time may make the results appear bias however this can allow other authors to carry on with the research and confirm if the method is proficient to work in the real world.

2.3 K-Means

Is one of the well-known unsupervised clustering algorithm that divides data into K clusters and makes sure that the data within the clusters are of the same kind while data in different clusters have zero or low similarities. That's how this technique is able to differentiant between malicious data and normal data.

Singhly (2015) proposed a K-means algorithm for anomaly detection and states that it improved accuracy. The author says firstly the proposed algorithm uses a metric method to reduce noise and then isolates points in the data set. This step alone reduces chances of a false positive alarm to be triggered as noise can be mistaken for an attack.

In the proposed model, the author says the KDD'99 dataset was used to conduct experiments and the data was normalized, randomized and fuzzified to bring it to a nominal scale between 0 and 1. After that process, features were partitioned into training and testing dataset (Singhly 2015). The proposed algorithm was compared against another machine learning algorithm called maximum probability (EM) and the results are presented in figure 8.

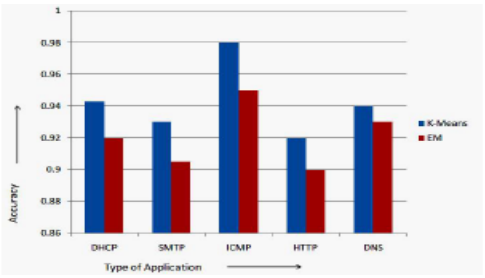


Figure 8 Comparison of accuracy of K-means and Em (Singhly 2015).

The author of this paper concluded that the proposed K-means algorithm performed better compared to EM.

The hypothesis and theories of this paper are not of good science because the author did not mention anything concerning how the conclusion was reached. This paper makes it hard to evaluate because all the necessary information was not noted.

2.4 Back Propagation and K-Means Algorithm

Different researchers say the human ability to think, reason and learn can be replicated in computers. Rashidan et. al. (2014) explains that ANN is a tool that is able to process a series of data and represent complicated relationship in a way similar to a human being.

Sen et. al. (2014) employed the idea and developed an architecture using Back Propagation Neural Network (BPNN) to detect anomaly that produces low false positive rates and high detection rates. The authors say BPNN is a unique type of ANN which consists of multiple layers containing interrelated nodes and an activation function.

In their proposed model, the authors specify that they used the KDD'99 dataset to conduct experiments and it was compared against other models. Modification was done to the KDD dataset to bring it to a nominal scale between -1 and 1. Then the new dataset was randomized, normalized and trained using 1000epochs. A matrix to calculate accuracy, detection and false positive rate was formulated (Sen et. al. 2014).

The BPNN architecture was implemented in the C Language and the program was executed on a Linux Operating System. The graph in figure 9 was extracted from table 4 and table 5 presented below.

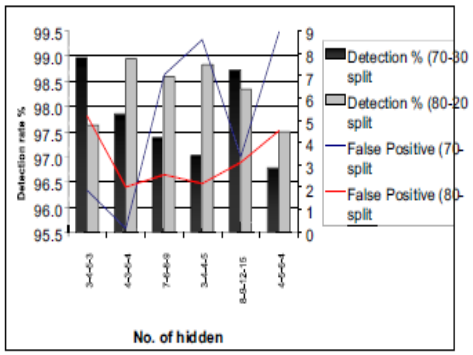


Figure 9 Testing rate obtained in different hidden layers (Sen et. al. 2014)

No. of hidden layer	No. of hidden nodes	Learning factor	MSE	Training			Testing			Execution time (sec)
				False Positive	Accuracy	Detection %	False Positive	Accuracy	Detection %	
4	3-4-5-3	0.015	0.016081	0.32	0.98	99.58	1.85	0.97	98.97	156.38
4	4-3-5-4	0.04	0.013247	0.38	0.99	99.30	0.14	0.97	97.84	138.24
4	7-6-8-9	0.04	0.005879	0.47	0.99	99.12	7.06	0.96	97.37	305.39
4	3-4-4-5	0.07	0.007967	0.91	0.99	98.29	8.61	0.95	97.02	124.92
4	8-9-12-15	0.035	0.006093	0.70	0.99	99.05	3.39	0.96	98.72	524.3
4	4-5-6-4	0.09	0.008846	0.97	0.99	98.18	8.97	0.96	96.77	165.89

Table 4 Dataset spilt with 70% for training & 30% for testing using 1000epochs (Sen et. al. 2014).

No. of hidden layer	No. of hidden nodes	Leaning Rate	MSE	Training			Testing			Execution time (sec)
				False Positive	Accuracy	Detection %	False Positive	Accuracy	Detection %	
4	3-4-5-3	0.07	0.007433	1.15	0.99	98.43	5.20	0.95	97.62	137.8
4	4-3-5-4	0.015	0.009033	0.47	0.99	99.37	2.01	0.96	98.94	158.76
4	7-6-8-9	0.09	0.005882	0.62	0.99	99.16	2.58	0.96	98.59	343.58
4	3-4-4-5	0.001	0.012686	0.63	0.99	99.16	2.16	0.96	98.84	144.28
4	8-9-12-15	0.034	0.00403	0.56	0.99	99.24	3.10	0.95	98.34	566.6
4	4-5-6-4	0.09	0.011712	1.31	0.99	98.21	4.53	0.95	97.50	188.51

Table 5 Dataset spilt with 80% for training & 20% for testing using 1000eponchs (Sen et. al. 2014)

The authors say the best detection rate was obtained in the 70-30% dataset with the 3-4-5-3 node combination, 8-9-12-15 is another good node combination. They further described that the best low false positive rate is in node combination 4-3-5-4. Therefore it is concluded that 3-4-5-4 node combination is of optimum results and it is the ideal node.

The authors of this paper stated that their model was compared against other models however the results were not presented. When conducting a comparative study it is imperative to incorporate the results of the other models in your study to prevent your effort from biasness.

The ability to interpret and present results is one of the most important aspects. In this paper the authors stated that their aim was to develop a model that will detect anomaly and produce low false positive rates and high detection rates of different categories of datasets. However they concluded that the node combination 3-4-5-4 produced optimal results but the graph shows that the node combination did not produce low false positives compared to other combinations rather it had the best detection rate which certainly does not meet the aim of their study. Furthermore two splits where used (70-30 and 80-20) yet in the author's conclusion only one split was discussed, which makes the result seem biased.

The criterion for choosing feature selection is not presented as well. It is important to produce and present all the important aspects so that when a different author carries on ones work they don't have to deal with missing critical data.

The validity is not well-founded because the KDD'99 dataset used no longer reflects on contemporary attacks; this may make the results seem bias. However implementing the BPNN architecture in a C Language on a Linux operating system may make the model feasible so other researchers can replicate the results and try to perfect it.

3 Conclusions

Machine Learning based anomaly detection techniques are discussed in this review paper. It consists of supervised and unsupervised techniques.

From the analysis of different surveyed researches presented above, it is certain that no technique can lessen the rate of false positives by a great margin as limitations are greatly encountered. There appears to be a common trend amongst all the experimental results presented that is the use of the DAPRA/Lincoln Labs packet traces and the KDD datasets. These datasets are not sufficient anymore as they represent data that was generated in 1998 and todays attacks are ever changing so this makes the results to come across as biased and insignificant.

It is well-known that network traffic is 'dynamic' it changes in a matter of seconds and minutes meaning an attack can occur within that time frame so it is important that techniques are trained with different data to validate that the system can adopt in such settings however in the experiments presented none of the authors took note of this which raises doubts regarding the methods.

It is concluded that the genetic algorithm by (Pal and Parashar 2014) is a viable method for the detection of malicious intrusions that needs to be further analyzed. The authors of this research paper show a lot of real world potential, with the use of logical diagrams and algebraic equations to back up their claims on the methods' functionality.

The comparison between various learning techniques will allow experts to find the best machine learning technique that is well-defined, more effectively and efficiently.

References

Gondal M., Makik A. and Kham F., 2015 'Network Intrusion Detection using Diversity-based Centroid Mechanism,' *IEEE 12th International Conference on Information*, Volume 7, Issue 1, pages 150-155.

Liu P., Gumus F., Sakar C., Erdem Z. and Kursun O., 2014 'Online Naïve Baiyes Classification for Network Intrusion Detection' *IEEE ACM International Conference on Advanced IN Social Network Analysis and Mining*, Volume 4, pages 670-674.

Pal D. and Parashar A., 2014 'Improved Genetic Algorithm for Intrusion Detection System', *IEEE Fourth International conference on Computer Studies*, Volume 13, pages 835-839.

Rashidan S., Signh Kushwah R. and Kingsway C., 2015 'A study on Intrusion Detection in Wireless Networks by Using Genetic Algorithm Applications' *IEEE Sixth International Conference on Computational Intelligence and Communication Networks*, Volume 2, pages 759-752.

Sen N., Hope R. and Chattopadhyay M., 2014 'An effective Back Propagation Neural Network Architecture for the Development of an Efficient Anomaly Based Intrusion Detection Systems' *IEEE Sixth International Conference on Computational Intelligence and Communication Networks*, Volume 20, pages 130-134.

Singhly H., 2015 'Performance analysis of unsupervised machine learning techniques for network traffic classification' *IEEE International Conference on Computing Concepts*, Volume 6, pages 23-26.

Xiao L., Chen Y. and Chang C., 2015 'Bayesian Model Averaging of Bayesian Network Classifiers for Intrusion Detection' *IEEE Annual International Computer Software and Application Conference Workshop,* Volume 13, pages 128-133.

Xiaxong X. and Ning R., 2014 'The Study of Cyber Crimes through the Internet and Solutions for such Attacks' *IEEE Sixth International Conference on Networking*, Volume 20, pages 75-80.

Xu C. and Srihari M., 2014 'Bayesian Network Graphical Mode against Intrusion Detection' *IEEE 10th International Conference on Networking*, pages 19-25.

A Critical Evaluation of Current Research on Techniques Aimed at Improving Search Efficiency over Encrypted Cloud Data

Kgosi Dickson

Abstract

The scalability, reliability and consistency of data in the cloud motivated a lot of data owners to outsource their data to the cloud servers. These includes very sensitive data that requires great security, meaning that the data had to be encrypted before being outsourced. Encryption of cloud data has now led to difficulties in accurate data searches. To tackle the aforementioned issue, a lot of researchers proposed different search techniques of which this paper critically analyzes and evaluate to further investigate if they indeed improve search efficiency over the encrypted cloud data. Focus will be on the three most popular techniques; Latent Semantic Analysis (LSA)-based multi-keyword ranked search which supports semantic multi-keyword ranked search, Efficient Multi-keyword Ranked Search (EMRS) and Privacy Preserving Multi-keyword Top-k Retrieval search scheme. The methods and experiments carried out by authors are evaluated, including their claims. At the end of this paper, recommendations on the most suitable search technique(s) are made basing on the evaluation results of this techniques.

1 Introduction

For the past decade, researchers have been toe to toe with the sole mandate of finding the most feasible way to search over encrypted cloud data efficiently, without compromising its security. The causal effect of all this is the fact that, in cloud computing, data owners resorted to encrypting their data, for security and confidentiality purposes, before outsourcing it to the cloud servers.

The main problem is that, the complexity of data encryptions available, make it very difficult to do normal searches over data like in normal plaintext because of the need to perform high level computations, before relevant documents can be retrieved. Thus, balancing search efficiency and data security/privacy in data retrieval proves to be very challenging.

A breakthrough in this research doses not only mean secure, effective and efficient searches over the encrypted cloud, but also gives birth to a very practical and economic data storage and retrieval on the cloud. It will mean saving time and resources on the users' side who trying to retrieve the exact data they want from the cloud servers.

Li et al., (2013) claim that currently, search queries in searchable encrption techniques suffer from returning less accurate results and most are singularly restricted to either boolean searches, single keyword searchable encyption or fuzzy keyword searches.

Goplani et al., (2015) base their research on reviewing existing techniques for searching over encrypted cloud data in order to highlight the most efficient methods, and investigate the possibility of integrating [some of] them to get the best possible search results accuracy and highest relevance ranking possible.

This paper critically evaluates three mostly proposed and widely known searchable techniques. Firstly, it critically looks at the methods used, experiments, tests results and conclusions made by the authors to investigate if they carry sound scientific value and rigor. Secondly, the proposed techniques are compared against each other. Lastly, sound arguments on the implications of applying these methods in the wide world are provided.

2 Current Searchable Encyption Techniques

This section evaluates three techniques aimed at achieving the main goal of search efficiciency over encrypted cloud data. Since search efficiency directly affects data security, every method will also be investigted if it does not compromise cloud data security as it seeks to achieve the best possible search efficiency. The paper evaluates methods using computation overheard, communication overheard and functionality, as they are mostly believed to make up a better improved search efficiency when all achieved.

2.1 Latent Semantic Analysis (LSA)-Based Multi-Keyword Ranked Search

Chen et al., (2014) proposed a LSA-based multi-keyword ranked search technique that utilizes the semantic structure over encrypted cloud data to retrieve the most relevant data, and not only the exact data searched but also the data semantically related to the queried data. The algorithm is made up of two vital algorithms (semantic analysis and k nearest neighbor algorithms).

The first algorithm, being latent semantic analysis, supports multi-keyword queries and performs result ranking through the use of singular-value decomposition (SVD). The SVD is used to compute the relationship between search terms and documents (Chen et. al., 2014, pg. 326). Vectors made up of Term Frequency values, which are used as indexes of documents, are used in analyzing a large matrix of term-document connotation data. This large matrix is used to build a reduced semantic vector space in which terms and documents closely related are placed near one another. The second algorithm, k-NN, was used to calculate the inner product so to achieve relevant ranked results in a way that does not compromise data privacy. It computes the distance between a database point and a query point with the goal of finding the neighbors to the query point.

Experiments were carried out on a real MED dataset, in which 200 documents were used for testing purposes. A computer running Windows 7 with Core 2.83GHz Processor was used and the experiment was implemeted in C++ language. The F-measure was used to weigh the results of the experiments;
F = 2(precision×recall) ÷ (precision + recall)

A graph comparing the proposed LSA-MRSE against the original MRSE was produced in which two variables were used, number of documents against the F-measure. The two techniques were compared to measure their performace in producing accurate and relevant ranked results.

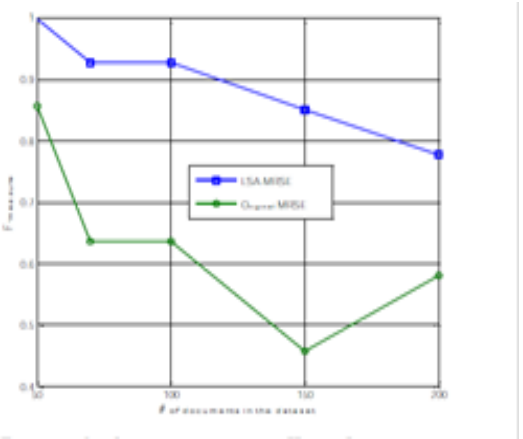


Figure 5 Performance comparison of the LSA-MRSE and the Original MRSE techniques (Chen et al., 2014)

The claim made was that the proposed scheme achieves higher scores than the original MRSE when using the F-measure since the proposed technique returns all relevant documents (including those semantically related to the search terms) when compared to the Original MRSE that returns only exact maches queried terms.

The technique proposed by the authors builds upon already existing methods as they already highlited that authors like Deerwester (1990) and Cao (2011) adapted the methods before, which proves that the technique has credibility. There is scientifc merit achieved through the use of pre-existing techniques for performance comparison even though it would have been more reliable if more than one or two pre-existing techniques were used for comparison with the proposed one. They also stated that their algorithm does not compromise security of the queried data through the use of k-NN algorithm, but no experiments were carried out to back this claim. They just laid out the algorithm without supporting it with tests.

The tests results were presented very well but the questionable part is the fact that test were only carried on only one platform (Windows 7) and on a 2.83GHz processor. The significance of the results remain a problem because they were carried out on only one platform, instead of different platforms, which would have yielded reliable results that would then be compared against each other to better investigate the efficiency of the technique.

Nevertheless, the paper made great adavancements since the authors managed to make improvements on the overall recall rate and search results return what the user searched for or results relating to the search query. E.g a user search for “internet” also gets results for

"network" which can be bad news for a particular user who only want only the exact results from search quey or good news for one who was not sure of the query phrase they used.

Concluding, more experiments need to be carried out to further give more validity and reliability to the results, most especially the security part which does not have any experiments and test explicitly provided. More work need to be done in providing a technique that allows for multi-user environment and synonym queries.

2.2 Efficient Multi-Keyword Ranked Search

Li et al., (2015) proposed a searchable encryption called Efficient Multi-keyword Ranked Search (EMRS) scheme which uses blind storage to conceal search user access pattern over encrypted mobile cloud data. It also incorporates the k-NN and relevance score techniques to achieve improved accurate and ranked search results.

The authors point out that to achieve great search efficiency and accuracy, and taking query privacy into consideration, a search technique is required to support multi-keyword searches, provide relevance-based order ranked search results based on the search query. In addition, the technique should be able to return search results with minimum delays, no matter how large the searched cloud database is.

Therefore, Li et al., (2015) designed their technique in a way that when a user searches the cloud server, they receive a secrete key from the data owner which they use to compute a trapdoor that includes a keyword-related token (stag), encrypted query vector and key k (symmetric encryption algorithm). The cloud server use the aforementioned variables to access index F and calculates relevant scores using the query vector before returning descriptors of relevant top-k files. Descriptors are used to return the encrypted files in the blind storage system. CP-ABE is used to make decryption of the files possible for the searcher.

Extensive test were carried out to investigate the efficiency of the algorithm. The authors compared their proposed algorithm against those of Cash et al., (2013); Cao et al., (2014) and Naveed et al., (2014) on functionality, communication and computation overhead since the authors concluded that they are all vital for the high performance, efficiency and robustness of the technique. Firstly, Comparisons on multi-keyword, Result Ranking and Relevance Scoring functionalities are made and results of these functionalities are shown if Figure 2 below.

	[10]	[11]	[13]	EMRS
Multi-keyword	√	√		√
Result Ranking		√		√
Relevance Scoring		√		√

Table 6 Comparison basing of Functionalities (Li et. al., 2015)

Documents in the National Science Foundation Research Awards Abstracts dataset of 1990-2003 were randomely selected and used to do test on the computation overhead of the proposed technique on a 2.8GHz processor computer. Trapdoor generation were carried out on a 1.2GHz smart phone. Results of the test are shown below.

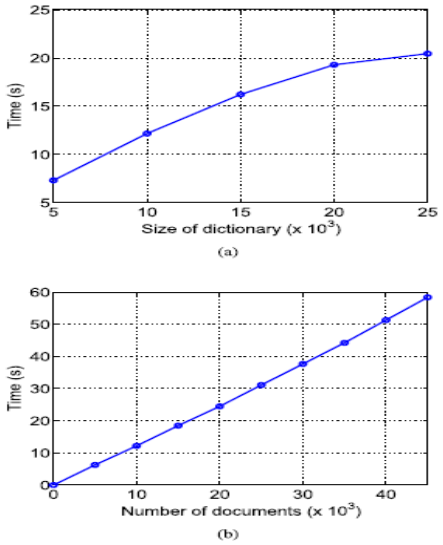


Figure 6 Time cost of calculating the relevance score (a) different dictionary sizes with the same # of documents (m=1000) (b) different # of documents with same size of dictionary (Li et al., 2015)

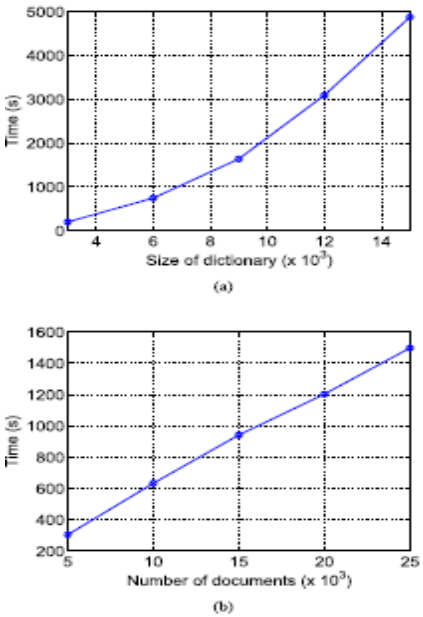


Figure 7 time cost for calculating the encrypted relevance vectors (Li et al., 2015)

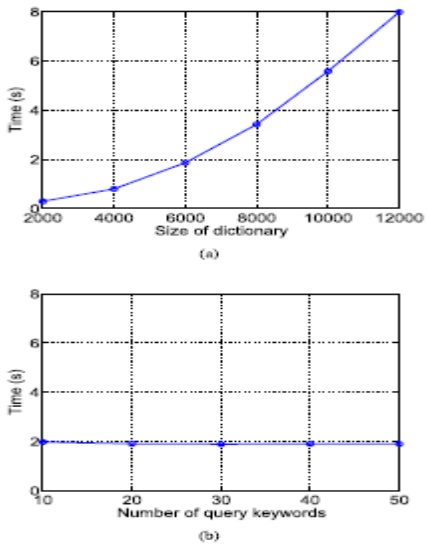


Figure 8 time cost for calculating the trapdoor on a real smart phone (Li et al., 2015)

As for search efficiency, Li et. al., (2015) compared their technique against that of Cao et al., (2014). Figure 6 below depicts the results.

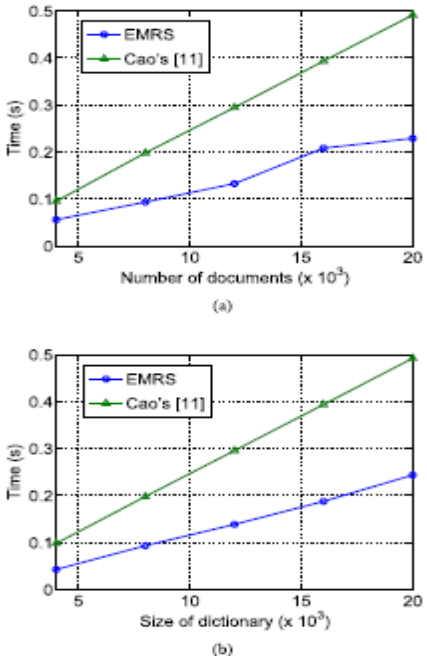


Figure 9 Time cost for searching on the cloud server (a) different # of documents, same dictionary size and same number of search keywords (b) different dictionary sizes, same number of documents and same number of searched keywords (Li et al., 2015)

	Number of Rounds	Size of Message(bit)
Cash's [10]	many	$\varrho\|\varpi\|\|G\|$
Cao's [11]	one	$2(d+2)\eta_q$
Naveed's [13]	one/two	$\alpha * size_\omega * n_b$
EMRS	one/two	$2(d+2)\eta_q + 2\alpha * size_\omega * n_b$

Table 7 Communication overhead results (Li et al., 2015)

The authors concluded that their technique performs better than the existing methods in terms of functionality and computation overhead, which means that their method gives the highest search efficiency possible as depicted in the experiments they conducted.

Considering the well-articulated requirements of a good searchable encryption technique, clearly stated and relevant variables under study as well as a properly laid out methodology of executing the set requirements proves that the authors' method is very valid and reliable.

The work done by the authors allow for reproducibility because they provided all the necessary variables, steps and environment for undertaking the experiments. They show all their results in both graphs and tables which shows reliability in their work as there transparency in their work.

The sample size of the data used for testing purposes was adequate with the authors clearly showing the number of documents and size of dictionary used for each test. An adequate number of queried keywords are also provided in search efficiency. All this proves how scientific sound the tests are. By splitting the algorithm into small tasks, they were able to test for every aspect of the algorithm, which are search efficiency, query privacy/security and multi-keyword ranked search.

Since a balance between search efficiency and security is required, authors have to show that their methods do not compromise security in their quest to give the most search efficient method. The authors however claimed that their technique gives the best security when compared with existing techniques, unfortunately, no experiments were conducted to back this claim. Therefore, further experiments on the security analysis of the technique need to be carried out and explicitly shown in the publication because none were shown, only results of the technique's security performance were displayed.

It can be concluded that further experiments need to be carried out on security part of the method as well as search efficiency (using even larger datasets), which is the main goal of the algorithm, to prove that indeed the results can be truly relied on. To even bring higher credibility, the experiments should be carried out by an external body to eliminate bias.

2.3 Privacy-Preserving Multi-Keyword Top-K Retrieval Search over Encrypted Cloud Data

This improved technique was proposed by Aashi & Bhaggiraj (2015) with the aim of solving the problem of balancing search privacy and multi-keyword ranked search over encrypted cloud data so as to come up with a highly efficient search technique.

The technique incorporates the "coordinate matching" principle which makes use of inner product similarity. The inner product similarity computes the number of query keywords appearing in a document to calculate the similarity of that particular document to the search query. Document ranking is then performed through Trapdoor generation, relevance score calculation and results ranking respectively. The user then uses the TOPLSELECT algorithm to retrieve the most relevant files according to their search query.

Aashi & Bhaggiraj (2015) carried out tests to verify their proposed method. Firstly, they carried tests on the time cost of sending a search query on different file sizes. Secondly, they experimented for computation overhead since they believed that a lower computation overhead proves the method to be effective. Lastly, they tested for storage overhead; time cost to store and retrieve the stored data. They compared their results with the original MRSE 1. Test results are show below.

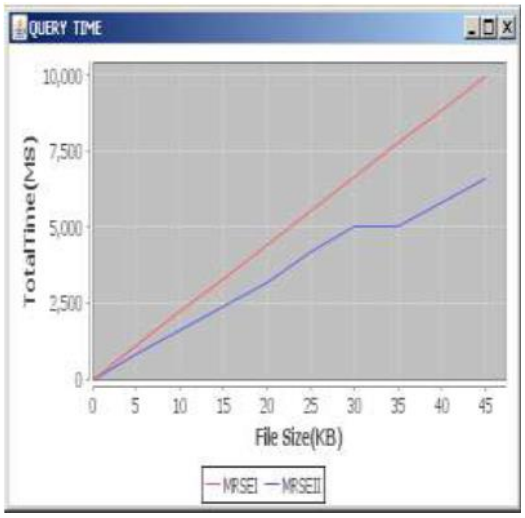


Figure 10 Query time (Aashi & Bhaggiraj, 2015)

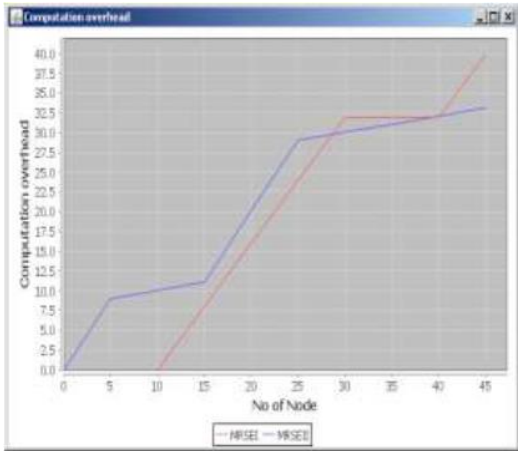


Figure 11 Computation Overhead (Aashi & Bhaggiraj, 2015)

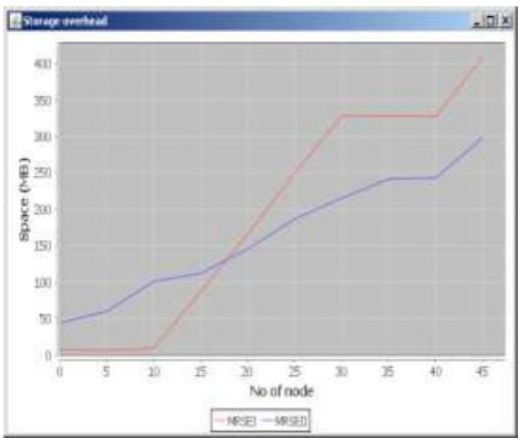


Figure 12 Storage Overhead (Aashi & Bhaggiraj, 2015)

The authors concluded that their method achieves low computation and communication overhead after performing thorough experiments on real-world datasets.

The proposed method lacks scientific merit in the sense that it was only evaluated against its original version (MRSE 1) during experimentation, meaning that it would not be valid to conclude that it achieves the highest search efficiency. The authors also claim that their method has low communication overhead but no tests were carried out to justify the claim, not even communication overhead results were presented.

The dataset used for testing the technique was not adequate enough to claim that the method excels because only up to 45kb of data files was used. The whole experiment lacks reproducibility because the authors did not provide adequate algorithms used in their technique, only one algorithm for TOPSELECT was presented, and no resources used for carrying out the experiments were presented.

In conclusion, further experimentation needs to be done on large scale datasets to justify the authors' claims of low communication and computation overhead. The proposed method also needs to be evaluated against other searchable encryption techniques to validate the experimental results. Variables that might have effect on the performance of the technique like size of dictionary and number of query keywords need to be explicitly covered in the experiments.

3 Comparison of Searchable Encryption Techniques

The Latent Semantic Analysis algorithm consumes less time in data retrieval but does not show much improvement from its original MRSE algorithm, which does not necessarily mean bad news. The use of SVD for computing relationship between search terms and documents consumes time in data retrieval beside the fact that it helps greatly in retrieval of relevant documents. The only great improvement it comes with is the high precision and recall accuracy in data retrieval which it achieves through the use of semantic analysis. It is greatly affected by the size of the dataset; a very large latent semantic space results in poor precision and recall accuracy.

Compared with EMRS and Top-K, EMRS achieves the highest search efficiency with minimum delays, no matter how large the dataset is, see Figure 6 compared to LSA algorithm which is highly affected by the size of the dataset. By implementing the inverted index *F* from the blind storage, EMRS achieves improved computation overhead leading to improved search efficiency.

Looking at the above results by Aashi & Bhaggiraj (2015), the top-k algorithm achieves the greatest search efficiency through very low query time and computation overhead. The time taken to perform trapdoor generation, relevance score and document relevance ranking are very low but compared to other aforementioned techniques, this method used very small datasets. Therefore, more experiments are required to further validate the robustness of this method.

In conclusion, a few methods from this algorithms like semantic analysis, TOPSELECT, k-NN etc. can be hybridized to come up with more robust algorithm that provides the highest search efficiency with compromising search query privacy, confidentiality or integrity.

4 Implications of Applying Searchable Encryption Techniques in the Real World

Considering the need by businesses to provide effective and efficient searches over encrypted cloud data without compromising search query privacy, implementation of this methods would come as a great achievement to organizations and their customers. However, more work need to be done first in experimentation of this methods to validate their efficiency before they can be implemented.

All the three techniques requires a programmer to at least have insight on encryption and decryption algorithms, sorting and searching algorithms as well as

the cloud infrastructure. This could create jobs directed towards teaching this skills. Method by Chen et al., (2014) requires a programmer to have knowledge on Singular Value Decomposition. Other than that, the methods are not that complicated for a programmer who possess the aforementioned skills.

5 Conclusions

This paper has critically reviewed three of the most proposed searchable encryption techniques with regard to their performance and security but mostly directed towards their search efficiency. Three authors were picked and from the literature they presented, it was mostly theoretical than practical. All showed that improved search efficiency can be achieved but further experiments are required to really validate their proposed schemes as well as their claims.

Looking at the evaluated techniques, it is difficult to draw solid conclusions on which of the methods provides better search efficiency without compromising security because not enough experiments were carried out give clear conclusions. Performance tests were carried out on different platforms with mostly on different variables which makes comparisons on computation time difficult.

Nevertheless, experiments and tests carried out by Li et al., (2015) carried more rigor because they layed out very well with enough and relevant variables e.g Number of documents against time, Size of dictionary against time. The problem is that, not enough experiments were carried out on the security aspect of the technique and the dataset need to be increased to further bring credibility to the results. Therefore more experiments are required on the method. The techniques by Aashi & Bhaggiraj (2015) and Chen et al., (2014) however, propose very good theories but lacks greatly experimental wise. More experiments are needed to support the authors' claims.

Nevertheless, different components of all the three proposed techniques can be brought together to achieve much improved search efficiency without compromising query privacy.

References

Aashi, Q. H. S. & Bhaggiraj, S., 2015. 'Improving privacy multi-keyword top-k retrieval search over encrypted cloud data.' *International Journal of Engineering and Computer Science,* 4(4), pp. 11385-11390.

Cao, N., 2011. 'Privacy-preserving multikeyword ranked search over encrypted cloud data.' s.l., INFOCOM, *Proceedings IEEE, Shanghai*, pp. 829-837, 10-11 April

Cao, N., Wang, C., Ming, L., Ren, K., Lou, W., 2014. 'Privacy-preserving multi-keyword ranked search over encrypted cloud data.' *IEEE transactions on parallel and distributed systems,* 25(1), pp. 222-233.

Cash, D., Jarecki, S., Jutla, C., Krawczyk, H., Rosu, C. K., Steiner, M., 2013. 'Highly-scalable searchable symmetric encryption with support for Boolean queries.' *International Journal for Cryptologic Research ,* 8042(1), pp. 353-373.

Chen, L., Sung, X., Xia, Z. & Liu, Q., 2014. 'An efficient and privacy-preserving semantic multi-keyword ranked search over encrypted cloud data.' *International Journal of Security and Its Applications,* 8(2), pp. 323-332.

Deerwester, C. S., 1990. Indexing by latent semantic analysis. *JASIS,* 41(6), pp. 391-407.

Goplani, A., Vaswani, J., Kukreja, S. & Anjali, Y., 2015. 'A review on techniques for searching and indexing over encrypted cloud data.' *International Journal of Emerging Technology and Advanced Engineering,* 5(1), pp. 523-532.

Li, H., Liu, D., Dai, Y., Luan, H. T., Shen, S. X., 2015. 'Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage.' *IEEE transactions on emerging topics in computing,* 3(1), pp. 127-138.

Li, M., Yu, S., Ren, K., Lou, W., Hou, T. Y., 2013. 'Toward privacy-assured and searchable cloud data storage services.' *IEEE Transaction Network*, 27(4), pp. 56-62

A Critical Analysis and Evaluation of Current Research on Credit Card Fraud Detection Methods

Lebogang Otto Gaboitaolelwe

Abstract

Various credit card fraud detection methods are being used in the financial sector to determine fraudulent transactions. Researchers have however, had to come to the fore and propose credit card methods that are more effective than those currently used in industry. This paper provides a critical review of current research on credit card fraud detection methods that have been proposed by various researchers, including discussions on the experiments carried, their validity and implications of the results. Ultimately, the paper provides a comparison of the work carried out gives guidance for future research into the area.

1 Introduction

Nowadays the credit card is one of the most popular methods of payments. People use the credit card to pay for different services or goods, from simply buying groceries, paying for food at the drive in, to booking for a vacation at some exotic location. Reardon et. al. (2012) provides a potent explanation as to why use of the credit card has become so popular, the researcher attributes this to the need of people to have access to their 'monetary assets' at any moment they may require, a convenience provided by the credit card. However this exponential growth in use credit cards has presented an opportunity for criminals to defraud users of their money using various methods, these include skimming, phishing, counterfeit credit cards, unethical behavior from employees of a credit company and simple theft, as advanced by (Duman et. al. 2013). Fraud has many negative consequences, according to (West and Bhattacharya 2016), fraud has the potential to reduce consumer confidence, destabilise the economy and adversely affect cost of living

Pavía et. al. (2012) posits that, because of the credit card fraud problem, many users are interested in methods that provide early alerts and some sort of prevention. To this end, researchers and credit companies have had to come up with such. Ghazali et. al. (2014) suggests that solutions aimed at both fraud prevention and election should have, beforehand considered the primary causes of fraud and current methods aimed at fraud detection and prevention .This paper will explore several methods that have been proposed by various researchers. Alowais and Soon (2012) explore the feasibility of methods known as personalized and aggregated models. Wong et. al. (2012) discusses the use of artificial immune systems for countering credit card fraud, the method is augmented in various ways out find the most optimal augmentation. Duman and Ozcelik (2013) propose and present a fraud detection method that uses genetic algorithm and scatter search. Mishra et. al. (2013) propose the application of the hidden markov model in solving the problem.

The rest of this paper consists of two sections; Detection methods section which will discuss and evaluate methods and ideas proposed by various authors. The final section is the conclusion, which provides the summary of the discussion and gives reference for future research

2 Current Credit Card Detection Methods

This section reviews four credit card fraud methods that have been proposed by different researchers. The proposals are discussed from various context, such as the way the method operates, validity of the experiment and implication of the results.

2.1 Personalised Model (and Aggregated Models)

Alowais and Soon (2012) carried out an empirical study to explore the feasibility of a method known as the personalised model, in detecting credit card fraud comparative to aggregated models. This is a prediction model created using transaction history of an individual whilst aggregated model is built using transaction history of different individuals. In order to carry out the experiment, the credit card history of three individ-

uals was obtained in order to build the data sets, participants are also asked to provide answers to an online questionnaire, which requests demographic data, the different transactions they make on the internet using credit card information and transaction they usually make at physical point of sale where they are present. The transaction dataset comprises of details about purchases done through a customer's credit card, such as amount spent, time, date and location.

Personalised models were created for each of the three participants and an aggregated model referred to as Real Transaction Aggregated Model using the same data, an alternate aggregated model called Questionnaire Transaction Model is also built using data obtained from the questionnaire (the aggregated models are built differently).The table below shows the training data sets used to build the different models

Datasets	Number of legitimate transaction	Number of fraud transaction	Total number of instances
Individual Model 1	91	10	101
Individual Model 2	80	18	98
Individual Model 3	151	20	171
QTAM	4003	921	4924
RTAM	318	48	366

Table 8. Displays dataset details (Alowais and Soon, 2012).

The data sets are constructed using answers given by the actual owners of the credit cards. It must be noted that the research tested the methods in relation to; its accuracy, precision, recall, F measure, TP and FP. Results of the experiment are captured in the tables below, they show the performance of both the personalized and aggregated models respectively. Two kinds of classifiers; random forest and naïve beyes were both used to determine which of the classifier improved the effectiveness of the method.

Models	Personalized Model 1		Personalized Model 2		Personalized Model 3	
	RF	NB	RF	NB	RF	NB
Accuracy (%)	91.09	**96.04**	86.17	**96.81**	89.47	**95.91**
Precision	0.919	**0.959**	0.882	**0.969**	0.881	**0.958**
Recall	0.911	**0.96**	0.862	**0.968**	0.895	**0.959**
F-Measure	0.877	**0.958**	0.828	**0.967**	0.862	**0.959**
TP	0.911	**0.96**	0.862	**0.968**	0.895	**0.959**
FP	0.811	**0.271**	0.584	**0.135**	0.751	**0.179**

Table 9. Represents Personalised Model Performance, (Alowais and Soon, 2012).

The tables below shows how both the Real Aggregated Model and Questionnaire Transaction Aggregated Model

Metrics	Random Forest	Naïve Bayes
Accuracy (%)	**96.18**	95.08
Precision	**0.96**	0.95
Recall	**0.96**	0.95
F-Measure	**0.96**	0.95
TP	**0.96**	0.95
FP	**0.2**	0.18

Table 10. Represents performance of Real Transaction Aggregated Model, (Alowais and Soon, 2012).

Metrics	Random Forest	Naïve Bayes
Accuracy (%)	**89.48**	84.08
Precision	**0.889**	0.849
Recall	**0.895**	0.841
F-Measure	**0.889**	0.844
TP	**0.895**	0.841
FP	**0.333**	0.316

Table 11. Represents Performance of Questionnaire Transaction Aggregated Model, (Alowais and Soon, 2012).

The results show that the personalised (method performed better when using the naïve Bayes classifier whilst both aggregated models used performed better using random forest classifier. The experiment also shows that the aggregated models perform better than the personalised model. However with the dataset used being simply too small for the results may not be conclusive, the researcher should have considered that a larger dataset may possibly give different results. The assumption is that a practical credit card detection method would have to deal with very large dataset. Another concern is in relation to the logic of the method, the method is premised on an individual(or individuals) having some sort of transactional history

using their credit card, this therefore means that the method cannot immediately work for a new user who has just been issued with a credit card. This is a fatal vulnerability of the method, it would leave new user exposed to attack, and therefore the method in its current design offers no absolute solutions. In relation to tools used, (Alowais and Soon 2012) used the latest technologies and seemingly reasonable computer memory to carry out the experiment, the i5 processor is a new technology in the computer hardware market; this offers a realistic view of whether the method is practical.

2.2 Application of Artificial Immune Systems for Credit Card Fraud Detection

Wong et. al. (2012) proposes the use of artificial immune system to detect credit card fraud, the method applies biological principles in order to achieve this. It attempts to detect fraud in the same way biological organisms detect attacks to their bodies. The diagram below displays components that make up the system.

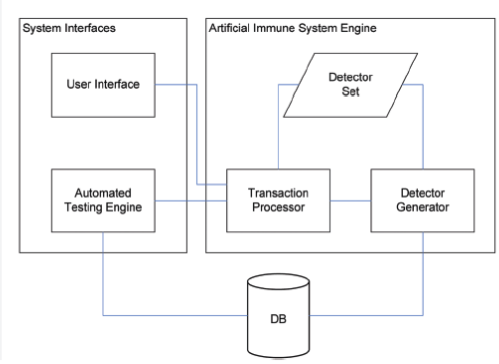


Figure 13 Displays Artificial Immune System for Credit Card Fraud Model, (Wong et al 2012).

The model Artificial Immune System Engine consist of two subsystems; detector generator and the transaction processor, the transaction processor is responsible for processing credit card transactions and determining the legitimacy of a transaction whilst the detector is responsible for generating new memory detectors for anomalous transactions that are already defined, this concept is similar to vaccination, it also generates new detectors using negative selection. The system interface allows users to interact with the engine in order to input transactions that are to be catagorised as either anomalous or genuine transactions. According to (Wong et. al. 2012) the systems should ideally achieve a high detection rate for anomalous transactions and a low false positive rate (in the context of the research, a false positive represents a legitimate transaction that has been inaccurately identified as anomalous). An experiment was carried out against this aim, this was done by carrying out several simulations in a computer laboratory using actual credit card transaction data. The simulation mimicked the interaction a user would make in a real system, performance statistics were captured by the database for later analysis. The method was tested using various configurations, for example there is the basic configuration of the method, in other scenarios, and certain features such as mutation are enabled. Table below displays results from the experiment.

Anomaly type	transactions	negatives	positives	rate (%)
Test run 1 (basic representation)				
Lost card fraud	17	16	1	5.9
Stolen card fraud	36	31	5	13.9
Skimmed fraud	37	27	10	27.0
Mail/phone fraud	74	69	5	6.8
Test run 2 (basic representation – mutation enabled)				
Lost card fraud	17	16	1	5.9
Stolen card fraud	36	28	8	22.2
Skimmed fraud	37	27	10	27.0
Mail/phone fraud	74	69	5	6.8
Test run 3 (basic representation – vaccination enabled)				
Lost card fraud	17	14	3	17.6
Stolen card fraud	36	17	19	52.8
Skimmed fraud	37	16	21	56.8
Mail/phone fraud	74	43	31	41.9
Test run 4 (basic representation – mutation enabled – vaccination enabled)				
Lost card fraud	17	13	4	23.5
Stolen card fraud	36	23	13	36.1
Skimmed fraud	37	16	21	56.8
Mail/phone fraud	74	42	32	43.2
Test run 5 (improved representation)				
Lost card fraud	17	12	5	29.4
Stolen card fraud	36	20	16	44.4
Skimmed fraud	37	21	16	43.2
Mail/phone fraud	74	40	34	45.9
Test run 6 (improved representation – mutation enabled)				
Lost card fraud	17	10	7	41.1
Stolen card fraud	36	22	14	38.9
Skimmed fraud	37	8	29	78.4
Mail/phone fraud	74	29	45	60.8
Test run 7 (improved representation – vaccination enabled)				
Lost card fraud	17	8	9	52.9
Stolen card fraud	36	11	25	69.4
Skimmed fraud	37	12	25	67.6
Mail/phone fraud	74	23	51	68.9
Test run 8 (improved representation – mutation enabled – vaccination enabled)				
Lost card fraud	17	8	9	52.9
Stolen card fraud	36	9	27	75.0
Skimmed fraud	37	12	25	67.6
Mail/phone fraud	74	18	56	75.7

Table 12. Displays results from the experiment, (Wong et. al. 2012).

Test run	Detection rate (%)	False positive rate (%)	False positive ratio
1 IMV	71.3	7.1	2.21:1
2 IV	67.1	3.7	1.21:1
3 IM	57.9	6.0	2.28:1
4 I	43.3	2.7	1.37:1
5 BMV	42.7	6.0	3.09:1
6 BV	45.1	4.8	2.34:1
7 BM	14.6	0.7	1.08:1
8 B	12.8	0.6	1.05:1

I, Improved representation; B, basic representation; M, mutation enabled; V, vaccination enabled.

Table 13. Displays derived performance statistics, (Wong et. al. 2012)

The results show that a fully augmented Artificial Intelligence System achieves a detection rate of 71.3% using actual transaction data. This demonstrates that the method may be feasible, the researchers had however not predefined a success scenario. This makes it

tricky to therefore make a conclusion on the overall performance of the system with any authority. Use of actual data for the experiment provides some validity of the results obtained hence they can be deemed as significant to the real world. The research by (Wong et. al. 2012) states how the method is modelled after an immune system of biological organisms, however it omits to mention whether the inherent inefficiencies of that system may manifest themselves on the artificial immune system. A biological system does also have its own weaknesses.

2.3 Application of Genetic Algorithm and Scatter Search for Credit Card Fraud

Genetic algorithm and scatter search are concepts adopted from natural evolution, they use the principle of survival of the fittest. Individually the genetic algorithm works by creating an initial number of solutions, further solutions are then created using cross over and mutator operators, the new solutions created are more robust as the generations go on. The procedure iterate until a pre- determined number of generations have passed. The scatter search uses a reference list (set of solutions) and combines the solutions in order to formulate a new one. Duman and Ozcelik (2011) propose a detection method that combines the genetic algorithm and scatter search, referred to as the Genetic Algorithm Scatter Search (GASS). The algorithm largely follows steps of the genetic algorithm and incorporates key components of scatter search.

The researchers carried out a test in order to evaluate the method, the experiment was carried out in a Bank facility in Turkey. 1050 fraud transactions were used for the training dataset whilst a sample one hundredth of all the legitimate transactions were used. The number of legitimate transaction were later reduced in order to improve run time. Table below show data set that were used for the experiment

	# of frauds	Ratio of legitimate transaction taken	Imbalance
Set-1	1050	0.01	236.0
Set-2	1050	0.001	23.6
Set-3	1050	0.0002	4.7

Table 14. Displays composition of dataset used in experiment, (Duman and Ozcelik 2011)

Carrying out the experiment in an actual banking environment improves the credibility of the results hence it was a well thought out consideration by the researchers. Although the training dataset was reduced due to run time issues, by only eliminating legitimate transactions, the variation pattern of the fraudulent transactions was preserved, the number of legitimate transaction was also extremely large hence it reasonable to eliminate the data than eliminating fraudulent transactions which were extremely small in comparison.

2.4 Application of Hidden Markov Model for Credit Card Fraud

The Hidden Markov model works by identifying the spending profile of a credit card user. The profile includes quantity of the items being bought, shipping address and billing address. To address a potential privacy concern, the method does identify what types of items were purchased. Monetary amounts being spent are used to create clusters of training datasets, for example, there could be a cluster for small values, medium values and large values. Whenever a transaction occurs the Hidden Markov model processes whether there is a drastic variation with regards to the spending profile. If a transaction is deemed suspicious an alert is made to the credit card company and credit card holder.

The researchers carried out an experiment to test the effectiveness of the method. The Baum-Welch algorithm was used to estimate parameters of each credit cardholder. Training is performed offline whereas detection is wholly an online process. The table below compares the probability of a genuine transaction occurring versus a false transaction using mean distribution.

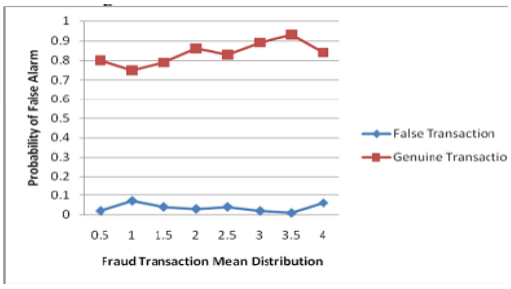


Figure 14 False alarm probability comparative to fraud transaction mean distribution

From the graph it is observed that there is a correlation between genuine and false transaction, whenever

probability of a legitimate transaction is declining, and the probability of false transaction increases. This helps identify any false detection of a transaction as fraudulent. Therefore when the probability of false detection is more than threshold probability, an alarm for fraudulent transaction is sent, the transaction is also declined.

Mishra et. al. (2013) does not explain the concept of spending profile adequately, which is a fundamental aspect of the method, spending patterns are influenced by a myriad of factors hence a credit card user may from time to time spend outside the scope of what is considered their spending profile, for example, people may spend large sums of money at certain periods of the year, and little sums at other periods of the year. It is not clear whether the Hidden Markov model makes considerations for such events. The researchers also posit that the 'hidden markov model is a perfect solution for addressing fraud transactions through the credit card', this is however not reflected by the results, the researchers have seemingly exaggerated the effectiveness of the method.

3 Conclusions

From a conceptual point of view, all the proposed methods seem to be well thought out. The idea of using spending profiles in order to detect transactions that may be fraudulent is informed by sound logic, both the personalised, aggregated and application of hidden markov model used this idea, although its success may rely heavily on whether the profile has been defined accurately. The experiments carried out demonstrated that the methods may be feasible however the way testing was conducted for some methods does not instill total confidence in the results, for instance, it would have been ideal for the other researchers to have adopted the approach by (Duman and Ozcelik 2011) by using an actual banking facility, this improves credibility a method as it provides a more realistic view of how a method would perform in the real world. In the future, it would be interesting to see researchers develop a method that does not use credit card history only to determine fraudulent data. For instance, before a customer is given a credit, they could be a facility where they are asked to define what should warrant an alarm. Methods that use historical data do not work effectively for a certain period of time when the customer has just been issued with the card.

References

Duman, E. & Ozcelik, M, J., 2011. 'Detecting Credit Card Fraud by Genetic Algorithm and Scatter Search.' *Expert Systems with Applications.* pp 13507-13063.

Ghazali, M, Z,Rahim, M, S,.Ali, A., Abidin, S., 2014.'A preliminary study on fraud prevention and detection at the state and local government entities in Malaysia.' *Procedia - Social and Behavioral Sciences*.,pp. 437 – 444.

Mishru, S. M., Panda, S., Mishra, A. S., 2013. 'A Novel Approach for Credit Card Fraud Targeting the Indian Market.' *International Journal of Computer Science Issues*, vol 10, no.3, pp 172-179.

Mohammed, I. A. & Soon, L.-K., 2012. 'Credit Card Fraud Detection: Personalised or Aggregated Models'. *FTRA International Conference on Mobile, Ubiquitous and Intelligent Computing*. FTRA., pp. 114-119.

Pavia, J. M., Ernesto, J. V.-F., Gabriel, F.-E., 2012. 'Credit Card Incidents and Control Systems'. *International Journal on Information Management*., pp.501-503.

Reardon, S., Nance, B., Mcombic, K.,. 2012. 'Visualisation of ATM Usage Patterns to Detect Counterfeit Card Usage'. *Hawaii International Conference on System Science*., pp. 3081-3088.

West, J. and Bhattacharya, M. 2016. 'Intelligent financial fraud detection: A comprehensive review.' *Computers & security*., pp. 47–66.

Wong, N., Ray, P., Stephens, G., Lewis, L., 2012. 'Artificial Immune Systems for the Detection of Credit Card fraud: An Architecture, Prototype and Preliminary Results.' *Info Systems*, pp53-76.

Yusuf, S., Serol, B.,Duman, E. 2013. 'A cost Sensitive Decision Tree Approach for Fraud Detection.' *Expert Systems With Application* .pp. 5916-5923.

Evaluation of Research in Automatic Detection of Emotion from Facial Expressions

Olorato D. Gaonewe

Abstract

Automatic detection of emotion has become a vital advancement in the world of technology as it has many applications. Several methods have been proposed to improve this technology, this paper analyses, evaluates and compares three of these methods. The methods evaluated are: Fully automatic recognition of the temporal phases of facial actions; a local approach for negative emotion detection, a novel automatic facial expression recognition method based on AAM and Feature Extraction and Facial Expression Recognition Based on Bezier Curve. The analysis of these methods proves that there are a lot of limitations in current methods to detect emotion from facial expressions even if some of the best methods can be combined. Further research and work is recommended in conclusion.

1 Introduction

Science has granted a wish (solution) to those who cannot read and comprehend facial emotions to effectively communicate and respond to social norms. This is made possible by research into automatic detection of emotion, using facial expressions. A breakthrough in this not only could generate financial income but also create immense benefits in the health section, in human computer interaction, education and others (Libralon et al 2014). Many glitches have been encountered in the pursuit of perfecting this science though.

Various researchers like Su et al (2014); Alonso-Martin et al. (2013) and Mavadati et al. (2014) allude that a great deal of detection and recognition of emotions methods cannot deal with variable unconstrained head poses, different illumination methods and natural emotions, making such systems less robust. Another problem is that the very creation of these systems is so expensive. Other problems highlighted by Alonso-Martin et al. (2013) include difference in emotion intensity and lack of spontaneous expressions use.

Robust detection and recognition in 2D cameras lacks in some conditions and thus Savran et al. (2013) says the use of 3D data expression helps alleviate problems of head orientation and illumination, but further research is being performed. Mavadati et al. (2014) goes on to add that further research into the use of spatial and temporal patterns to help in detecting and recognizing spontaneous expressions needs to be done as thorough analysis hasn't been performed yet. They add that using real-time and automatic learning systems can also help to that effect. Valstar et al. (2015) of FERA 2015 says that there is a need for further research in the automatic estimation of expression intensity.

Most methods are created on the foundation of Ekman, (1971)'s FACs system, his six emotions classification and lastly Anastasios et al. (2008)'s three phase technique, which divided the system in to: face appearance extraction, facial feature extraction to denote the facial expressions and analysis and classification of expression.

This survey paper will investigate into the current issues in automatic detection of emotion from facial expressions and research attempting to solve them, also critically evaluate four methods proposed by researchers to reach sound and solid scientific conclusions. The papers to be critically reviewed are: Fully automatic recognition of the temporal phases of facial actions; a local approach for negative emotion detection, a novel automatic facial expression recognition method based on AAM and Feature Extraction and Facial Expression Recognition Based on Bezier Curve.

2. Evaluation Facial Expression Research Papers

The following is a critical evaluation of four research papers. The scientific merit of the methods and experiments are evaluated then compared.

2.1 Fully Automatic Recognition of the Temporal Phases of Facial Actions

Valstar and Pantic (2012), this method only functions well if the first frame of an input video sequence shows a non-occluded neutral face in frontal view. In the first frame, a face region is detected using a facial point detector based on Gabor-feature and boosted classifiers which localizes about 20 facial fiducial points. These points are tracked throughout the sequence with the help of a tracking technique extracted from particle filtering with factorized likelihoods (PFFL).The tracking data from above is used to detect the presence of 22 action units. The tracking and identification of the action units is executed by using a combination of various methods being: Gentle Boost ensemble learning, support vector machines (SVMs) and HMMs. This also helps in determining the temporal activation model as a sequence of temporal segments for activated action units.

Five different sets of experiments were carried out to analyse and evaluate the performance of the scheme devised. The experiments evaluate different parts of the scheme: facial point detector, facial point tracker, AU (Action Unit) detector, AU temporal activation model detector and the six universal facial expressions.

To evaluate the performance of the facial point detector, two experiments were conducted: the first one used the first frames of 300 randomly selected image sequences from the CK-db (database) and the second, used first frames of 244 sequences from the MMI-db part 1 (database). The experiment of the CK-db images was evaluated using the threefold cross validation and the MMI-db, the facial point detector was trained using the CK-db images and then tested on the MM-db images. Automatically located facial points were compared with manually annotated points for evaluation. The researchers claim to have achieved an average of 93% recognition for the CK-db and 96% for the MMI-db and managed 20 facial feature points. Below are the results to this experiment to back up the claims made.

	MMI	CK		MMI	CK
A	0.784	0.920	G	0.982	0.950
A1	0.976	0.960	G1	0.982	0.990
B	0.976	0.960	H	0.976	0.980
B1	0.952	0.990	H1	0.976	0.970
D	0.569	0.960	I	0.904	0.970
D1	0.802	0.950	J	0.928	0.910
E	0.928	0.960	K	0.964	0.930
E1	0.958	0.900	L	0.952	0.800
F	0.982	0.910	M	0.904	0.900
F1	0.982	0.830	N	0.952	0.980
Average for all points:				**0.922**	**0.930**

Table 15 Classification rate of point detection on MMI-db facial expression CK-db (Valstar et al 2012).

The experiment for facial point tracker was conducted to test the preciseness of the PFFL point tracking algorithm used in the scheme. If there was an occlusion, its location point was calculated based on its location from the previous frame where the relevant point was still visible. In spite of this, there were occluded expression present which had a huge effect on the results.

To test the AU detector, a frame-based AU detection evaluation was conducted. The MMI-db was tested for 22 AUs and were detected using geometric feature based approach. Not all 22 AUs were present in required numbers in the CK-db. All investigations or tests were conducted by leave one subject out cross validation, this is said to make sure the system is trained as subject independent. Below are the results of the experiment.

AU	Videos	Frames	Cl. Rate	Recall	Precision	F1
1	22	1006	0.972	0.679	0.728	0.703
2	25	1092	0.961	0.628	0.629	0.628
4	38	1839	0.942	0.582	0.707	0.639
5	19	874	0.949	0.317	0.375	0.344
6	27	1241	0.952	0.695	0.583	0.634
7	15	772	0.963	0.319	0.510	0.392
9	15	636	0.968	0.503	0.477	0.490
10	17	719	0.955	0.266	0.321	0.291
12	17	1004	0.950	0.548	0.482	0.513
13	14	782	0.974	0.668	0.650	0.659
15	15	854	0.944	0.412	0.344	0.375
16	18	717	0.947	0.230	0.229	0.229
18	16	568	0.974	0.593	0.523	0.556
20	15	871	0.964	0.696	0.554	0.617
22	15	696	0.964	0.536	0.467	0.499
24	15	536	0.955	0.497	0.503	0.500
25	105	5401	0.909	0.810	0.831	0.821
26	32	1597	0.875	0.198	0.179	0.188
27	15	800	0.983	0.720	0.819	0.766
30	15	736	0.972	0.438	0.588	0.502
43	15	750	0.973	0.520	0.657	0.580
45	107	1243	0.956	0.668	0.625	0.645
46	6	130	0.913	0.723	0.667	0.694
Avg:			**0.953**	**0.532**	**0.541**	**0.533**

Table 16 Subject-independent cross-validation results for AU Activation detection (Valstar et al 2012).

To evaluate the six universal emotions, 171 videos taken from the CK-db were used for testing. The videos were picked with the criteria that two coders were

to achieve consensus on identifying the emotion shown in the video. The table below shows the results.

	ANGR	DISG	FEAR	HAPP	SADN	SURP
ANGR	2	3	2	0	9	1
DISG	1	19	1	1	4	1
FEAR	1	4	15	5	2	1
HAPP	1	0	3	33	0	1
SADN	4	2	1	0	16	1
SURP	0	1	1	1	0	34
Cl. rate	0.118	0.704	0.536	0.868	0.667	0.919

Table 17 Confusion matrix of emotion detection on the CK-db (Valstar et al 2012).

The researchers of this paper say that it is difficult to differentiate the emotions angry from sadness. They claim that the reason is that both movements use similar facial movements (brow movements). They go on to say that, fear is mostly confused with either disgust or happiness. They argue that, from a geometric point of view the reason is the downward motion of the lip corners. They conclude that, four points is not enough to capture the different shapes of the mouth and that increasing it to eight or more would improve it. That it would allow a geometric approach to better distinguish emotions without confusion.

The researchers of this paper claim that their scheme outperforms all other proposed methodS**,** which they compared their results with on the MMI facial expression database**.** They also claim that from the methods devised to detect temporal segments in the CK-db their method scores the highest. They go on to say that the geometric approach is very capable and suited for detecting four temporal phases of an AU with precise results. They say their system has been tested on several databases and it shows generalization. There is no evidence to back up their claims though, so they are inappropriate. The researchers claimed that their method scored highest against other existing methods, but these claims have no grounds as they were not tested in a controlled experiment.

This method has scientific merit as it is based on various techniques from previous works by other fellow researchers. A few of the methods from previous works are: temporal segments and the PFFL algorithm by Pantic and Patras; FACS and fiducial point detector by Vukadinovic and Pantic. This method was well researched and thought through it had contingency plans put in place to try avoid limitations. However, the method has limitations as it cannot handle heavy occlusions and fails if an occlusion is present in the first frame of the video sequence. Also images can only be recognized from a near front view and if an image in an angle is detected, the system fails.

The experiments and testing were of high standards, there were 5 experiments for every part or function of the system. The variables tested and tools used were appropriate. Four different data sets were used: MMI-db, CK-db, triad data and DS118. The sample size was appropriate for the experiments also. These data sets consisted of different data subjects of different gender, age and race; some were of drunken subjects and some of heart disease patients. They also consisted of different expressions ranging from the six universal ones by Ekman, to spontaneous facial expressions, spontaneous human behavior and volition facial displays. However, some of the videos were in grayscale which makes the method difficult to be used for real life applications. The facial expressions were made on command, which means some of the emotions were faked and of low intensity.

This method has potential to be great but needs more research. The researchers need to cater for different head orientations, real life images, videos and not posed or on command. They need more AUs in order to reduce the confusion in emotions. Also maybe increase the temporal segments from 4 as they clearly don't yield accurate results in a geometric based method. If all the limitations could be addressed, a great method that could be used in real applications could emerge. This method has contributed greatly to the subject area, as it has shown other researchers how to diligently conduct research.

2.2 A Local Approach for Negative Emotion Detection

Lablack et al. (2014), devised an emotion detection method that focuses on negative emotion as few method have done so. He says that detecting negative emotions can help applications with user experience as they would react to the user's emotions in a particular situation. He also says anger is often confused with disgust, creating confusion and leaving a gap in knowledge.

This system is based on the existing method, FACS by Ekman, it makes a local analysis to find common patterns, and a global analysis of every region area, to extract hints that show the activities of the muscles involved in a negative emotion. The method has two major phases being: image pre-processing and regions of processes.

Image pre-processing handles the extraction of a normalized face in a set size, shape and illumination and depicts only the necessary face region. It uses Boosted Haar and Viola Jones face detection algorithms. It also uses neural network approach to identify and locate

the exact position of the pupils and uses the vertical position of the eyes to determine the orientation of the face. It also uses histogram equalization to improve the image's contrast and reduce its intensity in order to normalize it. The picture below depicts this method.

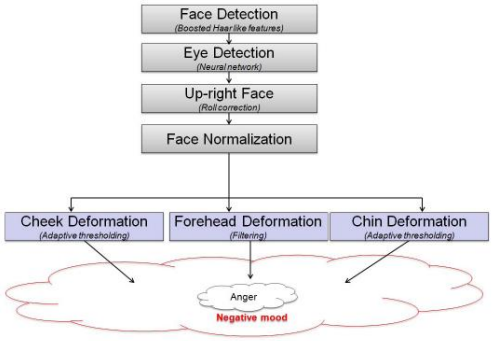


Figure 15 Approach/ method for negative emotion detection (Lablack et al 2014).

The region of processes performs texture analysis, so as to identify local patterns caused by negative emotions e.g. forehead area, vertical lines above the nasal root and lower face region. Gabor filters are used to analyze the texture and local binary patterns to further describe them and a threshold of the pixels used.

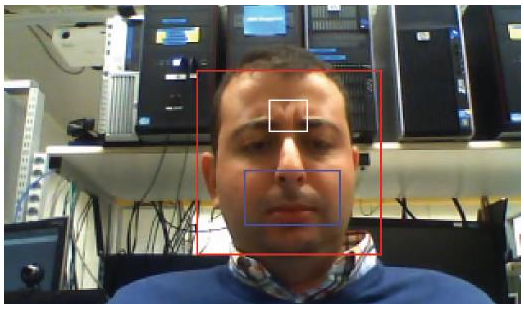


Figure 2 Region of processes method, (Lablack et al 2014).

Two datasets from the Karolinska Directed Emotional Faces (KDEF) database were used to test the scientific merit of the method in two experiments. The data sets consisted of static images and videos from a webcam acting different emotions (anger, sad, disgust, surprise, neutral, happy and afraid). The dataset was made up of 70 participants, who acted the 7 above emotions; all of them were 3 meters away from the camera. All of the images were 562 by 762 in resolutions.

The Researchers claim that they have correctly detected and recognised negative emotion, at a 95% rate. The results (pie chart) below acts as evidence

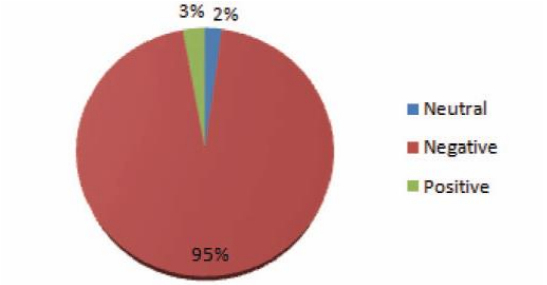


Figure 3 Results of the method performed on KDEF

The video experiment was designed to test the system for instances where the user is very close to the camera like in video conferencing or e-learning. The camera was embedded in a screen and 50 cm from the participant. The screen showed a video of the participants displaying various emotions. The participants were given guidelines to show or act all 7 emotions, especially negative emotions. Each participant was given 10 minutes in front of the camera.

Just like most proposed methods, this method only uses frontal view images and does not support full head orientation, aging, wrinkles to accommodate real life situations. However, in the illumination scenario, this method tries to solve the problem of lighting during the face normalization phase and also tries to address the head pose and orientation issues but can only go up to 30 degrees. The experiments used to test this method lack scientific merit as they have limited context on how they were performed and what was tested. The sample size used is also not enough to confidently make conclusions on the method. Therefore, there is a need for further work and research on it. However, I applaud the attempt in head orientation.

2.3 A Novel Automatic Facial Expression Recognition Method Based on AAM

To improve the accuracy and effectiveness of automatic recognition Wang et al (2013) introduced this method. It starts by detecting faces in images by a cascade classifier, then segmenting, extracting and normalizing them using Active Appearance Model (AAM). To recognize and characterize facial features Gabor filters are used and SVMs work with them to recognize facial expressions. The image below depicts the method.

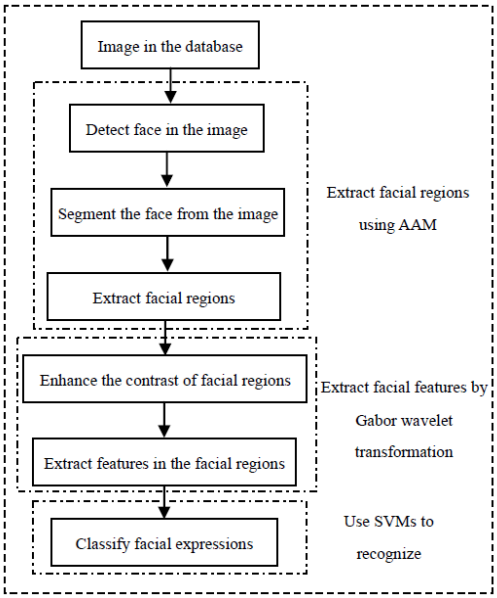


Figure 4 overall method proposed by Wang et al (2013).

Two databases are used to run experiments to validate the method, JAFFE and YALE. The JAFFE db has 213 images of 10 Japanese woman posing to a certain criteria of the common 7 emotions. The Yale db consists of 15 people posed in 60 expressions creating 165 images. Three experiments are conducted to validate the merit of the method.

Using the JAFFE data set a tenfold cross validation is performed. Some of the dataset is used to train SVM classifiers. The results of the experiment are below, average rate and confusion matrices of classification are represented.

Method	Ten-fold cross validation
LBP[31]	80.7±0.5
LD_iP[32]	85.9±1.8
Gabor[33]	79.7±4.2
LDN^G[34]	88.7±0.2
R=1	87.3±1.2
R=3	89.7±0.5

Table 18 recognition rate using ten-fold (Wang et al 2013).

R=1	AN	DI	FE	HA	NE	SA	SU
AN	87.5	0.0	0.0	0.0	2.78	3.23	0.0
DI	6.25	96.15	3.03	0.0	0.0	3.23	0.0
FE	0.0	3.85	87.88	0.0	0.0	3.23	3.57
HA	0.0	0.0	0.0	92.59	11.1	6.45	0.0
NE	0.0	0.0	0.0	0.0	77.78	6.45	0.0
SA	6.25	0.0	6.06	3.7	5.56	77.42	0.0
SU	0.0	0.0	6.06	3.7	2.78	0.0	96.43

Table 19 Confusion classification matrix for the whole face (Wang et al., (2013).

For a group test experiment, the JAFFE database is used. There is a huge margin of error where happiness is recognized as neutral and most of the expressions are recognized as sadness except for surprise. The authors say this is because it is difficult for humans, let alone machines to recognize small or subtle facial expressions.

The third experiment is the independent subject test, where only one subject is tested and the others are used for training. Gabor bank and texture are used for testing. The Yale database is used to provide the dataset. The recognition rate for this experiment is very robust but confusion of emotions still takes place.

The authors claim that this method is more accurate than others and that its computation amount and time is low which makes it preferential to others.

Research does show that it is faster but it confuses a lot of emotions with others, it is not entirely accurate. Although the authors made a claim of their method being accurate and preferential, they did not present an experiment to back their claim. Gabor filters cause noise and hinder the proper extraction of expressions and recognition. The dataset used from the JAFFE database is not adequate as it consists of only Japanese females; rather use a dataset consisting of individuals of different genders, race and age. It could help in identifying if the system handles occlusions well. With that said, this method needs further research in identifying ways of accurately extracting facials expressions and reducing confusion in recognition.

2.4 Feature Extraction and Facial Expression Recognition Based on Bezier Curve

Bao and Ma (2014) proposed this method based on previous ones, to extract features of facial expression and recognize them based on the Bezier curve.

This method aligns the feature points and then extracts the features of the facial expression and recognizes it. This is so, to get the displacement characteristics and shape features making neutral expression. After feature point calibration, ASM is used to model and locate necessary feature points like coordinates of facial corner feature points.

The JAFFE database was used, it contained 213 emoticons (expression) expressed by 10 Japanese woman. The women expressed six different emotions including neutral. A leave one method is used to train data and test the proposed technique. The Bezier curve works by marking points on the eyes, mouth and eye

brows to try measure the corresponding changes of expressions. The control point reversing algorithm is used to get control of points. The Bezier shape feature of expression is then extracted is get the feature vector.

This experiment used a neutral face as a template, using position and direction variables of other expressions corresponding to neutral expressions as feature vectors. Support vector machines and neural networks algorithm were used to recognize and classify expressions/emotions. The table below shows results of the experiments.

TABLE III. BEZIER+SVM

EXP	HA	SA	SU	DI	AN	FE
HA	100%	0	0	0	0	0
SA	0	80%	0	0	0	0
SU	0	0	100%	0	0	0
DI	0	20%	0	100%	0	0
AN	0	0	0	0	100%	0
FE	0	0	0	0	0	100%
rate: 95.65%						

Note: A total of 23 test set expression image, 22correctly identified.

Table 20 Results for Bezier with SVM (Bao and Ma 2014).

TABLE IV. BEZIER+NN

EXP	HA	SA	SU	DI	AN	FE
HA	100%	0	0	0	0	0
SA	0	100%	0	0	0	0
SU	0	0	100%	0	0	0
DI	0	0	0	100%	20%	0
AN	0	0	0	0	60%	0
FE	0	0	0	0	20%	100%
rate: 91.3%						

Note: A total of 23 test set expression image, 21 correctly identified.

Table 21 Bezier with NN (Bao and Ma 2014).

TABLE V. COMPARISON OF THREE METHODS

Method	Gabor+SVM	Bezier+SVM	Psoica+AMM
rate	83.8%	95.65%	95.8%

Note: Gabor + SVM data cited from the literature [12], Psoica + AMM data cited in the literature [13].

Table 22 Comparison of methods (Bao and Ma 2014).

This method lacks scientific merit and its results are not valid as the experiments lack preciseness, validity and repeatability. Only one dataset was used for experiments, which consisted of only Japanese women. This limited the results of the system from people of different genders and race. There is no information about occlusions, illumination, poses or picture resolutions in the experiments. This paper lacks structure and more clear information of how the method works and its experiments need to be added. The paper has made no compelling arguments and has not contributed to the subject area.

3. Comparisons

All the methods have confusion with identifying or mistaking some emotions as others. This is because of subtle expressions or facial movements which look the same. Wang (2013), did an experiment on classification confusion which accurately recognizes some emotions, making it the better method in this case.

In terms of computation time, Wang (2013), is the only author who stated the performance of their method in that regard. The other methods don't consider performance and efficiency.

Valstar et al., (2012), this method is limited in terms of occlusions and have not indicated a way of overcoming them like the two other systems. Valstar et al., (2012), this method fails if there is an occlusion in its first frame of the video stream. Most of the methods use frontal view images, gray scale pictures and images and don't cater for head orientation.

4. Real World Implications

The success of automatic detection of emotions from facial expressions could benefit a lot of industries and have various applications. Many authors come to a consensus that it can be used in human computer interaction to help give users a satisfactory user experience and to help collect data on user difficulties or dislikes during their interactions. It could also be applied in computer gaming by adapting the game to the user, says Alonso-Martin et al. (2013). He goes on to say Microsoft has already applied it in their game Kinect. Still in the computing world, it can be applied in computer-machine communication between a human and a robot. This could mean a robot being able to understand humans and working effectively with them.

Another helpful application is in the education system, like in e-learning for long distance students and in classrooms to gauze the concentration and understanding of students. A vast use of these methods can be applied in the health section, to automatically detect a patient in pain, monitor the state of patients with dementia to detect depression incidents and for people with autism and schizophrenia.

5 Conclusions

To create an efficient and effective automatic detection of emotion from facial expression, a lot of methods can be combined. Rather some features from different methods can be extracted to achieve accurate emotions recognition.

The two most promising methods that can be combined to come up with an ultimate method is the Valstar (2012) and Wang (2013) but a few features like region of processes from Lablack et al., (2014)'s

method could be extracted. The region of processes could help with the confusion of emotions.

Even with the two methods combined, the system wouldn't be entirely accurate, further research needs to be conducted on handling subtle facial muscle movements to reduce confusion of emotions. Other areas which need further work and research are handling of occlusions, variable head poses, lighting, and use of spontaneous expressions, also naturally colored images. And most importantly, the systems should be created for everyday environment applications. Research says the use of 3D technology could really improve automatic detection of emotions, therefor it is recommended that this technology be considered.

References

Alonso-Martın, F., Malfaz, M., Sequeira, J., Gorostiza, F., J., 2013, 'A Multimodal Emotion Detection System during Human–Robot Interaction' *Journal on Sensors* Vol 13 pp 15549-15581

Bao, H., Ma, T., 2014, 'Feature Extraction and Facial Expression Recognition Based on Bezier Curve' *2014 IEEE International Conference on Computer and Information Technology* pp 884- 887, Xi'an, 11--13 September.

Lablack, A., Danisman, T., Bilasco, M. I., Djeraba, C., 2014 'A local approach for negative emotion detection' *2014, International Conference on Pattern Recognition,* pp 417-420, Stockholm ,24--28 August.

Libralon, L. G., Romero, F. A., 2014, 'Mapping of facial elements for emotion analysis' 2014, *Brazilian Conference on Intelligent Systems,* pp 222-227, Sao Paulo ,18--22 October.

Lin, L. K., Huang, T., Hung, C. J., Yen, Y. N., Chen, J. S., 2013, 'Facial emotion recognition towards affective computing-based learning' *Library Hi Tech* 31 (2): 294- 304

Mavadati, M. S., Mahoor, H. M., 2014, 'Temporal facial expression modelling for automatic action unit intensity measurement' 2014 *22nd International Conference on Pattern Recognition*, pp 4648- 4653, Stockholm, 24—28 August.

Savran, A., Gur, R., 2013, 'Automatic detection of emotion valence on faces using consumer depth cameras' *Computer Vision Workshops, 2013 IEEE Conference (ICCVW)* pp 75-82, Sidney, NSW, 28 December.

Su, L., Levine, D. M., 2014, 'High stakes deception detection based on facial expressions' *Pattern Recognition (ICPR) 22nd International Conference,* pp 2519-2524, Stockholm, 24-28 August.

Valstar, F. M, Pantic, M., 2012, 'Fully Automatic Recognition of the Temporal Phases of Facial Actions*' IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 42(1):28-43, February.

Valstar, F. M., Almaev, T., Girard, M. J., McKeown, G., Mehu, M., Yin, L., Pantic, M.,
Cohn, F. J., 2015, 'FERA 2015 Second Facial Expression Recognition and Analysis Challenge' *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops,* Vol 06, Ljubljana, 4—8 May

Wang, L., Li, R., Wang, K., 2014, 'A Novel Automatic Facial Expression Recognition Method Based on AAM*' Journal of Computers,* 9(3): 608-617, March.

A Critical Evaluation on Methods of Increasing the Detection Rate of Anti-Malware Software

Thomas Gordon

Abstract

Malware has been a threat to computer systems for a very long time, and this threat has caused a range of anti-malware methods to be researched and developed. This paper describes a critical evaluation of proposed anti-malware methods, the experiments during development and results of these experiments, and claims made by the authors, as well as comparing these methods and concluding which method is most suitable for real world use.

1 Introduction

Malware has been a problem for a very long time, in fact, "malicious activities on the Internet are one of the most dangerous threats to Internet users and organizations" (Lee & Lee 2014). Vida et al. (2015) reported that global damages due to malicious software reached $370 million, and this problem is only getting worse as time passes. Alam et al. (2015) showed that there were 403 million new malware created in 2011 alone, with this number increasing drastically each year.

The massive amount of new malware that is created each year makes it impossible for current anti-malware software to detect every piece of malware. Shahzad et al. (2013) described that anti-malware software cannot be efficient and effective unless it achieves a high detection rate, a low false positives rate and detects malware quickly while being resilient to any evasion attempts. These characteristics can be hard to achieve however, as "the simplest obfuscation technology can fool current commercial malware scanners" (Ding et al. 2014).

This research paper will discuss different methods of malware detection. The experiments, claims, results and conclusions will be critically evaluated based on their detection rates, and whether they are viable methods for use within the real world.

2 Malware Detection using Application Programming Interface (API) Calls

API calls consist of a set of tools, routines, and protocols which are used when building a software application. These calls determine how an application behaves, and are therefore commonly used when creating anti-malware software.

2.1 API Call Graphs

Elhadi et al. (2014) proposed a method using API calls to detect malware, which they believed would offer enhanced detection over other methods.

The proposed system uses operating system resources and API calls in order to compare programs to existing malware samples by transforming them into an API call graph. A graph matching algorithm was then used to compare similarities with the input call graph and malware call graph samples.

A dataset of 416 malware and 98 benign programs were collected, which was used to test the proposed system. Results were collected for the proposed system, and also for two other malware detection systems (see figure 1). The amount of malware detected was recorded, as well as the similarity rate of this malware. Elhadi et al. (2014) deemed that in order for a program to be seen as a virus, it must have a similarity of 50% or higher when compared to a virus sample.

	Max similarity	Min similarity	Success	Failure
(a) Statistics for virus group				
Method (Virus vs. Virus)				
N-gram	98.97	28.96	156	9
(Lee et al., 2010)	97.47	57.44	165	0
Proposed system	82.57	53.84	165	0
(b) Statistics for Trojan group				
Method (Trojan vs. Trojan)				
N-gram	87.71	18.46	94	23
(Lee et al., 2010)	95.88	34.72	109	8
Proposed system	76.74	38.69	111	6
(c) Statistics for Worm group				
Method (Worm vs. Worm)				
N-gram	93.97	16.46	119	15
(Lee et al., 2010)	97.81	41.36	130	4
Proposed system	82.25	45.65	130	4

Figure 1 – Table of Results (Elhadi et al. 2014)

The results showed that the proposed system successfully detected more malware compared to other methods tested, having only failed to detect 10, however it is also noted that although the proposed system has a higher detection rate it can take up to 16 seconds to run a query on each file.

Elhadi et al. (2014) claims that the proposed solution shows significant improvement over previous attempts at malware detection using call graphs, with a 98% detection rate and 0 false positives rate.

The test strategies successfully show that the proposed system shows improvement over previous systems. However, the dataset used to test the proposed system only contains 3 categories of malware, which doesn't take into account all types of malware available. This lowers the validity of the results, as a detection rate of 98% may not be achievable with a wider dataset. The benign programs in the dataset were obtained from Windows XP, which is now outdated and rarely used. As a result more tests may be needed with more up to date programs.

Although the method shows promise, it also has its limitations and further research is needed before the method will be viable for real world use. Currently the only tests that have been conducted have been done with a very small dataset with outdated files, so tests with a larger and more up to date dataset may be required. In addition to the small dataset, results showed that this method is currently significantly slower than other methods and can take up to 16 seconds to complete one query, which is much too long a time for a real world scenario.

2.2 Feature Generation from API Calls

Salehi et al. (2014) proposed a different method of malware detection using API calls, instead of using the API calls to create a graph like the method proposed by Elhadi et al. (2015), they are instead used in feature generation.

When the API calls are extracted from a program, they are used to create sets of features. These feature sets are looked at by the system and any important features in the set are removed. The feature set is then used to create vectors of the files in the dataset. These vectors are used to reduce the number of features in the feature set so that it can be classified. 10-fold cross-validation was used in the classification stage, and it is accepted as being a suitable method of classification for malware detection.

The method was tested using 385 benign programs, and 826 samples of malware. The malware samples consisted of 7 different types of malware (figure 2)

Type	File Count	Sample Percent
Constructor	190	15.6
Trojan	91	7.5
Backdoor	173	14.3
Virus	42	3.5
HackTool	146	12.1
P2P-Worm	108	8.9
Exploit	76	6.3
Malware	**826**	**68.2**
Benign	**385**	**31.8**

Figure 2 – Dataset Files (Salehi et al. 2014)

Three sets of tests were done in order to test the system; each test used a different set of data. The first test (API-List) utilized the name of API's, the second test (ARG-List) used arguments, and the third test (API-ARG List) used both the name of the API's and arguments.
In all three instances the proposed method was tested against the 1184 programs mentioned earlier, McAfee antivirus was also tested as a comparison for the results.

Accuracy of all feature sets in comparison to McAfee anti-virus.

McAfee	API-List	ARG-List	API-ARG-List
94.4	93.6	98.4	97.6

Figure 3 – Results of experiments (Salehi et al. 2014)

The results of the experiments (figure 3) showed that of the three sets of data used, ARG-List was the most successful, achieving a 98.4% detection rate (with around a 3% false positives rate). This backs up Salehi et al. (2014) claims that the system outperforms McAfee's anti malware software.

The approach used to test the software was very detailed, and looked over many variables. The dataset that was used contained a wide range of malware types, and results shown in all three tests not only show the overall detection rate, but also show the detection and false positive rates of various malware classifiers. This increases the validity of the results shown.

Despite the high detection rate of the proposed method, Salehi et al. (2014) acknowledge that the method still needs more work. Although the proposed method has been thoroughly tested, further tests with larger datasets and unknown malware are still needed.

As a result, this method is not yet viable for use in the real world.

3 Malware Detection using Register Contents

Ghiasi et al. (2015) proposed a framework for malware detection which compares register contents to detect malware, named Dynamic VSA.

The framework worked by retrieving the registers generated before and after an API call is made. Since the number of registers are far too many to process efficiently, a small set of dataset samples (prototypes) are extracted from the registers, which represented the behavior of the whole dataset file. These prototypes where then put through a matching phase, which compares similarities between the prototype and malware samples to determine whether the prototype is a benign program or malware.
In order to test the proposed framework, a dataset consisting of 850 malware (7 different types) and 390 benign programs was used. The register values selected for the matching phase were those that appeared in 98% of dataset files in order to give an acceptable amount of features while only slightly decreasing any accuracy. Testing was carried out several times with many different thresholds in order to get accurate results, and each threshold was measured for its detection, false positives and f-measure rates.

The results (figure 4) showed that on average the proposed framework achieved a similar detection rate to current anti-malware software, and at its best outperformed this software.

	ESET NOD32	Kaspersky	Avira	McAfee	Our detection	
					Average	Best
Detection rate	0.878	0.976	0.920	0.964	**0.959**	**0.978**

Figure 4 – Detection rate results (Ghiasi et al. 2015)

Ghiasi et al. (2015) claims that the proposed system shows significant improvement over their previous works, with a 96% detection rate and 4.5% false positives rate. They also claim that it outperforms the updated versions of commonly used anti-malware software.

The test results show that at its best, the proposed framework achieved a detection rate of 97.8%, however this is not an accurate representation of the systems detection rate, as on average the system only achieved a detection rate of 95.9%. It is shown in the test results that anti-malware software by McAfee and Kaspersky achieved a higher detection rate than this, and so the claim that the proposed framework outperforms commonly used anti-malware software cannot be justified.

Although the proposed framework shows promise when dealing with known types of malware, it is uncertain whether it would be able to detect malware that is unknown. In order for this method to be viable for real world use, further tests would need to be conducted with unknown malware, and a larger set of data would be needed.

4 Malware Detection using Malware Analysis Intermediate Language (MAIL)

Alam et al. (2015) proposed a framework for detecting malware in real-time, named the MARD framework. This framework works by turning program samples into MAIL and translates this into a behavioural signature. This signature is checked for its similarity to malware and the system moves on to the next program sample. There are two methods the framework can use to turn a sample into a behavioural signature; this section will only focus on one of these methods (the ACFG detection technique).

In order to test their framework two different datasets were used. These datasets consisted of a smaller dataset of 250 malware and 1101 benign programs and a larger dataset of 1020 malware and 2330 benign programs. Test results showed that detection rates between 94% and 99.6% was achieved, with false positive rates between 3.1% and 4.5% depending on the dataset and training set sizes (figure 5). When compared with other malware detection methods, the proposed system achieved similar detection rates (figure 6)

Malware detection results for smaller dataset.

Training set size	DR	FPR	MMP	MMA	Real-time
25	94%	3.1%	0.86	0.96	✓
125	99.6%	4%	0.85	0.97	✓

Malware detection results for larger dataset.

Training set size	DR	FPR	MMP	MMA	Real-time
204	97%	4.3%	0.91	0.96	✓
510	98.9%	4.5%	0.91	0.97	✓

Figure 5 – Test Results (Alam et al. 2015)

System	Analysis type	DR	FPR	Data set size Benign/Malware	Real time	Platform
MARD-ACFG	Static	98.9%	4.5%	2330/1020	✓	Win & Linux 64
API-CFG (Eskandari and Hashemi, 2012a and Eskandari and Hashemi, 2012b)	Static	97.53%	1.97%	2140/2305	X	Win 32
Call-Gram (Faruki et al., 2012)	Static	98.4%	2.7%	3234/3256	X	Win 32
Code-Graph (Lee et al., 2010)	Static	91%	0%	300/100	X	Win 32
DTA (Yin and Song, 2013)	Dynamic	100%	3%	56/42	X	Win XP 64
Model-Checking (Song and Touili, 2012a)	Static	100%	1%	8/200	X	Win 32
MSA (Vinod et al., 2012)	Static	91%	52%	150/1209	X	Win 32
VSA-1 (Leder et al., 2009)	Dynamic	100%	0%	25/30	X	Win 32
VSA-2 (Ghiasi et al., 2012)	Dynamic	98%	2.9%	385/826	X	Win XP 64

Figure 6 – Comparison to other methods (Alam et al. 2015)

Alam et al. (2015) claim that this method shows superior results, and also supports automatic detection on Linux and 64 bit windows unlike other methods tested.

The test strategies used have been very thorough, having multiple test cases with different thresholds and dataset sizes, however when compared to other methods the results show potential bias, as many of the results either have a much smaller dataset and therefore unrealistic detection rates of 100%, or have a significantly higher false positives rate, which would deem the method much less effective in the real world. In addition to this, none of the 9 methods tested are capable of running on the same platform as the proposed system, and were tested on either 32 bit systems or Windows XP whereas the proposed system was tested on a 64 bit machine running windows 8.

5 Conclusions

Currently there are no anti-malware methods which can offer complete protection to a system. This means that every system is always in danger of malware attacks, and all of the methods mentioned above have their limitations. As a result, it is up to whoever is in charge of a network to ensure that the correct measures are put in place to lessen the damage caused by any malware attacks.

The methods discussed show that there are many different approaches to detecting malware. Even though none of these methods are perfect, all of them show promise, and with more research and testing could become much more viable for real world use.

The research conducted by Elhadi et al. (2014) and Salehi et al. (2014) shows that there is a lot of promise in detecting malware by using API calls, with both methods achieving a high detection rate of 98%.

The method presented by Elhadi et al. (2014) compares call graphs constructed from malware samples and API calls. Their research involved a lot of complicated theory and extensive research into the patterns of graphs, and results from their experiments were very promising. Despite this, there is still a lot of work that must be done before the method is ready for use in the real world. During testing the dataset used was of a small size, and was outdated. In addition to this, the method was significantly slower than other methods tested, and the dataset only contained three types of malware. Before use in the real world the method would need further research to increase efficiency and speed up detection time, and further tests with larger datasets with unknown malware would be needed.

Research conducted by Salehi et al. (2014) uses API calls differently, and extracts features from them in order to identify if a program is malicious or not. This differs from the approach of Elhadi et al. (2014), and as a result most of the research was based around API functions and behaviours. Results of their experiments were very detailed and outlined the many different thresholds that could be used, however Salehi et al. (2014) acknowledge that the method still needs more testing with larger datasets and unknown malware samples before it can be tested in a real world scenario.

The research conducted on the Dynamic VSA system by Ghiasi et al. (2015) shows extensive research into the behavior monitoring of register contents. Results of their experiments showed that the method has a lot of promise and these results are backed up by graphs and tables. Although this method has been tested against known types of malware, it has not been tested against unknown malware, and as a result further tests with unknown malware, and larger datasets would be required before this method can be used in the real world.

The research conducted on developing the MARD framework by Alam et al. (2015) shows a lot of real world potential, as the method is designed to detect malware in real time. The test strategies implemented are very thorough, using several datasets and parameters when testing to get accurate results. However, results shown from comparisons with other methods show a potential bias, which damages the credibility of the rest of the results. Before real world testing can be accomplished, the framework must be tested against unknown malware samples.

Though all of the methods mentioned need more research and testing done before they can be used in the real world, as none of them have been tested

against unknown software, however, of all the methods that were looked into, the MARD framework by Alam et al. (2015) has the most potential for real world use, as it detects malware attacks in real time. Having said this, the MARD framework is not perfect, and therefore system administrators would have to consider other anti-malware methods to be used alongside it.

References

Alam, S, Horspool, R, Traore, I, & Sogukpinar, I 2015, 'A framework for metamorphic malware analysis and real-time detection', *Computers & Security*, 48, pp. 212-233

Ding, Y, Yuan, X, Tang, K, Xiao, X, & Zhang, Y 2013, 'A fast malware detection algorithm based on objective-oriented association mining', *Computers & Security*, 39, Part B, pp. 315-324

Elhadi, A, Maarof, M, Barry, B, & Hamza, H 2014, 'Enhancing the detection of metamorphic malware using call graphs', *Computers & Security*, 46, pp. 62-78

Ghiasi, M, Sami, A, & Salehi, Z 2015, 'Dynamic VSA: a framework for malware detection based on register contents', *Engineering Applications Of Artificial Intelligence*, 44, pp. 111-122

Lee, J, & Lee, H 2014, 'GMAD: Graph-based Malware Activity Detection by DNS traffic analysis', *Computer Communications*, 49, pp. 33-47

Salehi, Z, Sami, A, & Ghiasi, M 2014, 'Feature: Using feature generation from API calls for malware detection', *Computer Fraud & Security*, 2014, pp. 9-18

Shahzad, F, Shahzad, M, & Farooq, M 2013, 'In-execution dynamic malware analysis and detection by mining information in process control blocks of Linux OS', *Information Sciences*, 231, Data Mining for Information Security, pp. 45-63

Vida, R, Galeano, J, & Cuenda, S 2015, 'Vulnerability of state-interdependent networks under malware spreading', *Physica A: Statistical Mechanics And Its Applications*, 421, pp. 134-140

An Evaluation of the Effectiveness of the Advanced Intrusion Detection Systems Utilizing Optimization on System Security Technologies

Carlos Lee

Abstract

Intrusion Detection Systems (IDS) are one of the first line of defense against attackers. Numerous studies have been developed to fight against this ever growing threat. This paper critically evaluates three key areas that directly affect IDS: Data Mining, Quality of Service and Data Filtering. This paper will compare and discuss the work that has been developed and further work that could be carried out to enhance the performance of IDS such as Multicore Layer Processes with the use of Fuzzy filtering systems to focus and enhance protection to networks and users.

1 Introduction

Intrusion Detection Systems (IDS) are one of three main lines of defense against attackers in system security; the other two being prevention and correction. It is imperative that IDS is at the highest level of technical advancement. Waleed Bul'ajoul (2015), clarifies that new methods and equipment are being developed in the fight against the ever growing threats to daily lives in security.

Even though new methods are continuously being developed against these attacks, hackers and intruders are also raising their abilities in gaining access into these security technologies (Kelton, et al., 2015)**.** This brings the ever growing challenge on system administrators to prevent intruders from gaining unauthorized access.

Wei-chao Lin (2015) presents a detection system based on nearest neighbors' path. Although the theories represent intrusion detection systems, data mining and machine learning techniques, the research focused on one link-to-link peer, Niva Das & Tanmoy Sarkar (2014) also takes a comparable view to the one-to-one link looking at host and system based intrusion detection systems. This results in inadequate justification towards this overall research paper.

Amal Saha (2014) takes a similar approach to Wei-chao Lin (2015) based on the Machine Learning technique however, expanding on the concept of making the systems scalable rather than link-to-link. Although this shows good understanding in the missing topology, Amal Saha (2014) has only based this on a concept for further development.

This survey paper will take a critical approach on intrusion detection systems, how they operate when optimizing key areas in preventing attacks, specifically focusing on the areas of Parallel Technologies, Quality of Service (QoS) and Data Filtering. With the enhancement to these key areas, they will be able to aid with the protection to networks and users, by developing new methods and theories in these specialized sectors.

2 Parallel Technologies

One method of parallel technologies that has been proposed to improve IDS security, is Fuzzy systems and pairwise learning method by Salma Elhag et al (2015) is a technique that overlays different technologies such as QoS and data mining to improve misuse of detection towards IDS.

Salma Elhag et al (2015) believes that this research will be able to bridge the gap between computing languages making them able to communicate smoother across border protocols enabling higher interpretability. They also believe that with research into the learning scheme they will be able to detect abnormal attacks, thus enhancing overall security.

In order to test their method, Salma Elhag et al (2015) used a data set from Lincoln Labs LAN network acquiring approximately 5 million TCP dump files. They only used 10% of these due to the sheer volume. They clustered the information found into categories Figure 1.

Class	SubClasses	Size (distribution %)
Normal	Normal	97,278 (19.6911)
DOS	back, land, neptune, pod, smurf, teardrop	391,458 (79.2391)
PRB	ipsweep, nmap, portsweep, satan	4,107 (0.8313)
R2L	ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster	1,126 (0.2279)
U2R	buffer_overflow, loadmodule, perl, rootkit	52 (0.0105)

Figure 16 TCP Dump files (Salma, et al., 2015)

The learning stages for FARC-HD that are used to be able to achieve this higher level of accuracy are displayed in Figure 3

Salma Elhag et al (2015) claims that with a multi-objective algorithm they would be able to maximize their metrics on the performance of Genetic Fuzzy Systems

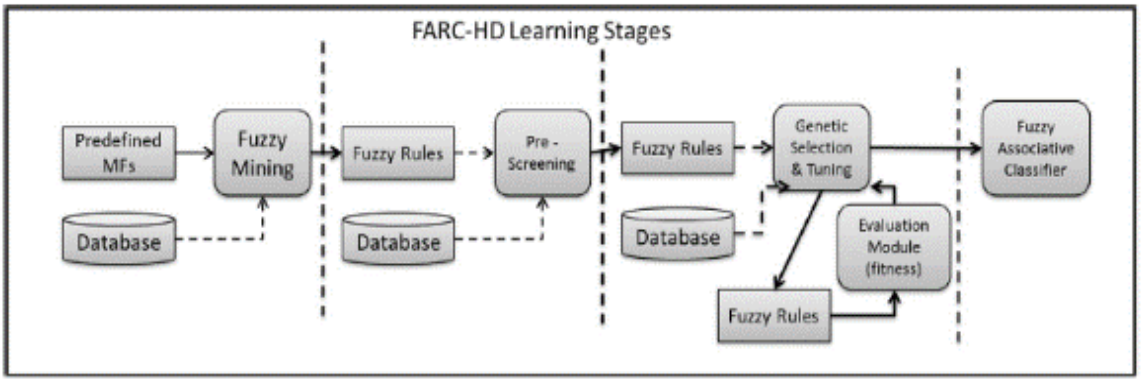


Figure 3 FARC-HD Learning Stages (Salma, et al., 2015)

By using the information that have been gathered from Figure 1. They developed six algorithms on different characteristics, based on misuse detection; these being:

- Accuracy
- Mean F-Measure
- Average Accuracy
- Attack Accuracy
- Attack detection Rate
- False Alarm Rate

The data output from data gathered gave Salma Elhag et al (2015) enough information to be able to conduct further experiments on Fuzzy Associated Rule-Based Classification for High-Dimensional One-vs-One (FARCHD-OVO) and Fuzzy Associated Rule-Based Classification for High-Dimensional (FARC-HD). This shows that there are clear improvements to FARCHD-OVO over FARC-HD Figure 2.

Classifier	Acc	MFM	AvgAcc	AttAcc	ADR	FAR
FARCHD-OVO	99.00	84.12	**89.32**	**86.70**	97.77	.1910
FARCHD	98.30	84.26	87.76	84.77	96.17	.2947
MOGFIDS	92.77	61.68	62.19	53.15	91.41	1.6599
GFS-GCCL	98.68	77.87	85.59	82.12	97.49	.5214
GFS-Pitts	98.64	75.55	86.07	82.69	97.26	.4016
GFS-IRL	98.64	**85.42**	85.18	81.57	97.16	.3777
Boost-FAR	97.61	74.26	67.47	59.36	94.13	**.0845**
C4.5	**99.59**	81.81	87.79	84.79	**99.29**	.2062

Figure 2 FARCHD-OVO over FARC-HD (Salma, et al., 2015)

(GFS) depending on the context that is given. They claim that this can prove, with the use of FARCHD-OVO their results are considerably more accurate especially in F-measure, the average accuracy and the false alarm.

The research produced by Salma Elhag et al (2015) argues that, the enhancement of Fuzzy systems is well justified, as seen by progression of new technologies. Their experiments are well demonstrated and produce clear improvements towards IDS. They carried out many experiments with a large data set in order to get as much accurate information as possible.

Although the experiments they conducted were well presented. They could have gotten much more accurate information by using more data sets as the one they used, they only used a small percentage of them, and this data set was also from 1998 meaning that the data is far outdated. This is bad practice due to the fact that there is other data sets that can be used. Salma Elhag et al (2015) could have developed more recent data sets in the LAB environment to give his research more accuracy. Tieming et al (2014) took a similar approach by using an old data set such as KDD99 for evaluation. They acknowledged this was an old data set and would not have current IDS network information, so they also used a data set called CDMC2012 to make sure they were able to get the most accurate data possible. Salma Elhag et al (2015) could take this approach to be able to further strengthen their method.

3 Quality of Service

When investigating research based on Quality of Service (QoS) by optimizing a system to be able to make it function at a higher capability. Waleed Bul'ajoul (2015) has taken this approach by utilizing QoS in Cisco Switches to increase performance in Network Intrusion Detection and Protection Systems (NIDPS)

Waleed Bul'ajoul (2015) understands that the optimization of NIDPS will be able to decrease the amount of packets that are lost through the result of heavy traffic or high speed network attacks based on DoS and DDoS attacks.

In order to test their theory on NIDPS they use Snort IDS being the most popular NIDPS at the time of testing. They prepared their testing area in a laboratory environment, Figure 4 displays the network structure that was used.

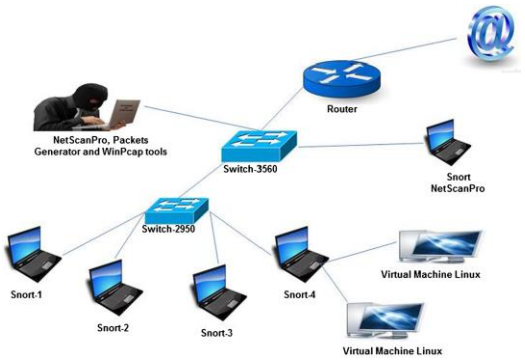


Figure 4 Network structure for testing (Waleed, et al., 2015)

They conducted several tests on the network using Snort to show performance with and without QoS configured on their network. They also conducted a test on QoS with parallel technologies.

The tests without QoS (Figure 5), shows the performance is affected by high-speed and heavy traffic, whereas Figure 6 demonstrates that is it able to cope with the volume of packets sent more efficiently. They claimed however it is not perfect due to the capability issues of Snort not being able to handle large amounts of data at one time.

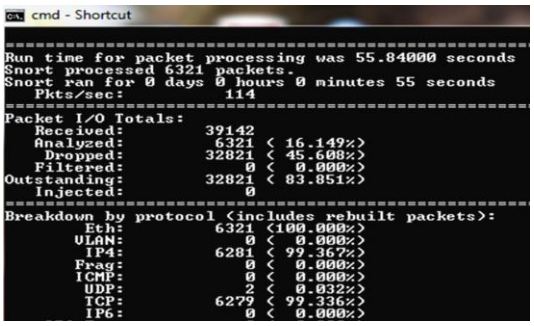


Figure 5 Testing Without QoS (Waleed, et al., 2015)

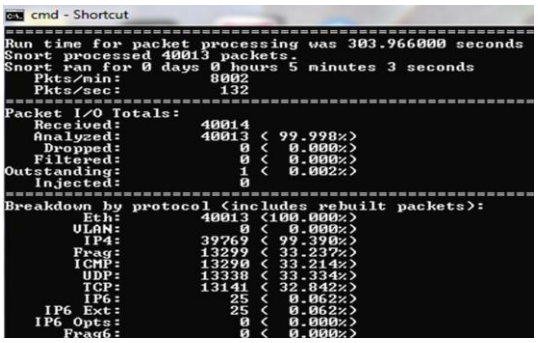


Figure 6 Testing With QoS (Waleed, et al., 2015)

Waleed Bul'ajoul (2015) demonstrated a solution to the problem by layering multiple systems to cope with the volume of information that is sent, by the means of multicore processors (Figure 7). There is an extensive improvement to the number of packets that are analyzed, in this case 100%.

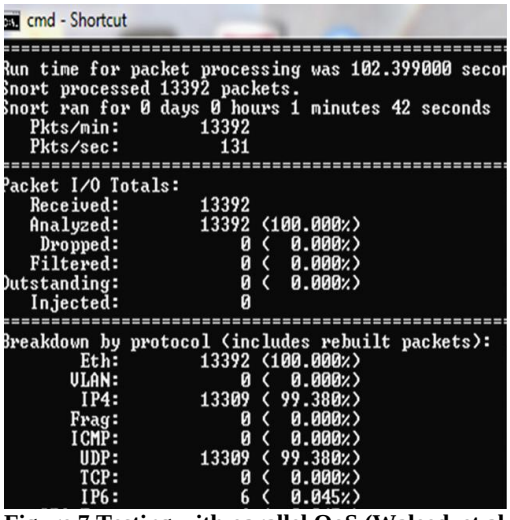


Figure 7 Testing with parallel QoS (Waleed, et al., 2015)

Waleed Bul'ajoul (2015) believes that with the use of different processing powers there will be a greater achievement of NIDPS Figure 8.

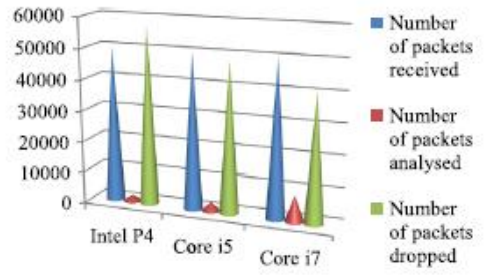


Figure 8 Snort using different processors (Waleed, et al., 2015)

"IDPS is considered to be one of the best technologies to detect threats and attacks" (Waleed, et al., 2015) He suggests that with further research into the development into multicore processors, advancements can be made to NIDPS to be able to cope with higher volumes and faster attacks on networks.

Waleed Bul'ajoul (2015) theory on further research on the improvement of multi-layer processors to handle increased packets sent into a network, is well thought out and the testing conducted to prove this on the theory is well produced. Waleed Bul'ajoul (2015) approach to testing his method for optimizing QoS in NIPDS is well justified with thorough testing using Snort, Omar Al-Jarrah and Ahmad Arafat (2015) conducted similar experiments that further strengthen Waleed Bul'ajoul (2015) theory. E.Bharathi and Dr.A. Marimuthu (2014) also agrees that with the development into QoS in multicasting it would be able to satisfy the requirements of high-speed information networks.

To further develop their method and theory they could use different IDPS instead of using Snort. They could also use different equipment in place of Cisco Catalyst 3560 to see how different equipment is able to handle their method, giving a wider range of scientific justification. To conclude, Waleed Bul'ajoul (2015) method is well presented with evidential testing to back their theory.

4 Data Filtering

Weizhi (2014) focuses on the use of Data Filtering to be able to optimize the performance of intrusion detection systems. Weizhi (2014) believes that IDS suffers from specific limitations in large scale network environments, these being packet overload, expensive signature matching and false alarms. Weizhi (2014) aims to develop an enhanced filter mechanism (EFM) to be able to filter good information from bad, focusing in three specific areas, these being;

- Context-aware blacklist-based packet filter
- Exclusive signature matching component
- KNN – based alarm filter

Focusing on these three areas they aim to combine them in working together to be able to enhance performance (Figure 9). They based their experiments on the software Snort where they used two different data sets, one being DARPA (lab environment) and the second using real live data. Their first testing on DARPA shows the improvement with the use of their EFM in place (Figure 10).

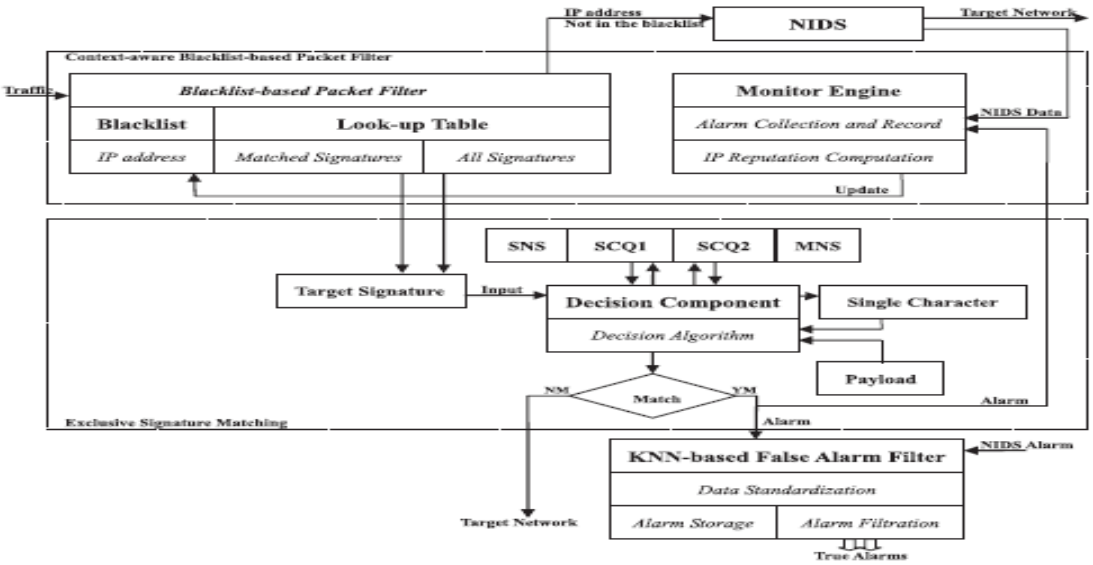


Figure 9 EFM (Weizhi, et al., 2014)

Table 1 – Evaluation results on the DARPA data set: the time consumption of Snort under *Situation1* and *Situation2*.

Week day	Week2	Week4	Week5
Situation1: Snort time consumption (s)			
Monday	14.11	7.33	13.30
Tuesday	16.72	–	17.86
Wednesday	5.58	13.75	16.36
Thursday	13.61	16.87	27.91
Friday	11.62	11.56	38.50
Situation2: Snort time consumption (s)			
Monday	10.42	5.67	10.43
Tuesday	12.51	–	13.27
Wednesday	4.34	10.24	12.66
Thursday	9.85	10.67	18.50
Friday	9.40	8.56	27.45

– Situation1: deploying Snort without the EFM.
– Situation2: deploying Snort with the EFM.

Figure 10 Improvement from EFM (Weizhi, et al., 2014)

The testing done was conducted over a five week period, where three of the weeks were used. With their use of signature matching component, they were able to get more accurate matching with the use of Snort as seen in Figure 11.

Table 3 – Evaluation results on the DARPA data set: the time consumption of Snort and the *exclusive signature matching component*.

Week day	Week2	Week4	Week5
Time consumption of Snort (s)			
Monday	14.11	7.33	13.30
Tuesday	16.72	–	17.86
Wednesday	5.58	13.75	16.36
Thursday	13.61	16.87	27.91
Friday	11.62	11.56	38.50
Time consumption of exclusive signature matching (s)			
Monday	9.42	4.96	9.12
Tuesday	13.15	–	12.45
Wednesday	4.42	11.05	11.89
Thursday	9.65	11.20	17.44
Friday	8.35	9.43	28.43

Figure 11 EFM using exclusive signature matching (Weizhi, et al., 2014)

Weizhi (2014) managed to save time in the filtering of files on their DARPA data sets by using his recommended method (Figure 12).

Table 4 – Evaluation results on the DARPA data set: the percentage of time reduction by comparing *Snort* with the *exclusive signature matching component*.

Percentage of time reduction (%)			
Week day	Week2	Week4	Week5
Monday	33.24	32.33	31.43
Tuesday	21.35	–	30.29
Wednesday	20.79	19.64	27.32
Thursday	29.10	33.61	37.51
Friday	28.14	18.43	26.16

Figure 12 Percentage of time saved. (Weizhi, et al., 2014)

For their second set of data Weizhi (2014) used real data sets based on Honeypot which is was deployed in their lab a CSLab. They had to use the software WireShark to be able to capture packet information as it is filtered through EFM. The experiments that were conducted were based on the same format as DARPA, however the tests were based on information of three days as seen in figure 13.

Table 8 – The percentage of time reduction of *Snort* with the *exclusive signature matching component* on the real data set.

Percentage of time reduction			
Real data set	DAY1	DAY2	DAY3
Percentage (%)	35.76	38.48	32.81

Figure 13 Real data set percentage of time reduced (Weizhi, et al., 2014)

Weizhi (2014) believes that with the optimization of the three main factors discussed above this will to enhance the performance of NIDS with the utilization on EFM. The results that they demonstrate shows that EFM can provide overall improvement without affecting the whole network security.

However, Weizhi (2014) makes claims that DARPA is the only well labeled data set that is available even though it is over ten years old. As seen in different experiments by Mahbod, et al (2009) this is not the case. This could be due to bias by Weizhi (2014) , partly due to a pervious evaluations they conducted with some of their previous work. Weizhi (2014) fails to state the amount of information that is being used for their DARPA tests which they did with their Honeypot data and claims that no results were present over several different weeks, resulting in Weizhi (2014) selecting their data range. The DARPA data set that was used was conducted over a three week period, where as their second source of data from Honeypot was only conducted over three days. There was a large amount of information that was presented from the Honeypot data however it is an unfair experiment, because of this change of time in data.

Although the data sets that were used by Weizhi (2014) seem biased the experiments they conduct on the information they had, was vigorously tested and with further development Weizhi (2014) EFM's theory could potentially optimize IDS's.

5 Comparison

There are comparisons that can be made between the three theories that have been presented in this paper, these being: Parallel Technologies, Quality of Service and Data Filtering. Although there are separate technologies optimizing specific areas of protection, if Salma Elhag et al (2015), Waleed Bul'ajoul (2015) and Weizhi (2014) were to take into account these

different areas, they would be able to further enhance their theories into network security. If Weizhi's (2014) theory of Data Filtering, extracting good data from bad, was layered with Waleed Bul'ajoul's (2015) theory of optimizing QoS in multicore processors, it would give a greater return in time taken for the system to process data, with a majority of the bad data already being filtered.

6 Conclusions

Intrusion Detection Systems (IDS) are a vital area in being able to help prevent attacks on networks. However, as networks are never 100% secure due to the development of new methods of attacks, continuous research is needed to be able to fight this ongoing battle.

In this paper there has been critical evaluations on three different types of Intrusion Detection Systems; Parallel Technologies, Quality of Service and Data Mining. Although these are not the only areas that are important to IDS they provide key fundamentals to areas that are of great importance to protect systems and their users.

Salma Elhag et al (2015) has taken an approach into Parallel Technologies that were the focused on Fuzzy Associated Rule-Based Classification for High-Dimensional One-vs-One (FARCHD-OVO) systems was presented. They demonstrated extensive test results from their method that was used and showed a clear improvement to intrusion detection systems. Although their testing was well generated, their lack of resources used has flawed their method. If they were to conduct further experiments on more data sets, their method would be further justified.

Optimizing Quality of Service in IDS to enhance performance was introduced by Waleed Bul'ajoul (2015) where he took the approach of using multicore processors to be able to handle large amounts of data as well as high speed data sets in DoS and DDoS. The research that was conducted by Waleed Bul'ajoul (2015) was focused on their selected area, because of this they demonstrated some well justified tests to back up their theory, although they did not conduct tests on different types of equipment to provide further evidence. The results they developed could potentially be used in a real life environment, to increase security systems.

The research conducted by Weizhi (2014) looked into the use of Data Filtering to separate good data from bad data that is sent to a network. The experiments the authors conducted appeared to be well presented. However, there were major flaws in the way these tests were carried out due to the claims made about only being able to use one data set which seemed to be biased towards their previous work they did. Although the concept on filtering data for networks is a key role, Weizhi (2014) fails to 100% accomplish this with his approach to the solution.

The quality of the research that has been presented in this paper concludes that the authors discussed have shown to have made contributions to IDS's. However, more testing and research needs to be conducted in all areas that have been looked into. Nonetheless if the work by Salma Elhag et al (2015) and Waleed Bul'ajoul (2015) were put together if could make a valuable contribution in network security.

References

Amal, S. & Sugata, S., 2014. 'Application Layer Intrusion Detection with Combination of Explicit-Rule-Based and Machine Learning Algorithms and Deployment in Cyber- Defence Program'. *Advanced Networking and Applications*, 6(2), pp. 2202-2208.

Chen, T. & Xu Zhang a, S. J. b. O. K., n.d.
E, B. & Dr.A, M., 2014. 'QoS based hybrid swarm intelligent intrusion detection system for network security'. *Journal of Theoretical and Applied Information Technology*, 69(2), pp. 257-274.

Kelton, A.P. Costa; Luis, A.M. Pereira; Rodrigo, Y.M. Nakamura; Clayton, R. Pereira c; João, P. Papa; Alexandre, Xavier Falcão., 2015. 'A nature-inspired approach to speed up optimum-path forest clustering and its application to intrusion detection in computer networks'. *Information Sciences*, Volume 294, pp. 95-108.

Mahbod, T., Ebrahim, B., Wei, L. & and Ali A, G., 2009. *'A Detailed Analysis of the KDD CUP 99 Data Set'*. Ottawa, ON, IEEE, pp. 1-6.

Niva, D. & Tanmoy, S., 2014. 'Survey on Host and Network Based Intrusion Detection System'. *Advanced Networking and Applications*, 6(2), pp. 2266-2269.

Omar, A.-J. & Arafat, A., 2015. 'Network Intrusion Dectection Systems Using Neural Network Classification of Attack Behavior'. *Journal of Advances in Information Technology*, 6(1), pp. 1-8.

Salma, Elhag; Alberto, Fernández; Abdullah, Bawakid; Saleh, Alshomrani; Francisco, Herrera.,

2015. 'On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems'. *Expert Systems with Applications,* 42(1), pp. 193-202.

Tieming, C., Xu, Z., Shichao, J. & Okhee, K., 2014. 'Efficient classification using parallel and scalable compressed model and its application on intrusion detection'. *Expert Systems with Applications,* 41(13), pp. 5972-5983.

Waleed, B., Anne, J. & Mandeep, P., 2015. 'Improving network intrusion detection system performance through quality of service configuration and parallel technology'. *Journal of Computer and System Sciences,* 81(6), pp. 981-999.

Wei-Chao, L., Shih-Wen, K. & Chih-Fong, T., 2015. 'CANN: An intrusion detection system based on combining cluster centers and nearest neighbors'. *Knowledge-Based Systems,* Volume 78, pp. 13-21.

Weizhi, M., Wenjuan, L. & Lam-For, K., 2014. 'EFM: Enhancing the performance of signaturebased network intrusion detection systems using enhanced filter mechanism'. *Computer Security,* Volume 43, pp. 189-204.

An Evaluation of Current Research on Data Mining Techniques in Decision Support

Keamogetse Mojapelo

Abstract

Data mining, which is also known as knowledge management, is one of the core knowledge discovery in today's databases of decision support systems. For a very long time, businesses have been using decision support systems only to query management reports instantly, like OLAP which is mostly used for decision making that is effective. The growth of complex data today has led to the use of data mining techniques for business forecasting and determining customer's preferences and business trends. This paper provides a thorough evaluation of data mining techniques to give reliable business trends and patterns in order to make informed decision making in large and complex databases.

1 Introduction

Today's businesses are categorised by huge amount of data needed for daily decision making support. There are a lot of applications that are being used to extract data or information from its data sources that are different. In this fast pace of live, data sources needs to be queried instantly so that operations of the business are controlled. As such the expansion of businesses has led to rise of data management systems to be put in place to handle transactions and analytical queries.

Data warehousing has proved to be a core solution in decision making in organisations of businesses (Elena et. al. 2012). Abello and Romero (2008) adds on to say that although data warehouse seems to be a solution, it solitary contains managerial devotions of data as such it only has limited intelligent mechanisms used for a business when it comes to decision making and prediction.

A major limitation in the Decision Support Systems is that they can only provide query reports of historical data and what is happening at the present but are insufficient in providing the future events or draw patterns from given data. As such it is difficult to make clear patterns from data that is being stored in real time databases and for forecasting in the business.

If this problem goes unsolved, businesses would be unable to discover information that is valuable which would be hidden in their data ware houses by converting this data to be more valuable and useful for knowledge, as such it could be hard to manage knowledge resource.

A call has been made to provide support systems that will be combined with data mining techniques because for many years companies have been using these techniques on separate and it was difficult to predict trends on the consumer's preferences to yield better profits.

With this research, evaluation of others researchers' work towards decision support through data mining will be evaluated to determine if there is a chance of improvement in business trends to opportunity discovery. Data mining with conjunction with decision support techniques will be fully evaluated in accordance to recent research. Lastly a conclusion will be derived based on the findings found on current decision support systems of data mining on previous sections.

2 Evaluation of Algorithms: Decision Support Systems using Data Mining Techniques

Organisations today have large amount of data which is useful for management reports, (Srinivasan 2011). Research shows that many organisations today have massive volume of data but they are unable to query and identify information that is valuable.

However Srinivasan (2011) suggests knowledge management as a way of data mining. This is because it has the capability to extract useful knowledge from large datasets.

This implementation shows that it can demonstrate the business intelligence of an organisation that is common for all data miners and also it can be used as a tool to further extend the knowledge the business has. According to Swathi et. al. (2012) data mining has appeared to have the capabilities of extracting data, transform data, and load transactions and even manage the data. All this could happen in a system that is multi-dimensional.

The sections that follows looks at different data mining techniques that are applicable to Decision Support Systems for future forecasting and pattern recognition in business intelligence.

2.1 DSS and Self Organizing Maps

Silwattananusam and Tuamsuk (2014) proposed a Self-Organizing Map (SOM) which is related to neural networks. This algorithm is explained to be a trend analysis with the capability to recognize customer patterns in web mining for businesses that uses online services.

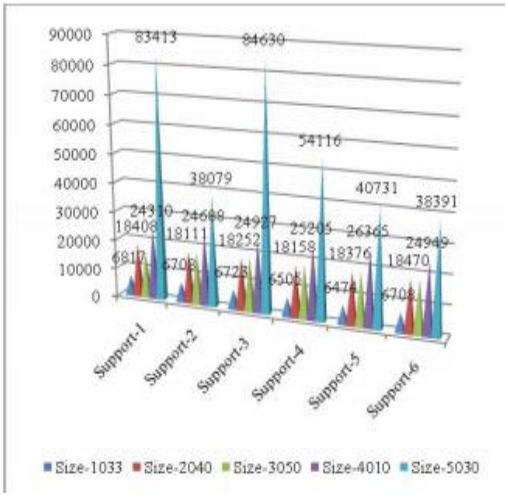


Figure 1 Running time in milliseconds (ms) on different support of different sizes (Sahu et. al. 2014).

Authors' states that every item on the above diagram belong to one page having a maximum and minimum Support value. If Support is less than and equal to a given Support, pruning is done to determine how frequent item is. Every Support is plotted against time taken.

The authors have failed to show that support with a higher size affects performance in real time databases so prediction of trends on customer preferences will take long time hence affecting decision making.

Support→	1	2	3	4	5	6
WSpan	83413	83210	84630	83397	84645	83990
FSTSOM	82009	83100	80745	54116	40731	38391
Percentage Improvement in Execution Time (in ms)	2%	0%	5%	35%	52%	54%

Table 1 WSpan and FSTSOM algorithms having different Support with a recorded size of 5030 that is being used (Sahu et. al. 2014).

The above graph (figure 1) shows functionality analysis with Support 1 through 6 suggested by (Silwattananusam and Tuamsuk 2014). The author states WSpan as the newly proposed algorithm to be pattern mining algorithm. Silwattananusam and

Tuamsuk (2014) concludes that WSpan algorithm is the fastest mining algorithm used for pattern recognition. The author's work shows that transversal patterns that are sequential with minimum/maximum weights can be unified with a web page and a support with a clustering that is effective.

2.2 Back Propagation and K-Means Algorithms

Mathuriya and Bansal (2012) proposed back propagation as a neural network method that is being used to mine data so that analysis can be made on the effect of technologic parameters that are structural. The data set and input attributes are determined by the process of knowledge engineering.

```
1. Initialize the weights in the network (often randomly)
2. repeat
* for each example e in the training set do
1. O = neural-net-output (network, e);
Forward pass
2. T = teacher output for e
3. Calculate error (T - O) at the output units
4. Compute delta_wi for all weights from hidden layer to
output layer;

Backward pass
5. Compute delta_wi for all weights from input layer to
hidden layer;
backward pass continued
6. Update the weights in the network
* end
3. until all examples classified correctly or stopping criterion
satisfied
4. return (network)
```

Figure 2 K means algorithm used in mining data for analysis, (Mathuriya and Bansal 2012).

These algorithms were evaluated and experiments were carried on them on the collected data sets. Accuracy was used to evaluate classification and clustering algorithm. The accuracy to be determined was the ratio of records that are correctly classified when testing's performed in response to records tested.

Two datasets were used for experimentation. Four hundred (400) records were used for the first data set while the second dataset had about eight hundred and fifty (850) records. 60% of the records in a dataset were placed in the rejected category and machine learning algorithms were able to determine the rejected data. The data set was planned and had the same number of records for the two experiments of two datasets. The machine learning algorithms had failed to notice records that were selected for a large extent.

Algorithm used	Accuracy %
K- means	72
Back propagation	80.40

Table 2 Algorithms results in which training and testing were done on Dataset1 (Mathuriya and Bansal 2012).

Algorithm used	Accuracy %
K- means	86.23
Back propagation	91.42

Table 3 Trained and tested results on Dataset2 (Mathuriya and Bansal 2012).

Algorithm used	Accuracy %
K- means	71
Back propagation	75.46

Table 4 Algorithm results which are trained and tested with Dataset 1 and Dataset2 (Mathuriya and Bansal 2012).

MATLAB was used to implement K Means and back propagation algorithms and Tables 2, 3 and 4 shows how accurate the results are. The results show that K means algorithm has poor accuracy hence its nature of data is not suitable for the problem area. The results demonstrates that back propagation yields better results as compared to K means algorithm.

Looking at this paper evidence has shown that good experimental skills were used for all the experiments that were carried out in an environment that was controlled. All of the experiments were carried multiple times so this reduced randomness in the output. These authors have practiced good science as their results are reproducible.

2.3 Neural Networks

Research by Oancea and Ciucu (2012) shows that using neural network can approximate functions that are continuous in a business. They further elaborate that neural networks have been used for some time to forecast financial data series in companies. The authors propose two networks being feed forward and recurrent neural networks and training algorithms in order to predict the exchange rate of currencies.

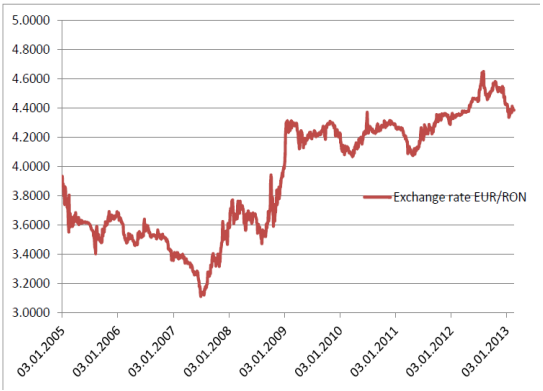


Figure 3 Data series results from the Bank of Romania (Oancea and Ciucu 2012).

The above graph is showing the corresponding exchange rates of EURO/RON currencies. Since the data is rough forecasting is difficult to be made so the following functions removes correlation.

$$R_n = \frac{\ln E_n}{\ln E_{n-1}}$$

Figure 4 Function used to remove correlations, (Oancea and Ciucu 2012).

For the learning rate to be improved, data pre-processing was added to make sure that data distribution was smoothed before any training was performed on the network. A logistic function below was used to normalize the data:

$$\tilde{R}_n = 1/(1 + \exp(\frac{R_n - \bar{R}}{\sigma}))$$

Figure 5 Logistic function used to normalize data, (Oancea and Ciucu 2012).

Oancea and Ciucu (2012) carried out experiment to determine the performance of these two algorithms. They used data series having exchange rates for the day provided by Bank Romania.

A feed forward neural network was built having an input layer, hidden layer and an output layer. Twenty (20) neurons were used as the optimum number in the input layer while forty (40) neurons in the hidden layer. A linear activation function was used for the input layer and output layer. Data sets were divided in to two groups in which 80% was used to train the data set and 20% used for test. Three algorithms were tested being backpropagation, iRPROP (resilient propagation version) and resilient propagation.

The results for this three training algorithms were similar and they showed an increase of speed in convergence with about 25% for IRPROP.

Recurrent neural network was also tested and was tested using Extended Kalman Filter (EKF).A multi-stream approach was used for training. Twenty (20) neurons were used as the input and ten (10) neurons were used in the hidden layer.

The training data was split in to 20 neurons each stream having two hundred (200) points. The error for forecasting was about 0.01% which is small so the results are reliable.

Looking at how these authors have carried out their experiments, conclusions can be that the research

carried out was good because more than one algorithm was used for forecasting exchange rates .
so this improves results obtained. Also since the dataset used for experimenting was in a real world environment from Bank of Romania, this made the tests to be unbiased.

However even though the authors (Oancea and Ciucu 2012) do not show a detailed procedure used for training the data, it is impossible to tell if the correct steps were taken to reach the final results even though it may show as if the results are credible. Again these authors have failed to show that the experiments were performed a numerous times, rather they only state that three (3) training algorithms were used for testing.

2.4 CRM Framework

In addition to the enrichment work using the prediction technique, (Femina and Sudheep 2014) presented a technique known as the CRM framework which is data mining frame work. Its strategy was to explore the usefulness of two used classification models that are used in data mining to forecast how customers behave in CRM application of a business.

In their experiments, (Femina and Sudheep 2014) compared performance measure of Naïve Bayes classifier to Multi-layer Perceptron Network (MLPN). To access performance, metrics such as sensitivity analysis and accuracy rate specificity were used.

Authors's dataset contained 45211 instances from a period of May 2008 to November 2010. Their set up contained a dataset with a percentage of 10% having a preprocessed dataset containing sixteen (16) variables used as inputs. A Weka tool was used in a Windows 7 environment having an Intel core of i3 with about 2.53Hz of a processor. The larger the database of the Naïve Bayes the better accuracy it computes and faster it is to train the models.

Authors used ten-fold cross validation to test the two models. The number of hidden layer is set using the heuristic defined as *a=round* (*M/2)* with *M* being the sum of both the classes and attributes.
The diagram below shows the results taken from the experiment classifying the given dataset with values for two given classifiers. Each classifier has its accuracy rate, True Positive rate which shows the amount of actual positives and False Positive showing incorrect positives that are classified as correct as well as the time taken to build the model

Classifier	Classification Accuracy(%)	True Positive Rate(TPR)	False Positive Rate(FPR)	ROC Area	Time taken to build models (s)
MLPNN	88.63	0.41	0.052	0.847	1767.75
NB	87.97	0.47	0.067	0.858	0.08

Table 5 Comparisons of results of classifiers over 10 runs, (Femina and Sudheep 2014).

Their results showed that the Multi-Layer Perceptron Neural Network (MLPNN) had a better accuracy over Naïve Bayes (NB) classifier. The authors' conclusions are that the tests shows that Naïve Bayes yields better True Positive Rate (actual proportions which are defined correctly) value as compared to the MLPNN classifier. Also from the results given, Naïve Bayes takes less time to build the model as it only takes 0.08 seconds compared to MLPNN which took about 1767.75 seconds.

Looking at the methodology used in this research, it has been observed that two different means of evaluating the experiment was used, as such consistency in the results was shown. The setup of the experiment was in a controlled way the tools used eliminated the factors that could lead to bias. Their methodology is properly documented so that other researchers may repeat it to improve results.

2.5 Particle Swarm Optimization

Just recently a new contribution to Decision Support Systems in banking business by the use of prediction models was proposed. Lin et. al. (2011) suggested a Particle Swam Optimization as a performance measure for banking businesses. They argue that the purpose of this model was to simulate social behavior in order to gain objectives that were precise for a given space that is multi-dimensional.

This research was aimed at discovering the optimal solution for particle change in searching direction to two factors that are being given which are the particle best experience and the best experience given all members. Particle Swarm Optimization (PSO) was implemented by the authors for determining the parameter and feature selection of the decision tree and for the Support Vector Machine (SVM).

To carry out their experiments to test accuracy measure on bank performance prediction, these authors used a Personal Computer (PC) that had a RAM of about 512 MB, an Intel Pentium IV having 3.0 GHZ and a XP Windows Operating system. All of their entire experiments were carried on Visual C++ environment.

The authors stated that two datasets were collected from a bank being, Bank of China and the dataset from this bank were derived from 44 banks in China from the year 2006 to 2010 and they were about 220 records. A cognitive learning factor, *C1* was used as well as the social learning factor, *C2* for PSO Support Vector Machine. These learning factors were set to 2 and 1 respectively. The learning factors for Particle Swarm Optimization for Decision Tree (PSO-DT) was set to 1, 5 and 2 in respective to *C1* and *C2* learning factors.

Authors stated that the number of iterations and particles was set to be between 10 and 1000 catering for both datasets. These authors used a k-fold approach to determine the accuracy rate of the classification. This k-fold approach divided the data into five (5) portions. Portions that were made through this division were categorised to ratio of classification of the original data.

Dataset	Fold	PSO–SVM			Grid search		
		C	γ	Accuracy (%)	C	γ	Accuracy (%)
1	1	28712.4528	0.1003	93.182	32768.0000	0.0001	81.818
	2	44.9900	0.1409	90.909	32768.0000	0.0078	93.181
	3	2240.2050	0.0020	97.727	8192.0000	0.0078	90.904
	4	44.9900	0.0020	84.091	32768.0000	0.0020	75.000
	5	7469.0438	0.0106	88.636	20488.0000	0.0078	81.818
	Avg.	-	-	90.909	-	-	84.544
2	1	44.9900	0.0578	90.244	32.0000	0.1250	87.804
	2	17505.0000	0.5005	85.366	2048.0000	0.0004	78.048
	3	44.9900	0.1030	82.927	2048.0000	0.0078	73.170
	4	44.9900	0.0020	82.927	32.0000	0.1250	82.822
	5	1519.5752	0.3156	77.500	8.0000	0.1250	60.000
	Avg.	-	-	83.793	-	-	76.369

Table 6 Results of Particle Swarm Optimization (PSO) against Grid Search with no feature Selection (Lin et. al. 2011).

Lin et. al. (2011) observed that the classification rate average of proposed PSO-SVM has improved classification accuracy of about 90.909% and 83.793% for dataset 1 and 2. The Grid search had

a classification accuracy of 84.5% for dataset 1 and 76.37% for dataset 2. The authors concluded that Particle Swarm Optimization had a better accuracy as compared to the Grid search in which it was tested against with the selection feature in inclusion.

Looking at research by (Lin et. al. 2011) it can be concluded that the authors clearly stated the hardware requirements they were going to use for their experiments. They also provided comparisons between PSO-SVM and Grid Search to improve reliability of the results got. However, their work does not clearly state the other dataset they were using to determine the bank performance prediction because they only mentioned the dataset collected from the Central Bank of China only. The number of records for the second dataset are also unknown so it is impossible to specify how many input variables were used and how many output variables were also used also this makes the outcomes of the results biased because it is not known what has happened to the other records.

3 Lessons Learnt

Business intelligence has seemed to a very big broad aspect in terms of analytics and mining of data. Data mining has proved to be important in spreading and increasing the chances of decision making through the use of discovering the very important relationships that are being hidden and so as their patterns which enhances the chances of decision making in categorised. Data mining in decision support uncovers unknown patterns through its techniques and it has addresses problems that seemed to be difficult to diagnose in the categorised. OLAP works on multidimensional data and it extract specific knowledge from the database.

4 Conclusions

This paper has shown that the scramble in participation of businesses today has increased data usage at a high alarming rate, hence making data mining a very important aspect in industries. Different algorithms have been tested throughout the body of this paper, and it has been shown that the effectiveness of decision support systems for today's businesses is not only based on extracting useful information instantly for decision making, but also for drawing patterns to make informed future recommendations using classification technique. It is being anticipated that the presentation of data mining techniques will improve the power of understandable correlations as well as patterns in data that is existing in data warehouses.

References

Abello A., Romero O., 2009, 'Online- Analytical Processing; Data Integration Flow for Business Intelligence' *International Conference on Engineering and Knowledge Management,* vol.145, no.12, pages 342-387.

Elena C., Parpandel D., Popa I., 2012, 'OLAP Queries and Parallel Processing', *IEEE 28th International Journal of Information and Decision Making,* pages 3456-3721.

Hamad D., Qader B., 2013, 'Knowledge-Driven Decision Support System Based on Knowledge Warehouse and Data Mining for Market Management', *International Journal of Application or Innovation in Engineering and Management* (IJAIEM), vol.3, Issue.1, pages 139-147.

Lin S., Shiue Y., Chen S., Cheng H., 2011, 'Applying Enhanced Data Mining Approaches in Predicting Bank Performance of Taiwanese Commercial Banks', *Expert Systems with Applications, International Journal of Computer Technology and Electronics Engineering (IJCTEE),*Vol.3, No.2, pages 11544-11550.

Mathuriya N., Bansal A., 2012, 'Comparison of K-means and Backpropagation Data Mining Algorithms', *International Journal of Computer Technology and Electronics Engineering (IJCTEE*), vol.2, Issue 2, pages 151-155.

Silwattananusam T., Tuamsuk K., 2012, 'Data Mining and its Applications for Knowledge Management: A Literature Review from 2007 to 2012', *International Journal of Data Mining and Knowledge Management Process (IJDKP*), Vol.2, No.5, pages 13-29.

Srinivasan S., Singh S., Kumar V., 2011, 'Multi-agent Based Decision Support Systems using Data Mining and Case Based Reasoning', *International Journal of Computer Science Issues,* vol.8, Issue. 4, no.2, pages 340-349.

Swathi B., Praveena M., Kavitha L., 2012, 'Harmonized Scheme for Data Mining Technique to Progress Decision Support System in an Uncertain Situation ', *International Journal of Research in Engineering and Technology*, Vol.1, No.3, pages 335-338.

Varde A., Rundesteneir E., Riuz C., Manirruzzaman, Sisson R., 2014, 'Data Mining Over Graphical Results of Experiments with Domain Semantics', *5th International Conference on Electrical and Computer Engineering* , vol.3, pages 1-9.

A Critical Investigation of the Cognitive Appeal and Impact of Video Games on Players

Kealeboga Charlie Mokgalo

Abstract

Over the years various video games have become an increasing part of our day to day lives, this has raised awareness and their influence has been discussed in recent research. This paper‘s focal point is to address the psychological lure players have towards respective video games, and discover why they are subconsciously drawn to playing specific video games. The report will also highlight the psychological impact or behavioural effects of respective video games on players. It considers, compares and critically evaluates all formal and informal forms of psychometric scales developed through the years to determine the proposed appeals and effects of video games on players. The Brainhex model and HEXACO model will be critiqued as they are the cornerstone for the analysis of cognitive pulls and effects in video game players. Furthermore the misconceptions about post-game aggression and human cognition in players will be put into perspective.

1 Introduction

Human cognition in video game players has for years been marred by uncontested assumptions and invalid findings. Video Game Design requires game developers to develop models that are used in Human-Computer Interaction (HCI) to identify player-video game interactions which are yet to be clearly defined in the scientific literature. This paper focuses on current research about player-video game interactions used to determine the cognitive motives players have for playing video games and the cognitive effects incurred thereof.

According to Coary et al. (2015) general types of players and traits of play must be identified in order for game designers to “understand their players”. Numerous video game researchers have established psychometric scales to measure the motives that people have for playing video games. These psychometric scales are usually limited by their focus on specific gaming genres, or players’ culture and philosophy, as well as “their lack of behavioural validation” (Caroux et al. 2015).

The onus of the study is to explore the results video game experience may have on an individual’s way of thinking or behaviour, considering people have different personalities, the field is volatile.

2 Players’ Motivations for Playing Video Games

This section is designed to review the proposed models that present the cognitive motivations players may have towards video games.

2.1 The Brainhex model

Bateman et al.’s (2014) Brainhex model presents seven different player typologies which are represented in Figure 1 and Figure 2 on the following page.

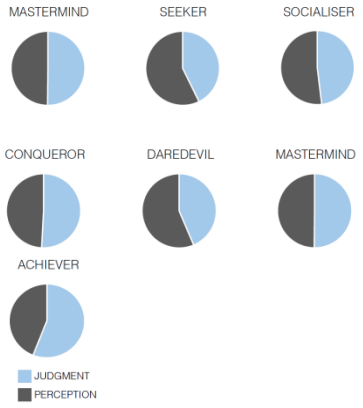


Figure 1 depicts the Judgement and Perception groups broken down by BrainHex primary archetypes (Bateman et al. 2014).

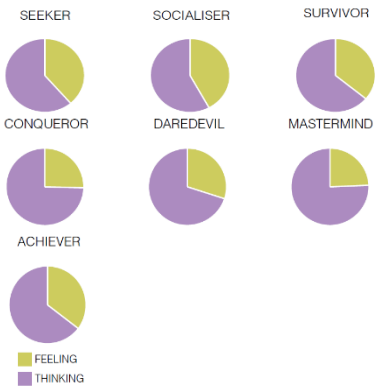


Figure 2 above illustrates Feeling and Thinking groups broken down by BrainHex archetypes (Bateman et al. 2014).

This paper has placed much emphasis on these archetypes, as they propose different types of cognitive pulls players may have towards respective video games.

Seeker

The "Seeker" is motivated by curiosity and interest, which are closely associated to the part of the brain "processing sensory information and the memory association area" (Bateman et al. 2014). The study suggests the brain releases "endormorphin, which triggers the pleasure centre" (Bateman et al. 2014). The "Seeker" can be related to explorative play and inquisitiveness "fostered for by attention to detail and role-play" (Bateman et al. 2014). As depicted in Figure 2, current research revealed the "Seeker", along with the "Survivor", "Achiever" and "Socialiser" archetypes have a greater preference for feeling than the other archetypes. This suggests that players fitting this archetype are less likely to pursue skill-oriented games, rather they play games for "moments of jaw-dropping wonder or beauty'' (Bateman et al. 2014). In Figure 1 the "Seeker" is depicted to have a greater preference to perception than the other proposed player typologies. This implies that players under this archetype are unlikely to be attracted to strategic, decision-making games.

Survivor

According to Bateman et al. (2014), the "Survivor" enjoys taking gameplay risks in search of a thrill. It can also be argued that this typology bares hallmarks of the "Seeker", as the player finds subjective fulfilment in feeling terrorized within gameplay. Bateman et al. (2014) support this claim by suggesting that epinephrine "enhances the effects of dopamine, the neurotransmitter triggered when rewards are received". As depicted in Figure 2 the "Survivor" along with the "Seeker", "Achiever" and "Socialiser" archetypes had a greater preference for feeling than the other archetypes. Figure 1 depicts that this archetype, same as the "Seeker", "Achiever" and "Socialiser" archetypes has a greater preference for perceiving than the other archetypes. From these results we can deduce that players fitting this typology will pursue terrifying games which provide moments of "heart-stopping fear while escaping" (Bateman et al. 2014).

Daredevil

As with the "Survivor" archetype, the "Daredevil" archetype enjoys a thrill from taking risks in gameplay solely as a positive experience. Again the neurotransmitter epinephrine can be seen as a reward enhancer. As shown in Figure 2, the study revealed the "Daredevil" archetype was less susceptible to feelings than the "Seeker", "Achiever", "Socialiser" and "Survivor" archetypes. Furthermore, Figure 1 depicts a lack of judgement preference amongst the "Daredevil" archetype therefore players fitting this typology will seek games that require little judgement but offer moments "of speed or vertigo" Bateman et al. (2014).

Mastermind

Players who fit into the "Mastermind" typology enjoy solving puzzles and devising decisive strategies. Bateman et al.'s (2014) research reveals, "The close relationship between the brain's decision centre and pleasure centre ensures that making good decisions is inherently rewarding" therefore the player may find favour in strategic games of a decision-making nature. Figure 2 depicts the "Mastermind" archetype as dominated by the thinking preference. Furthermore, Figure 1 depicts both Judging and Perceiving preferences of this archetype equally represented without differences. These results suggest that the "Mastermind" archetype will prefer games that make them think of solutions to problems and challenge them strategically.

Conqueror

According to Bateman et al. (2014) players who fall into this archetype are not "content with winning easily, they want to struggle against adversity". The study reports that when humans are faced with challenging situations, "their body produces epinephrine (adrenalin) and norepinephrine, the former producing arousal and excitement and the latter being associated with combative tendencies" (Bateman et al. 2014). These combative tendencies

and testosterone serve as a motivation for players to pursue difficult games and "persist in the face of challenging gameplay situations" (Bateman et al. 2014). Figure 2 suggests the "Conqueror" is dominated by the thinking preference, whereas Figure 1 reveals the Judging and Perceiving preferences of this archetype are equally represented without differences. It can be deduced that players fitting this typology are drawn to difficult games in which they can gain hard-fought victories.

Socialiser

The Brainhex model proposes that people are a major source of enjoyment for players fitting this typology. The research suggests that people connect to this archetype's social centre, "the principal neural source of oxytocin, a neurotransmitter in the brain demonstrated to have a connection with trust" (Bateman et al. 2014). This archetype along with the "Seeker", "Achiever" and "Survivor" are dominated by the feeling preference as illustrated in Figure 2. Additionally, Figure 1 depicts that both judgement and perception preferences were equally represented without differences in the "Socialiser" archetype. These results suggest players falling under this archetype play games with socialising as a primary motive, as they seek games that make them feel "an intense sense of unity with another player" (Bateman et al. 2014).

Achiever

The Brainhex model presents the "Achiever" typology which is primarily goal-oriented but long term in thinking. Subjective reports from players in the study considered under this archetype revealed their play is obsessive in its focus and they prefer games "amenable to ultimate completion, especially digital Role Playing Games (RPGs), who's self-adjusting difficulties ensure completion as a result of perseverance" (Bateman et al. 2014). The study reveals this archetype along with the "Seeker", "Socialiser" and "Survivor" is dominated by the feeling preference as depicted in Figure 2. Moreover Figure 1 depicts the achiever archetype as being considerably dominated by judgement. Players fitting this typology will pursue games that validate their gameplay progress as they strive for the challenge of completion. This archetype is supported by a study carried out by Cruz et al. (2015) that reveals a player typology, "The Completionist" where self-identified participants of this typology expressed "the drive not just to finish the game, but to achieve all of the tasks within the game". In Figure 3 the results of a study undertaken by Collins and Cox (2014) on the "average number of hours spent playing digital games per week separated by genre" reveals that 75% of the game genres are dominated by games that offer achievements, rewards and goals to support validation of players' progress. Such games include First-person shooters and a role-playing context according to Collins and Cox (2014).

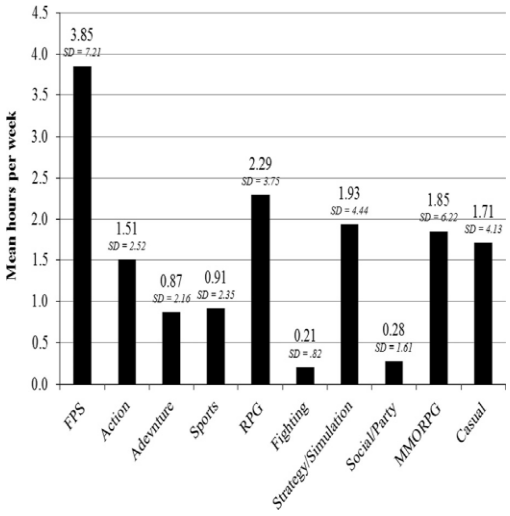


Figure 3 Mean number of hours spent playing digital games per week separated by genre. FPS ¼ first person shooters, RPG ¼ role playing games, MMORPG ¼ massively multiplayer online role playing games (Collins and Cox 2014).

2.2 The HEXACO Model

The HEXACO model is a six-factor model that encapsulates many ways in which individuals differ. The HEXACO model was used in a study by Monica and Zeigler-Hill (2015) as a metric for the basic dimensions of personality presented in the HEXACO model namely, honesty-humility, emotionality, extroversion, agreeableness, conscientiousness and openness to experience. The HEXACO model was applied to Bateman et al.'s (2014) proposed Brainhex typologies.

Honesty-Humility

This dimension was used to capture the degree to which players "exhibit fairness, sincerity, and modesty" Monica and Zeigler-Hill (2015). This dimension was not associated with any of the proposed Brainhex player typologies.

Emotionality

Monica and Zeigler-Hill (2015) used this dimension to capture the extent to which a player will be susceptible to worry, anxiety and other negative emotional states. The research by Monica and Zeigler-Hill (2015)

indicates this dimension has a weak negative association with Bateman et al.'s (2014) proposed "Daredevil" archetype, as its sensation-seeking aspects may be upsetting to players. This dimension furthermore displayed weak negative associations with Bateman et al.'s (2014) "Mastermind" and "Conqueror" player typologies, as well as weak positive associations with the proposed "Seeker" and "Achiever" archetypes. Players highly associated with this dimension are sensitive to emotional experiences and inclined to play explorative games, rather than those which challenge or excite them.

Extraversion

The Extraversion dimension takes into account characteristics such as the sociability, dominance, and talkativeness of players. Monica and Zeigler-Hill's (2015) study revealed that extraversion is closely associated with Bateman et al.'s (2014) "Socialiser" and "Daredevil" archetypes, as it had a fairly positive link with the "Socialiser" preference, and a weak positive association with the "Daredevil" preference as depicted in Figure 4. Extraversion encapsulates the sensation-seeking and sociability motivations some video game players may have. Monica and Zeigler-Hill's (2015) study also revealed that extraversion is weakly associated to the "Survivor", "Conqueror", and "Achiever" player typologies. It can be deduced that players with high levels of extraversion are motivated to play games for excitement, competition and social reasons.

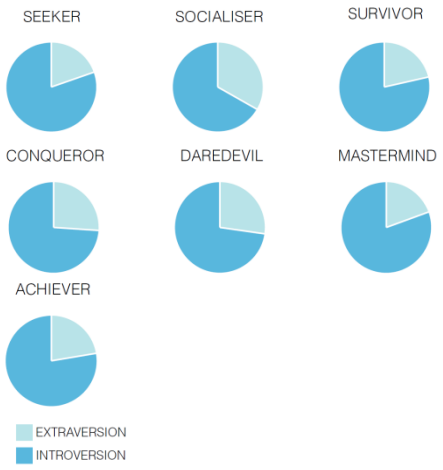


Figure 4 Extraversion and Introversion groups broken down by BrainHex primary archetypes (Bateman et al.'s 2014).

Agreeableness

According to Monica and Zeigler-Hill (2015) this dimension is used to capture player traits that represent the degree of their friendliness, warmth, and cooperativeness. This dimension, like honesty-humility was not associated with any of the Brainhex player typologies.

Conscientiousness

This dimension is defined by players' attributes such as carefulness, self-discipline and reliability. Monica and Zeigler-Hill's (2015) study reflected a weak positive association with Bateman et al.'s (2014) "Achiever" archetype. Results of Monica and Zeigler-Hill's (2015) study suggest that players with high levels of conscientiousness long for games that challenge their efficiency and organisation. The study also revealed that conscientiousness has a weak positive association with the "Mastermind" archetype and a weak negative association with the "Survivor" preference. It can be deduced that for players with high levels of conscientiousness, task-oriented or problem-solving games are more alluring than others.

Openness

This dimension will capture players' cognitive qualities such as curiosity, imaginativeness, and originality. This dimension was found by Monica and Zeigler-Hill (2015) to have a weak positive link to the Brainhex model's "Socialiser" archetype. Unsurprisingly, players with high levels of openness are helpful, trusting and sympathetic, whilst they maintain a preference for cooperative games.

2.3 Social Interaction

Social interaction is considered as a primary motive for video gameplay and can be closely linked to Bateman et al.'s (2014) "Socialiser" archetype. A survey by Herodotou et al. (2014) on "reasons for playing WoW", an online, multiplayer, role-play game reported 67% of respondents played the game for social reasons as depicted in Figure 3. Furthermore, research by Ferguson and Olson (2013) suggests social play was more common among players who felt stressed.

3 The Cognitive Effects of Video Games on Players

This section will discuss research on the cognitive influence video games may have on players. Research on long term and short term effects of video games will be presented.

3.1 Aggressive Behaviour

Research by Drummond et al. (2015) assesses the "short-term and long-term effects of video game

violence on players' cognition and behaviour". In their study, Drummond et al. (2015) use precursors to aggressive responses following aggressive gameplay and claim that aggression is fostered by the increased accessibility to aggression-related gaming scenarios and the continued rehearsal of this behaviour in gameplay. Drummond et al.'s (2015) research inspects the key implications of "postgame aggression flow from in-game aggression, and the generalised activation of aggression-related cognitive systems". As depicted in Figure 5 Drummond et al. (2015) used a first-person shooter (FPS) game to metric players' in-game control and aggression namely, Punish Aggression and Reward Aggression, the former can be considered unnecessary aggression whereas the latter is considered indispensible to the game.

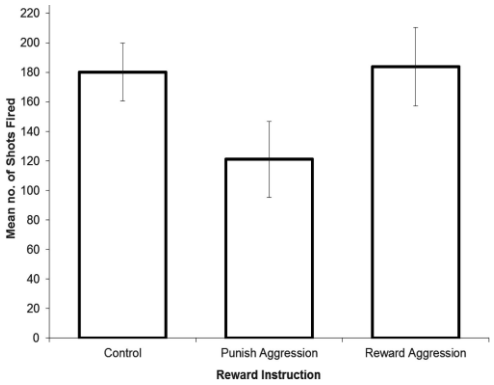


Figure 5 Results for investigating the flow of in-game aggression into post-game aggression (Drummond et al. 2015).

The results reveal that this reward-based, in-game aggression has little effect on post-game aggression, despite public perception of the contrary. This is evidently depicted in Figure 5 as the reward aggression visibly dominates the study, we can deduce that players who "experienced a reward structure encouraging in-game violence engaged in similar levels of postgame aggression to players' who experienced a punish-violence or neutral in-game reward structure" Drummond et al. (2015). Results from Drummond et al.'s (2015) study suggest that context-based game mechanics "increase or decrease postgame aggression without affecting in-game violence". The increased identification with aggressive game characters was previously linked to increase in self-activation and postgame aggression however the study by Drummond et al. (2015) had no influence on self-activation.

Hallmarks of aggressive behaviour are reflected in the research of Bateman et al., firstly when identifying the "Socialiser" typology's trusting persona, the study suggests that players fitting this typology will show anger (or *fiero*) when their trust is abused. In addition to the "Socialiser", Bateman et al. (2014) uses *fiero* to describe the ferocity used by some typologies in their research. When describing traits of the "Conqueror" it states, "They behave forcefully, channelling their anger to achieve victory and thus experience *fiero*" (Bateman et al. 2014).

3.2 Cognitive Flexibility (Visuospatial skills and Executive functions)

Research by Bavelier et al. (2014a) suggests that different video game genres affect players' cognitive functions differently. This section discusses research on the effects of violent video games and Real-Time Strategy (RTS) games on players' visuospatial skills and executive functions.

Action Video Games

A study accredited to Bavelier et al. (2014a) carried out experiments on respective players' visual attention to violent games and non-violent video games. The study appealed to the participants' "fundamental ability to select task-relevant items and filter out task-irrelevant items" (Bavelier et al. 2014a). As depicted in Figure 6, the visual attention is higher in test subjects that partook in violent video gameplay as compared to those that partook in non-violent video games. Bavelier et al. (2014a) suggest that the continued use of action video games will enhance players' sustained attention, vigilance and overall perception. Bavelier et al.'s (2014a) research examines the effects of video game training on players, observing the variations in pre-test and post-test results of players' ability to select and manage attention, memory, and planning. As depicted in Figure 6, action video game training results in players' improved executive functioning and cognitive flexibility as compared to the results of non-action video games on players' executive functioning and cognitive flexibility. A further study by Bavelier et al. (2014b) based on a visuospatial short-term memory task revealed that violent, Action Video Game Players (AVGPs) were more accurate than Non-Violent Video Game Players (NVGPs) for set sizes that were higher than 4, the general memory capacity found in these tasks. This supports earlier claims by underlining AVGPs "have better visuospatial memory performances than NVGPs" (Bavelier et al. 2014b).

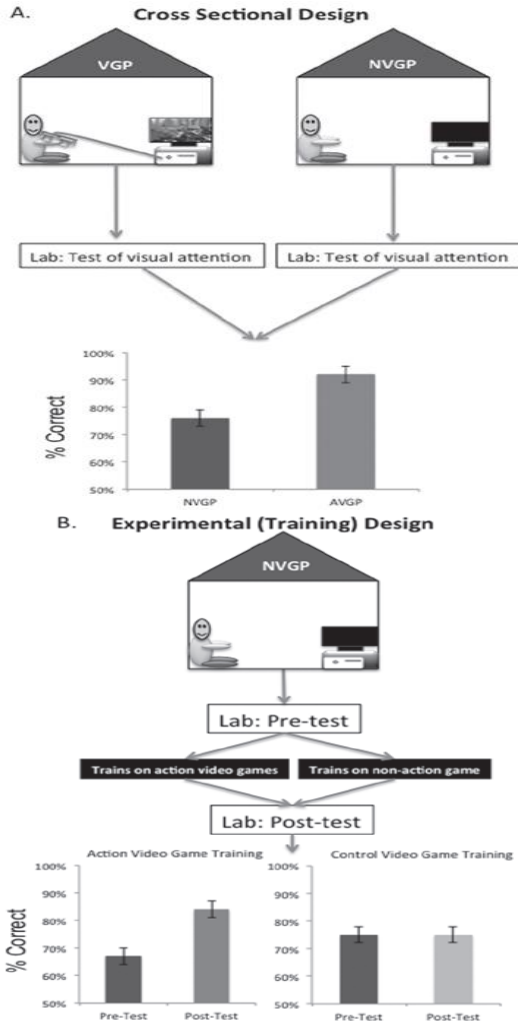


Figure 6 Test results for the effects of action video game play on perception and cognition (Bavelier et al. 2014).

Real-Time Strategy (RTS) Games

Research by Bavelier et al. (2014a) on the effects of RTS places an emphasis on players' "active maintenance of many different units, each having its own timing hierarchy and schedule". Once again video game training has been used by Bavelier, et al. (2014a) as forty (40) hours of RTS video game training "resulted in significant increases in cognitive flexibility, mainly players' improved working memory and reduced task switching cost". Fundamentally results provided by Bavelier et al. (2014a) on RTS video game training suggest "games that challenge different aspects of perception, attention, and cognition in a variety of contexts are likely to result in enhancements of these base abilities".

3.3 Social Isolation (Immersion)

Bowman et al.'s (2015) research has revealed that players' immersion in the video game world can lead to solitary behaviour in reality. The fantasy elements of video game use may help players meet needs not met in real life, thus reducing stress. In a study carried out by Book and Worth (2014), the HEXACO model was applied to immersion where it was predicted by the Openness to Experience dimension. It is therefore suggested by Book and Worth (2014) that players under this HEXACO dimension exhibited creativity and inquisitiveness, these traits are consistent with players who "fully engage in the fantasy game-world". Figure 4 depicts introversion as a dominant preference amongst Bateman et al.'s (2014) player typologies, suggesting they all have the potential to keep to themselves and avoid social interaction.

4 Discussion

The Brainhex model proposed by Bateman et al. (2014) encapsulates the cognitive motivation players have for playing video games and adequately classifies them as archetypes. The Brainhex model archetypes are adequately supported by validated psychological theories about the brain and the release of neurotransmitters namely endormorphin, dopamine, epinephrine (adrenalin) and norepinephrine. It can however be stressed that the model can be misunderstood as some participants of the study in particular may be misconstrued as a different typology, especially with the closely associated motives of some of these archetypes, namely the "Daredevil" and "Survivor". Furthermore, some participants may feel they do not belong to any of the proposed archetypes or rather fall under more than one archetype, which undermines the classifications Bateman et al. (2014) has attempted to put in place. It is for this reason Monica and Zeigler-Hill (2015) applied the HEXACO model to Bateman et al.'s (2014) proposed archetypes. Monica and Zeigler-Hill's (2015) research adds credibility to Bateman et al.'s (2014) Brainhex model as it adequately portrays the archetypes' association with the HEXACO traits, which either verified or expanded on a typology's proposed motivations. Monica and Zeigler-Hill's (2015) study can also be used to disassociate other motives from the Brainhex model.

The player incentives of challenge along with social interaction were not adequately represented in Bateman et al. (2014) or Monica and Zeigler-Hill's (2015) studies. Current research by (Cruz et al. 2015; Dickhäuser et al. 2015; de Salas and Lewis 2013; Neys et al. 2014) appropriately relate challenge to

achievements and the Social Determination Theory. An unbiased sample size was used by Collins and Cox (2014) to identify that 75% of participants played games for a challenge and de Salas and Lewis (2013) supports this claim with research introducing an achievement-driven player typology, "The Completionist". (Bowman et al. 2015; Herodotou et al. 2014; Ferguson and Olson 2013) use their research to discuss the social interaction motive derived from the Brainhex model's "Socialiser" archetype. An adequate sample size was used to suggest 67% of participants played WoW for social reasons. Although WoW is a Massive Multiplayer Online Game, research by Cruz et al. (2015) depicts this genre as the fourth most played amongst video game players.

Aggressive behaviour was investigated with the preconception that in-game aggression leads to post-game aggression or self-activation of aggression. The study by Drummond et al. (2015) rightly used a violent, First Person Shooter (FPS) game to discredit the claim that in-game aggression flows into postgame aggression, instead context-based game mechanics were found to have an effect on post-game aggression although they have no effect on in-game aggression. Drummond et al. (2015) paints the picture that violent games have little effect on players' post game aggression and self-activation of aggression. However, hallmarks of the Brainhex model were used to describe what raises aggressive behaviour in respective archetypes. Research by (Bavelier et al. 2014a; Bavelier et al. 2014b) has shed violent video games in a positive light as their experiments used AVGPs and NAVGPs to compare effects of violent video games on different types of players. Also the effects of non-violent video games were compared to those of violent video games. Bavelier et al.'s (2014a) experiments support the claims that violent video games improve cognitive flexibility and executive functions, along with the improvement of problem solving skills, estimation and strategy development by Real Time Strategy Games. Bavelier et al.'s (2014b) research used a reliable memory game to prove the superior memory abilities in action video game players.

Immersion in the gaming fantasy world was investigated by (Book and Worth 2014; Bowman et al. 2015) using the HEXACO model and this effect was validated by the Openness to Experience trait encouraging players to lose themselves within the fantasy worlds provided by games.

5 Recommendations

It would be prudent to utilise both Bateman et al.'s (2014) Brainhex model and the HEXACO model when investigating any motivations players may have as demonstrated by Monica and Zeigler-Hill (2015) in section 2.2 of this paper.

6 Conclusions

This paper has surveyed players' motives (Brainhex model, HEXACO model and social interaction) and the effects games have on players (Aggressive behaviour, cognitive flexibility and immersion) in an attempt to identify and evaluate the cognitive incentive and impact players incur. This area of human-computer interaction (HCI) is still in its infancy albeit the literature is vast but not all that reliable. The research by (Bavelier et al. 2014a; Bavelier et al. 2014b) can be considered ground breaking, as the misconception that violent video games affect players negatively was buried with astute evidence.

References

Bateman, C., Mandryk, R.L. and Nacke, L.E., 2014, 'Brainhex: A Neurobiological Gamer Typology Survey.' *Entertainment Computing*, 5(1):55-62, January.

Bavelier, D., Eichenbaum, A. and Green, C.S., 2014a, 'Video Games: Play That Can Do Serious Good.' *American Journal of Play*, 7(1):50-72, Fall.

Bavelier, D., Green, C.S. and, McDermott, A.F., 2014b, 'Memory abilities in action video game players.' *Computers in Human Behavior*, 34:69-78, May.

Book, A.S. and Worth, N.C., 2014, 'Personality and behavior in a massively multiplayer online role-playing.' *Computers in Human Behavior*, 38:322-330, September.

Bowman, N.D., Cohen, E. and Kowert, R., 2015, 'When the ball stops, the fun stops too: The impact of social inclusion on video game enjoyment.' *Computers in Human Behavior*, 53:131-139, December.

Caroux, L., Isbister, K., Le Bigot, L. and Vibert, N., 2015, 'Player–video game interaction: A systematic review of current concepts.' *Computers in Human Behavior*, 48:366-381, July.

Cavedon, L., Karpinskyj, S. and Zambetta, F., 2014, 'Video game personalisation techniques: A

comprehensive survey.' *Entertainment Computing*, 5(4):211-218, December.

Coary, S., Hou, J., Kahn, A.S., Lu, L., Meng, J., Osborn, J., Ratan, R.A., Shen, C. and Williams, D., 2015, 'The Trojan Player Typology: A cross-genre, cross-cultural, behaviorally validated scale of video game play motivations.' *Computers in Human Behavior*, 49:354-361, August.

Collins, E. and Cox, A.L., 2014, 'Switch on to games: Can digital games aid post-work recovery?' *International Journal of Human-Computer Studies*, 72(8-9):654-662, August-September.

Cruz, C., Fox, J. and Hanus, M.D., 2015, 'The need to achieve: Players' perceptions and uses of extrinsic meta-game reward systems for video game consoles.' *Computers in Human Behavior*, pages 62-75, September.

de Salas, K. and Lewis, I., 2013, 'Identifying types of Achievements.' *The 18th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games (CGAMES)*, pages 23 - 30.

Dickhäuser, O., Dinger, F.C., Hilbig, B.E., Müller, E., Steinmayr, R. and Wirthwein, L., 2015, 'From basic personality to motivation: Relating the HEXACO factors to achievement goals.' *Learning and Individual Differences*, 40:1-8, May.

Drummond, A., Nova, N. and Sauer, J., 2015, 'Violent video games: The effects of narrative context and reward structure on in-game and postgame aggression.' *Journal Of Experimental Psychology: Applied*, 21(3):205-214.

Ferguson, C.J. and Olson, C.K., 2013, 'Friends, fun, frustration and fantasy: Child motivations for video game play.' *Motivation & Emotion*, 37(1):154–164, March.

Herodotou, C., Kambouri, M. and Winters, N., 2014, 'Dispelling the myth of the socio-emotionally dissatisfied gamer.' *Computers in Human Behavior*, 32:23-31, March.

Monica, S. and Zeigler-Hill, V., 2015, 'The HEXACO model of personality and video game preferences.' *Entertainment Computing*, 11:21-26, November.

Neys, J.L.D., Jansz, J. and Tan, E.S.H., 2014, 'Exploring persistence in gaming: The role of self-determination and social identity.' *Computers in Human Behavior*, 37:196-209, August.

Evaluation of Computing Research Aimed at Improving Virtualization Implementation in the Cloud

Keletso King Mooketsane

Abstract

Virtualization in the cloud has grown in implementation over the past few years. Several techniques have been developed which aim to improve how data is stored and some security aspects are also considered. This research paper aims to analyze, compare and evaluate some of the techniques that have been implemented to improve how virtualization is used in the cloud. It mainly focuses on the Xen-hypervisor virtualization technique which uses the Open Nebula virtual machines management tool, the Graphics processing units (GPUs) virtualization using PCI pass-through mechanism and the Fine grained parallelism in graph traversal algorithms via lock virtualization on multi-core architecture. This paper evaluated the techniques mentioned, compared them and provided a recommendation on which technique is most suitable to be used in the real world.

1 Introduction

The rise in cloud computing has lead most modern organizations to implement cost-effective virtual environments into their operations. Goettelmann et.al. (2014) explained that to manage the security concerns in Information Communication Systems (I.C.T) is vital to guarantee the security in the organization as well as handling the costs. The lack of skills, expertise and experience often compromises the data that is stored in the cloud to a lot of security risks that may expose these organizations to unnecessary attacks, Ashford (2012). These may include virtual firewalls improperly configured, hypervisors/Virtual machine monitor (VMM) configured improperly, Data leakage through offline images etc. Pearce (2013) stated that, one major concern is with the lack of knowledge of how virtual environments work, most organizations are too focused on performance and costs resulting in security being overlooked or tagged only at the end.

Roy et.al. (2015) explained another security threat is when a virtual machine crashes, it is much harder to recover than when a workstation crashes because the crashed virtual machine becomes a corrupt binary file and unrecoverable by normal forensics standards. Security is a major contributor of VM crashes, such as compromising the hypervisor, using a remote cloud client and the threats from malicious insiders. Choi et.al. (2013) and the information security forum (ISF) identified key activities that should be carried out for good practice to secure virtual environments, they include establishing a policy when using virtual servers such as limiting the number of virtual servers that can run on one physical server and to control the number of business applications that can run on one server.

Yang et.al. (2013) proposed a system that claimed that it improved from using a single server to using more than one machine, each able to offer local computation and storage capabilities. Yan et.al. (2014) proposed a novel fine-grained lock mechanism and lock virtualization lock that claimed to map a large logical space to a small fixed physical lock space that can be located in the cache during runtime. Yang et.al. (2013) proposed a programming model that claimed to offer barrier synchronization, shared memory and a hierarchy of thread blocks. These solutions above all claim to address issues of security and improve network performance by insuring that data is protected, available when needed and it cannot be accessed by unauthorized users.

Therefore, the goal of this paper is to take a critical look at current issues that are present in the most recent research papers focusing on the methods that have been discussed above and evaluate the security threats that are faced in virtual cloud systems. A critical evaluation of the technical solutions will be done to determine the most suitable method that will offer a better and more secure cloud storage environment.

2 The Xen-Hypervisor Virtualization technique using the Open Nebula virtual machines management tool

Yang et.al. (2013) carried out research on the Xen-Hypervisor virtualization technology and the Open Nebula virtual machines management tool, they were designed to improve from using a single server to using

more than one machine, each offering local computation and storage capabilities. The Hadoop cluster/virtualization fault tolerance (VFT) has multiple worker nodes and a single master node that has a Data node, Task Tracker, Name node and Job Tracker, it uses the information of different rack names when duplicating data and attempts to keep different copies of the data on the different nodes. Its goal is to reduce the impact of a power outage of the rack or a switch off failure, i.e. when these events occur their data may be readable.

The researchers considered a practical problem, which was the single-point-of-failure issue which happens often in virtualized systems. Their experimental results confirmed that the downtime interval can be shortened greatly even if there is a failure. Their result showed that virtual machine tolerance is not only useful for Hadoop applications, but also for areas in cluster-based systems.

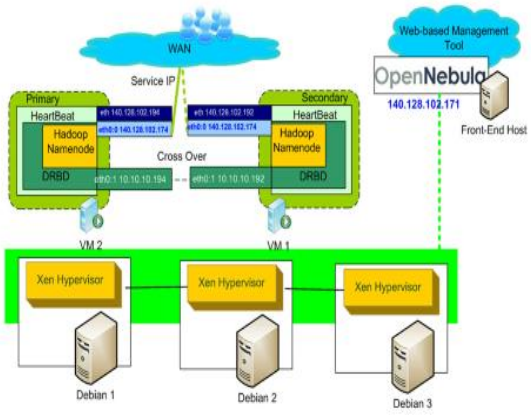


Figure 1: This diagram depicts an overview of the Open Nebula system. A cluster system was built with open Nebula to provide users a web interface to manage virtual and physical machines, Yang et.al. (2013).

The diagram's cluster has four (4) computers with similar specifications, hardware equipped with Intel i7 CPU 2.8 GHz, and four gigabytes of memory. 500 gigabytes disk, a Debian operating system and a gigabit switch to connect to the network.

The Hadoop Distributed File system was able to address the challenges that are faced when developing a distributed application such as hardware failures, if more than one piece of hardware is used, there is more chance of failure. The system was able to replicate redundant copies of data so that in the event of a system failure, a backup copy of the data will be available.

They claimed that they were able to offer local computation and storage capabilities, however, their system did not support automatic recovery for Name node failure which was a well-known and recognized single point of failure in their system Yang et.al. (2013). If their Name node machine failed then manual intervention would be necessary, it also did not support automatic restart and failover to another machine. They attempted to resolve the single-point-of-failure issues that surrounded the master node process such as the Job Tracker and the Name node. Their solution, called the virtualization fault tolerance (VFT) made use of Distributed Replicated Block Device (DRB) from LINBIT and the Heartbeat from the Linux- High Availability (HA) project, Yang et.al. (2013).They claimed that combining these projects provided them with a more reliable and highly available solution which would address their critical limitations. They used virtualization as a solution not only to improve their flexibility, but also to consolidate the workloads and to enhance the utilization of the server, by adapting dynamically to the client's demands, such as deploying new virtual nodes when demands increase and powering off when demands were low.

Based on what the authors presented, there still remains certain gaps about if their proposed system could support automatic recovery, which was a well-known point of failure in the Hadoop system. There were no experiments in their paper showing no results of their system to prove that they were able to provide information backup and addressing the hardware failures. They only showed the algorithms they presented and concluded by drawing attention to the fact that their system was able to manage virtual machines more effectively.

A wide variety of tests and results were produced to emphasize how their solution was able to improve flexibility as well as consolidate the workloads and enhance how the server was used, but it was also apparent that the aspect of security was not thoroughly investigated or even attempted. They did however present their results and evidence well to justify the claims that they made. They did not support their results against a similar system where a comparison could be made. It is also apparent that they tested on just Linux systems, this was not enough as there are other systems which could have been used to give a much better result.

To conclude, more experiments, research and more tests will need to be carried out on different virtualization systems to analyze and compare those systems to reach a more concrete solution. Which was done by Pek et.al. (2013) which showed their experiments and

results where they used more than one system to compare allowing the authors to reach a more representative conclusion.

3 Graphics processing units (GPUs) virtualization using Pass-through mechanism (PCI)

Yang et.al. (2013) proposed a scalable parallel programing model which was used to code highly parallel applications. Compute Unified Device Architecture (CUDA) from NVIDIA Graphics Processing Unit (GPU) provided a number of key abstractions namely barrier synchronization, shared memory and a hierarchy of thread blocks. GPUs have an important position in the future of cloud computing because applications requiring high computing abilities require high powered CPUs just as GPUs.

The Pass-through technology (PCI) uses a virtual machine added to a virtual environment which allows it to use the NVIDIA graphics card as well as the CUDA high performance computation. This enabled the virtual machine to have the virtual CPU and the GPU as well. Experiments were carried out on how the virtual machine performed which was expected to improve substantially. Yang et.al. (2013) stated that measuring the differences between the virtual machines as well as the physical machines using CUDA was carried out, investigating how the virtual machine would authenticate the CPU member using CUDA. A comparison of the performance of CUDA using different virtualization hypervisor environments, and only using the PCI pass-through once. The data from the experiments showed the authors which condition was more efficient using CUDA.

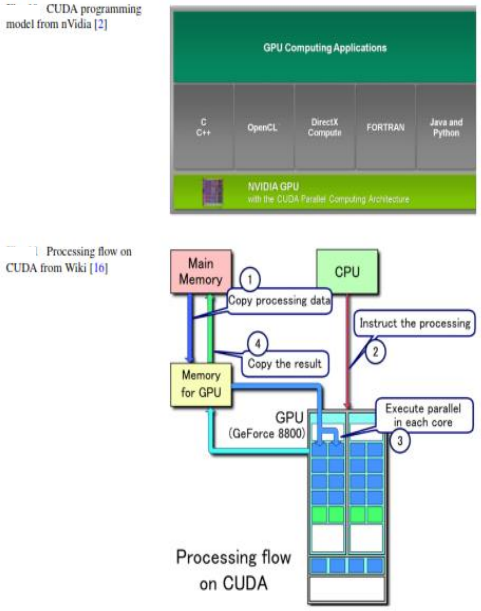


Figure 2: The diagram shows how the GPU virtualization is used with PCI pass-through mechanism, Yang et.al. (2013).

The authors claimed that they were able to conserve resources at the same time having similar performance of real physical machines, however their experiments did not consider the security aspects specifically on the key issues for secure virtualization such as the lack of visibility in the network traffic, performance sapping security overhead and if the security perimeter is breached, Zhu et.al. (2015). These are the major issues in making sure that information stored in the servers are secure, they also didn't address any security concerns that were present, they did however provide some algorithms that could be used for providing some sort of security, these algorithms did not provide any sort of detail on how they could work and there were no experiments or results, they only emphasized on how they would provide security but their claims were not backed up by providing results from experiments that could provide proof that their algorithms could actually improve on that particular aspect.

Experimental work for testing the performance of the PCI pass-through and the Graphic processing units and the results were provided, however, it is not confirmed how the systems performed as they were just providing general performance of their system and didn't go into much detail. Therefore, more thorough experiments will need to be carried out to prove the validity of their claims allowing for a fairer conclusion to be reached.

In conclusion, GPU's execution is identical to the virtual machines that use the PCI pass-through, it doesn't affect the number of CPUs that are used in the virtual machines and the GPUs still maintains the same behavior. Using the PCI pass-through when adding computing with GPU accelerators in virtual machines can save computing resources and have identical high operation than can be found in real machines.

4 Fine-grained parallelism in graph traversal algorithms via lock virtualization on multi-core architecture

Yan et.al. (2014) carried out research on the fine-grained data synchronization which was critical in order to examine the substantial fine grained parallelism in graph traversal.

The rise of graph analytics of large irregular graphs such as social networks means that it often suffers from large memory costs and poor locality due to the massive scale vertex and inherent random vertex access. Yan et.al. (2014) proposed a novel fine- grained lock mechanism and lock virtualization lock.
Their goal was to map the large logical space into a small physical lock space which can be located in the cache during runtime. The virtualization mechanism can effectively reduce the extra lock incurred memory cost, it can achieve this while preserving the high portability for the legacy cores by catering P-Threads like application programming interface.

Further analysis revealed that from random access pattern, the lock conflict rate was longer related to the size of the vertex that was set, only the numbers of the parallel threads and the physical locks were retained, this meant that the v-lock was independent from graph topologies. The v-lock's performance was evaluated in four (4) graph traversal algorithms namely Breadth – First Search (BFS), Single Source Shortest Path (SSSP), Connected Components (CC) and Page-Rank.

Experiments were carried out on the Intel Xeon ES eight-core processor, Yan et.al. (2014) stated that the processor revealed that, compared to the P-threads fine-grained locks, the v-lock required the lock's cache significantly and achieved 4-20% performance improvement.

Figure 3 shows the lock virtualization in parallel graph traversal algorithms, Yan et.al. (2014) where the x-axis depicts the number of threads (left).Performance and normalized to one thread with Coarse-lock. The Higher one is better. (Right) L3 cache (LLC) represents miss rate and the lock conflict rate.

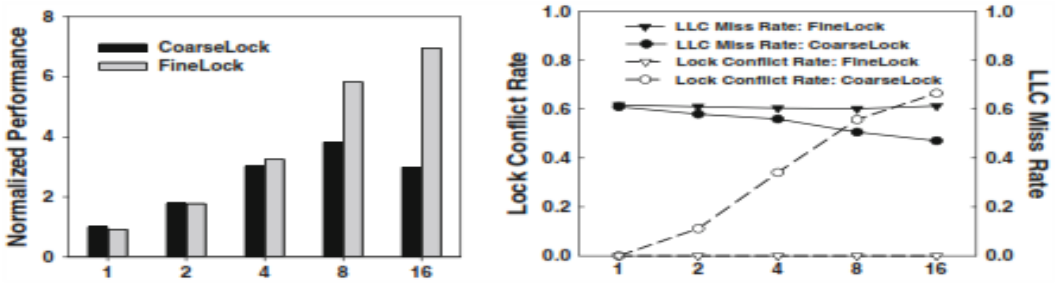


Figure 3: Yan et.al. (2014)

To conclude, a solution of the fine-grained locks i.e. v-lock was presented which its idea was lock virtualization. With the biggest size of the virtual lock space and the random access to locks comparing to the traditional fine-grained lock, the v-lock only needed a small amount of memory and achieved higher performance on the multi-core platforms.

The authors presented their algorithm well and described how they would carry out their experiments. They evaluated their V-lock with four (4) graph traversal algorithms namely Breadth-first search (BFS), Single source shortest Path (SSSP), Connected Components and the Page-rank algorithms. They presented their results well on how their algorithm performed by testing it with other author's algorithms. They showed their results in an experimental section which showed how their v-lock outperformed other fine-grained lock methods that were used, Yan et.al. (2014) also explained that for many applications with frequent lock requests such as Page-Rank, the v-lock showed more than 20% improvement on its performance. They also explained that their v-lock provided the same programming style as the lock methods in the P-threads. With their v-lock, there is no need to change the original structure of the codes. They also concluded well on describing that their v-lock was decoupled from

any specific graph vertex set or even the memory objects and it could easily be implemented into the hardware as an extension of multi-core processors.

For a more concrete conclusion about their algorithm to be made, more experiments will need to be carried out which should focus more on the security aspect of their algorithm. It is clear that their algorithm is a better improvement on the current algorithms used such as providing better performance and memory, more experiments will need to be conducted on how their algorithm can offer better security.

5 Conclusions

In this research paper, the most up-to-date research papers on the different virtualization techniques have been thoroughly analyzed and evaluated. The performance and some security aspects were also taken into consideration.

Research done on the Xen-hypervisor virtualization technique which used the Open-Nebula virtualization machine management tool done by Yang et.al. (2013) showed that, high availability could be achieved using the Hadoop Name-Node Active-Standby Architecture, allowing the service to be failed over if the primary node failed. In their analysis, their experiments and simulations showed that their management tool resulted in generating positive results, however these would still need to be compared in the real world to see how their system would function in an actual real-world environment. A potential of bias could result from only relying on results from experiments because everything is done in a controlled environment. Further analysis and testing will need to be carried out to reach a more representative conclusion.

More research was also done on the Graphics processing units' virtualization which used the pass-through mechanism by Yang et.al. (2013) which showed that the GPU's performance was identical to the native machine and the virtual machines which used PCI-pass through. The authors reached a conclusion that using the PCI-pass through to add computing with the GPU accelerators in the virtual machines would be able to conserve resources and have the same high performances that could be found in physical machines They also planned to test more GPU boards in PCI pass-through and to add GPU hot-plugs into the virtual machines in the future, Yang et.al. (2013). However, it is difficult to come up with a representative conclusion as their mechanism was not tested against other schemes to be able to compare them. Therefore, more experiments and tests will need to be carried out and the results compared with other mechanisms proposed by other authors.

Research on the Fine-grained parallelism in traversal algorithms via lock virtualization on multi-core architecture by Yan, et.al. (2014) showed that the author's solution of the fine-grained locks i.e. the V-lock only needed a small amount of memory and was able to achieve higher performance on the multicore platforms. This showed a lot of potential on how it could be implemented in the real world because their algorithm was compared/tested with other algorithms in its field. The results showed that their algorithm outperformed other lock methods that were used. However, more experiments will need to be carried out which should focus more on the security aspect of their algorithm.

The schemes/algorithms presented still need more work before they are implemented in the real world, but the Fine-grained parallelism method showed the most potential in that the authors presented their results against other related methods allowing for a more unbiased and accurate conclusion to be reached.

References

Ashford Warwick, 2012 'Seeking Nirvana: Virtualization without security risk'. *IEEE International Conference on Services Computing*, 5, Pages 19-22.

Choi Chang, Choi Junho, Kim Pankoo, 2013 'Ontology-based access control model for security policy reasoning in cloud computing'. *Department of Computer Engineering Journal*, Vol. 55, Pages 711-722.

Elio Goettelmann, Karim Dahman, Benjamin Gateau, Eric Dubois, Claude Godart, 2014 'A Security Risk Assessment Model for Business Process Deployment in the Cloud'. IEEE International Conference on Services Computing, 8, Pages 307-314.

Pearce Michael, Zeadally Sherali, Hunt Ray, 2013 'Virtualization: Issues, Security Threats and Solutions'. *ACM Computing Surveys*, Vol. 45, No.2, Article 17.

Pek Gabor, Buttyan Levente, Bencsath Boldizsar, 2013 'A Survey of Security Issues in Hardware Virtualization'. *ACM Comput.Surv*.45, 3, Article 40, 34 Pages.

Roy Arpan, Sarkar Santonu, Ganesan Rajeshwari, Goel Geetica, 2015 ‘secure the cloud: From the perspective of a service oriented organization’. *ACM Computing Surveys*, Vol.47, No.3 Article 41.

Yan Jie, Tan Guangming, Sun Ninghui, 2014 ‘Exploiting fine-grained parallelism in graph traversal algorithms via lock virtualization on multi-core architecture’. *Springer Science Business Media Journal* Vol. 69, Pages 1462-1490.

Yang Chao-Tung, Liu Jung-Chun, Hsu Ching-Hsien, Chou Wei-Li, 2013 ‘On improvement of cloud virtual machine availability with virtualization fault tolerance mechanism’. *Department of Computer Science Journal*, Vol. 69, Pages 1103-1122.

Zhu Qiang, Wang Hui-Qiang, Feng Guang-Sheng, Lv Hong-Wu, Wang Zhen-Dong, Wen Xiu-Xiu, Jiang Wei, 2015 ‘A Hybrid Reliable Heuristic Mapping Method Based on Survivable Virtual Networks for Network Virtualization’ *Discrete Dynamics in Nature and society*, vol. 2015, Article ID 316801, Pages 8.

A Critical Evaluation of the Technology Used In Robotic Assisted Surgeries

Botshelo Keletso Mosekiemang

Abstract

Robotics or Artificial Intelligence in general have become the centre of attention in the field of medicine and with the advent of technology, there is continuous and persistent birth of advanced and relevant technology aimed at aiding in the various robotic assisted surgeries. This paper describes a critical evaluation of the currently used technologies in robotic assisted surgeries, for instance Electromagnetic Navigational Bronchoscopy and Robotic-Assisted Surgery; an evaluation is made with regards to the experiments and results of the proposed prototypes as well as the claims made. There are conclusions reached with regards to the comparisons made amongst the highlighted technologies.

1. Introduction

As one of the most critical parts embodied in the anatomy of Medicine, Surgery is a very complex, demanding field of Science that requires immense care; precision and accuracy. It is therefore as a result important that the right and necessary measures are put in place to allow for this and hence the incorporation and use of intelligent agents like robots or even certain methods in performing the various surgical procedures.

According to Christie (2014) the integration of robotic surgery and electromagnetic navigational bronchoscopy allows for less infiltrative surgical procedures and also brings about benefits of early diagnosis as well as brief check-ins at the hospital with regards to the patients concerned.

Elsamnah et. al. (2014) stresses the issue of dealing with the problem of Line of Sight in image guided surgery by employing a multi-stereo camera system as opposed to a single stereo camera. The incorporation of a multi-stereo camera is said to be aimed at enhancing the accuracy of the image guided surgery. Hu et. al. (2013) mentions the use of a video see-through technique which integrates real and virtual objects and allows surgeons to actually see components of a physical environment under view as per the produced virtual objects with reference of course to image guided surgery. Jeevan et. al. (2014) describes a 3D pre-operative visualization for atrial transseptal puncture, which is actually another form of image guided surgery. Jeevan et. al. (2014) mentions how challenging it is to gain a passage to the left atrium and that the passage to the left atrium via the systemic venous system is actually the most advocated for as opposed to the highly retrograde arterial route. More cautious implementation of greater catheters and devices is possible as such. Jeevan et. al. (2014) as such intends to avail a proficient method for the purposes of target localization pre-operatively as well as 3D visualization with consideration to the core of the atrial transseptal puncture surgery, the "catheter tip".

Salgueiro et. al. (2015) proposes an assistive tool for breast reduction surgery, which is aimed at aiding doctors with capturing the original view of a particular patient's breasts and also the modelling of the desired breast outlook so as to avoid any kinds of risks before they can happen. The assistive tool, which is an interactive web-based tool (Salgueiro et. al. 2015) serves as a repository that makes available the much needed information before surgery and also makes available post information before the surgery is even completed.

Corenthy et. al. (2014) proposes haptically driven procedures for thorough reconstruction of dendritic spines and it is mentioned that these procedures allow for image processing stage ahead of the automatic segmentation stage, and thereafter avail the final outlook of the dendritic spines. The issue here is that dendritic spines are thin and hence a need for careful planning before any form of surgery can be undertaken. Brooks (2015) however, presents an evaluation of a number of robotic-assisted technologies in executing lobectomies with regards to early-stage lung cancer patients. Emphases are given in that robotic-assisted lobectomies are safe for patients with stage 1A or 1B lung cancer, and that some conventional methods are still needed in making these procedures more efficient (Brooks 2015). Silva et. al. (2014) on the other hand talks about availing an interactive tool, whose design is in reference to a latter existing left ventricle segmentation and matches lesions identified by the clinician to a myocardial segment. It is claimed that this tool renders

some feedback about the myocardial segments identified with lesions and their severity.

This paper will focus on critically looking into and evaluating a few of the current technologies manipulated in various kinds of robotic assisted surgeries. A discussion pertaining to their implementation will be made and aspects such as of experiments and results will be stressed. Ultimately, a conclusion will be made with regards to those currently used technologies.

2. Currently Used Technology in Robotic Assisted Surgeries

This section will stress at least three of the many other technologies used in Robotic Assisted Surgeries. The proposed solutions, their experiments and results will be critically looked into and evaluated.

2.1. Electromagnetic Navigational Bronchoscopy and Robotic-Assisted Thoracic Surgery

Christie (2014) stresses the manipulation of electromagnetic bronchoscopy as well as robotics, which enable surgeons to treat lung lesions on a single surgical procedure. ENB and robotics are actually a cherry on top of the latter, thoractomy and video-assisted thorascopic surgical (VATS) techniques for thoracic surgery. Christie (2014) mentions that the mandate of robotic surgery is to help provide minimally invasive procedure which will in turn lead to patients having speedy recoveries and enduring minimal pain during the course of the surgical procedures.

ENB consists of three phases, being the planning; registration and navigation phases, and it is claimed that surgeons can be able to diagnose and undertake a biopsy with reference to peripheral lung lesions which are rather complicated to arrive at via a conventional bronchoscope (Christie 2014). Furthermore, claims are made that techniques such as image-guided, steerable catheters in collaboration with navigational software are manipulated so as to produce a 3D structure resembling the patient's bronchial anatomy and figure 1 below depicts this.

Figure 1 Display of CT cross-sections; 3D views and guidance instructions for bronchoscope and catheter steering (Christie 2014).

Christie (2014) emphasizes the importance of both ENB and Robotics in thoracic surgeries by presenting a patient case report. This report is in relevance to a patient who suffers from abdominal pain, and whom is actually a smoker (Christie 2014). On one of the last stages of the surgery it is claimed that when the patient is at the cardiovascular ICU, he is ordered to suppress his own pain with the use of a patient controlled analgesia pump. This in my view seems to be insensitive and inconsiderate of the surgeons because usually surgeries make one weak to a point that they cannot handle much on their own. Therefore as such, there is need to employ other measures in addressing this issue. But despite this though, claims are made with regards to ENB in that it helps in avoiding the need to transfer patients to and from the radiology department for needle localization since the ENB procedure itself incorporates this function and this saves a lot of time. A nursing care plan for the patient undergoing surgery is provided and with the given parameters, one may use this as a guide to validate the results of ENB and Robotic-Assisted thoracic surgery. Even though so, given the sample size of which the surgery was carried out on, it is impossible to safely and firmly say that ENB and robotics are effective considering the fact that the surgery was performed on only one patient and also the nursing care plan availed was for that one patient. See figure 2 below for a sample nursing care plan.

TABLE 1. Nursing Care Plan for a Patient Undergoing Electromagnetic Navigational Bronchoscopy and Robotic-Assisted Thoracic Surgery

Diagnosis	Nursing interventions	Interim outcome statement	Outcome statement
Risk of imbalanced body temperature	■ Assesses risk of normothermia regulation. ■ Assesses risk of inadvertent hypothermia. ■ Assesses risk of inadvertent hyperthermia. ■ Identifies physiological status. ■ Implements thermoregulation measures. ■ Monitors body temperature. ■ Monitors physiological parameters. ■ Evaluates response to thermoregulation measures.	■ The patient's temperature is greater than 36° C (96.8° F) at time of discharge from the operating or procedure room.	The patient is at or returning to normothermia at the conclusion of the immediate postoperative period.
Risk of perioperative	■ Assesses baseline skin	■ The patient has full return	The patient is free from signs

Figure 2-Table of a sample Nursing Care Plan (Christie 2014)

2.2. Multi-Stereo Camera System To Enhance The Position Accuracy Of Image-Guided Surgery Markers

Elsamnah et.al. (2014) mentioned availing coloured reference markers as assistive tools in tracking systems for Image-guided surgery systems, with reference to a multi-stereo camera system. The coloured reference markers are used in collaboration with the multi-stereo camera system to aid in making image-guided surgery effective, and mostly accurate. It is mentioned that none of the latter similar, proposed methods, made use of coloured markers and hence the reason why the line of sight problem still prevailed and also the reason why coloured markers were incorporated in image-guided surgery to better it and make it worthwhile to the medical fraternity.

Elsamnah et. al. (2014) carried out an experiment to demonstrate the coloured markers tracking method with reference to a multi-stereo camera. This experiment was executed in three stages being: the detection of the circular objects in the image and recognition of the colour of each marker; determination of coloured marker in space (x, y, z) and lastly, using the multi-stereo camera. In the detection of circular objects in an image, Elsamnah et. al. (2014) emphasizes how a captured image needs to be enhanced in order for one to appreciate the presence of circular objects in the image. Techniques aimed at enhancing the captured image such as of resizing and unsharp mask are stressed and Figure 3 illustrates the outcomes and presentation of these techniques.

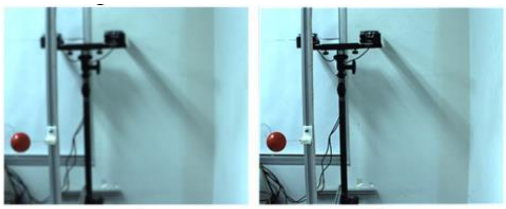


Figure 3 (a) Before sharpening the image (b) After sharpening the image using unsharp mask (Elsamnah et. al. 2014)

Elsamnah et. al. (2014) elaborates that prior to enhancing the captured images, circular objects become easy to detect and ways in detecting these circular objects are stressed and they include Hough Transform based formulated in Probablistic Pairwise Voting or Ellipse Growing. The circular objects are checked for some coordinates which are mainly in reference to their centres and there is need after all to determine the colour of these objects (Elsamnah et. al. 2014). Afterwards, the coloured markers are said to be determined in space x, y, z axis by performing techniques such as motion parallax (Elsamnah et. al. 2014). This part of the experiment demonstrates the functionality of the stereo vision and also calculations of the projection geometry for two camera systems.
In using a multi-stereo camera, it is highlighted that its mandate is to get rid of LOS and bring about accurate results of the object position (Elsamnah et. al. 2014). IDs cameras UI-1240LE-C were manipulated in this context and the setup was as on figure 4 below.

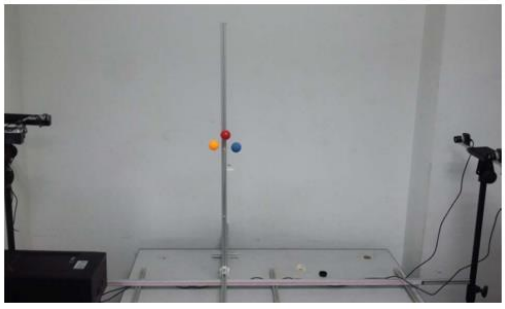


Figure 4 Display of IDs cameras UI-1240LE-C (Elsamnah et. al. 2014)

The system model was availed, showing the relevant details of operation of the multi-stereo camera based IGS system, see figure 5 below.

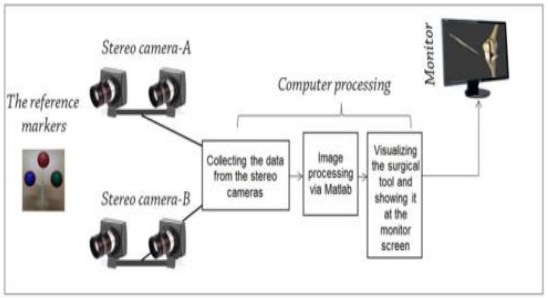


Figure 5 System model structure of the multi-stereo camera based on IGS system (Elsamnah et. al. 2014).

The experiment results attained, especially those with regards to 3D root mean square error (RMSE) showed that the multi-stereo camera was better of at giving the best accuracy as compared to the latter stereo camera. Two tables were presented showing the results of the RMSE of the markers in space at distances of 1000mm to 1100mm in two different platforms being that of the stereo camera and the multi-stereo camera.

From studying these tables, a lot of inconsistency in the values attained prevailed and questions arose as to why that is the case. It seems as if the experiments were not carried out properly and precisely, hence the appearance of great variations in numbers here and there in the tables. These numbers are obviously an influencing factor as per the RMSE of the camera system used and as such if the pattern of the numbers given is inconsistent as shown, then more questions may arise as to the truth and validity of the final RMSE reached. But despite all of that, the claims that the multi-stereo camera is more accurate than a stereo camera are valid since the RMSE of the multi-stereo camera is 2.88mm as opposed to the 4.75mm of the stereo-camera. See figure 6 and 7.

Distance (mm)	3D error (mm) Exp. 1	3D error (mm) Exp. 2	3D error (mm) Exp. 3	3D error (mm) Exp. 4
1000	1	-2	0	5
1010	9	-3	3	4
1020	5	6	9	6
1030	6	4	3	3
1040	4	0	-5	7
1050	7	-1	2	4
1060	2	9	-3	0
1070	-2	10	2	12
1080	1	0	5	3
1090	0	-1	0	0
1100	1	0	7	-4
RMS	4.89	5.97	4.42	5.52
RMS Total	4.75			

Figure 6-Table showing results of RMS of a stereo camera (Elsamnah et. al. 2014)

Distance (mm)	3D error (mm) Exp. 1	3D error (mm) Exp. 2	3D error (mm) Exp. 3	3D error (mm) Exp. 4
1000	1	-2	0	3
1010	2	-3	3	2
1020	2	1	9	6
1030	3	3	4	3
1040	3	0	-5	1
1050	2	-1	2	4
1060	1	3	-2	0
1070	-1	4	2	6
1080	-1	1	-1	3
1090	1	-2	0	0
1100	2	0	6	-2
RMS	1.88	2.22	4.05	3.36
RMS Total	2.88			

Figure 7-Table showing results of RMS of a multi-stereo camera (Elsamnah et. al. 2014)

This experiment and the results given are substantial enough to allow for other people to repeat it and validate its claims, but the prevalence of inconsistency in the values attained in the original experiments needs to be corrected by simply repeating these experiments once more and ensuring that there is little less difference in the values, and therefore as such eliminate the element of biasness with regards to people studying and evaluating those experiment results.

2.3. A Convenient Method Of Video See-Through Augmented Reality Based On Image-Guided Surgery System

Hu et. al. (2013) presents a prototype of an enhanced AR application based on an already existing IGS system better known as the Excelim-04 IGS system. Below is the frame of the proposed AR system on figure 8.

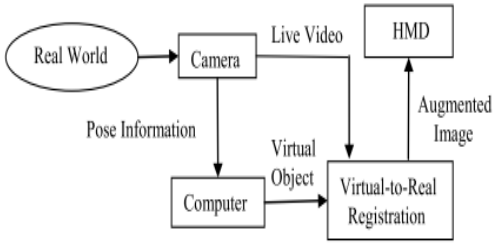


Figure 8-Frame of the proposed AR system (Hu et. Al. 2013)

Hu et. al. (2013) emphasizes that the AR system bears two cameras attached to the head-mounted display

(HMD) and that they take live video of the real world and transfer it to the computer. It is mentioned that further transfer of the video, which is the augmented video in this case, is made via video see-through HMD and is presented on some micro screens. It is claimed that for purposes of easy calibration with regards to video see-through, the eMagin Z800 3D Visor HMD is manipulated. Having two video cameras assembled on the HMD is said to allow for the birth of the much needed video camera system for purposes of augmented reality (Hu et. al. 2013).

The methods stressed in carrying out the experiment on AR include real scene calibration with regards to a skull exemplar. There are emphases on model-to-patient registration where point-pair registration function is manipulated in computing the conversion from the virtual model system to the marker system (Hu et. al. 2013). In implementing the AR system, it is stressed that once the application has began operation, inevitably the link between the marker and the real object will be defined (Hu et. al. 2013). It is mentioned that during surgery, the tracking of the camera movement on the HMD is appreciated via the use of a computer vision algorithm and that the display of the augmented medical scenes are rendered via the HMD micro screens in real time (Hu et. al. 2013).

Experiments on the proposed AR system were carried out on a skull exemplar, and images produced by the system regarding the landmarks selected from the virtual model of the skull are shown below on figure 9.

Figure 9 Illustration of the images rendered which are of the selected skull landmarks from the virtual model (Hu et. al. 2013).

Given these images though, it is likely possible, especially for someone who lacks the know-how of interpreting them, to fail to specify as to which capturing of the skull landmarks corresponds to which part of the skull. Therefore as such, more clarity should be given so that everyone is catered for in understanding and appreciating the results of the images produced. More images produced by the AR system are shown and in one of these images two illustrations of the skull are shown and one depicts the end product of the alignment of the skull surface, whilst the other one depicts a virtual brain bearing a tumor in red (Hu et. al. 2013). See figure 10 below for these illustrations.

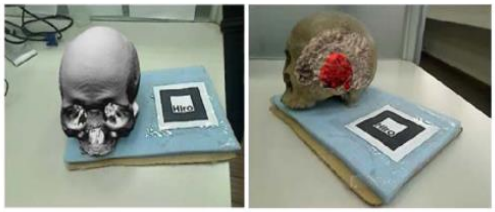


Figure 10 Additional images produced by the AR system (Hu et. al. 2013)

With the image whose emphases are on the alignment of the skull, Hu et. al. (2013) stresses that excellent accuracy is achieved as the resulting grey model adequately covers the skull's surface and this is very much true looking at the image given but there are no results regarding the accuracy being referred to, to fully support this. Proceeding on to the image depicting a virtual brain bearing a tumor in red, it is actually stated that the original data used for rendering this image is from another patient and not from the CT scans of the skull that is under assessment and that being the skull exemplar. Furthermore, it is emphasized that the skull used in the experiment is just an exemplar and therefore as such it does not consist of a brain nor a tumor and this brings about some kind of biasness or confusion because the experiment undertaken only specified the use of a skull exemplar and no other entity was mentioned at the beginning of the experiment. This raises questions of why this is being referred to and also questions of what the authors are trying to prove.

Hu et. al. (2013) goes on to say that the alignment of the brain and the skull is not achievable and this brings about even more biasness, and it therefore in a way confirms to one that the proposed system has some infidelity in performing such a task. It is mentioned that the AR system can work at a speed of 30 fps with resolutions of 640 * 480 pixels and without compromising performance, but no evidence is provided to prove that and as such those who would wish to verify this will not be capable of doing so.

Therefore as such, there is need for some adjustments in this experiment for extensive clarity.

3. Conclusions

A critical evaluation and analysis of the currently used technology in Robotic Assisted Surgeries has been made. The technologies stressed were the ENB and Robotics; multi-stereo camera system with reference to image-guided surgery markers and augmented reality based on image-guided surgery system. A number of factors were noted, such as coloured reference markers used in multi-stereo camera systems and also the optical tracker coordinate system and marker system of AR systems. The design of each of the technologies highlighted were mainly based on the limitations of the latter technologies, for instance, improving image-guided surgery systems with a multi-stereo camera which helps deal with the line of sight problem as highlighted before in the previous sections. The highlighted technologies were all tried and tested and their results provided and were all involved in real world experiments and hence providing results that can be used as a guide in verifying each one of the experiments with realistic objects such as of a skull exemplar in terms of AR.

Despite this though, there needs to be some extensive adjustments in terms of the experiments done on ENB and AR. Issues of sample size should be addressed more especially with ENB and issues of biasness should be eliminated with regards to AR, there should be one specific scope for the experiment undertaken.

References

Brooks, P., 2015. 'Robotic Assisted Thoracic Surgery for Early-Stage Lung Cancer: A Review.' *AORN Journal*, vol.102, no.1, pp.40-49.

Christie S., 2014. 'Electromagnetic Navigational Bronchoscopy and Robotic-Assisted Thoracic Surgery'. *AORN Journal*, vol.99, no.6, pp.750-764.

Corenthy, L., Garcia, M., Bayona, S., Santuy, A., Martin J.S., Benavides-Piccione, R., DeFelipe, J. & Pastor, L., 2014. 'Haptically Assisted Connection Procedure for the Reconstruction of Dendritic Spines.' *IEEE Transactions On Haptics*, vol.7, no.4, pp.40-49.

Elsamnah, F., Sediono, W., Khalifa, O. & Shafie, A, 2014, 'Multi-Stereo Camera System to Enhance the Position Accuracy of Image-guided Surgery Markers.' *The 5th International Conference On Computer & Communication Engineering*, Pages: 342-345.

Hu, L., Wang, M. & Song, Z., 2013, 'A Convenient Method of Video See-through Augmented Reality Based on Image-guided Surgery System.' *The Seventh International Conference on Internet Computing for Engineering and Science*. Pages: 100-103.

Jeevan, M., Jebavaj, R. & Khrishnakumar, R., 2014, 'In-vitro validation of image guided surgery system with 3D preoperative visualization for atrial transseptal puncture.' *The 18th International Conference on Information Visualisation*, Pages: 342-345.

Salguero, P., Abreu, S., Rolo, J. & Clain, S., 2015. 'An Interactive Web-Based Tool For Breast Reduction Surgery Simulation.' In *2015 International IEEE Symposium on 3D User Interfaces*. IEEE Symposium on 3D User Interfaces, pp. 181-182.

Silva, S., Madeira, J. & Santos, B.S., 2014, 'Computer-Assisted Myocardial Perfusion Assessment.' *The 18th International Conference on Information Visualization*. Pages: 336-341.

An Evaluation of Current Bio-Metric Fingerprint Liveness Detection

George Phillipson

Abstract

This research paper focuses on the critical evaluation of bio-metric fingerprint liveness detection to ensure the user is alive when authentication is taking place by checking their physical traits such as skin distortion, perspiration detection, pulse oximetry and finger vein recognition. An evaluation of each academic's research paper will be undertaking regarding their findings and how each method could be used in the real world to create a secure bio-metric fingerprint scanner for security applications. This paper will then make recommendations for which technique is most suitable for detecting the liveness of the user when being authenticated.

1 Introduction

Acara et.al (2013) stated that "users frequently reuse their passwords when authenticating to various online services" rather than having to remember multiple login details, this can lead to a user having multiple accounts hacked if their username or password is compromised. Because of this, research has been undertaking into various authentication methods such as bio-metric fingerprint detection. Galbally et.al (2012) stated that "the basic aim of biometrics is to discriminate automatically between subjects in a reliable way" and this can be done either by the user's physical or behavioural traits. Weaknesses in this type of authentication have been found when an "attacker (active imposter) tries to fake somebody else's identity by presenting fake samples of the person's traits" (Chingovska et.al. 2014).

1.1 Fingerprint Vulnerabilities

Bio-metric fingerprint security scanners have a vulnerability in that the fingerprint scanner can be spoofed. Reddy et.al (2008) claimed that with modest effort, many leading bio-metric scanners can be susceptible to fake fingerprints and easily spoofed, Sujan et.al. (2005) claimed that in laboratory tests, they were able to spoof bio-metric fingerprint scanners with Play-Doh with a success rate of between 45% and 90%.

Research by Espinoza et.al (2011) showed that with the correct type of equipment and time, casts of a finger could be produced so that to a bio-metric fingerprint scanner the spoof cast can look like a real finger. (See Figure 1 below)

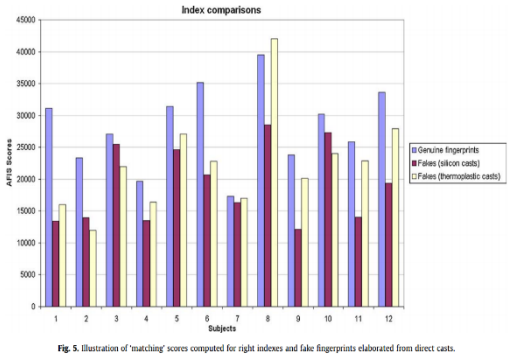


Fig. 5. Illustration of 'matching' scores computed for right indexes and fake fingerprints elaborated from direct casts.

Figure 1, Scores for fake finger made from cast (Espinoza et.al 2011)

Antonelli et.al (2006) argued that "Forging a fake finger is not as easy as some authors claim", but did admit that with time and the correct equipment, it can be done as no scanner that they tested was resistant to fake fingerprints. Espinoza et.al (2011) therefore concluded that without liveness detection, bio-metric fingerprint scanners can be spoofed. Research by many academics have proved this theory and because of this, research has been undertaking into how to add liveness detection into bio-metric fingerprint scanners to prove that the user is alive.

2.0 Skin Distortion Analysis

Research by Antonelli et.al (2006) looked into detecting the liveness of a user by their skin distortion by using video that has a camera resolution of 569dpi and 20fs to capture individual images of the finger as it is rolled over the sensor. Espinoza et.al (2011) argued that a scanner with a 500dpi could be spoofed and that the global trend is towards scanners with a resolution of 1000dpi.

Antonelli et.al (2006) hypothesis is that because the skin has peculiar characteristic such as the skins elasticity this will be unique to each person due to the bone structure and the underlining derma. In an attempt to prove their theory and because "no public available benchmark database could be used, due to the specific requirements of the fake detection algorithm" (Antonelli et.al 2006) they had to create a prototype bio-metric fingerprint scanner (see Figure 2 below)

Figure 2, Prototype bio-metric scanner (Antonelli et.al 2006)

and database to compare the images of the finger as it was rolled over the scanner and then they used Distortion map and Distortion code to validate whether the finger belonged to a living person or was fake. The experiments were carried out in laboratory conditions with the help of 45 volunteers, each volunteer had received a brief amount of training into how to use the scanner, but even with this training Antonelli et.al (2006) stated that the volunteers had a high error rate due to incorrect finger movement over the scanner. (See figure 3 below)

Main error cause	Num. of seq.
Incorrect finger movement	55
Too fast movement	15
Insufficient deformation due to low finger pressure	8
Wrong estimation of the centre	11
Other	11

Figure 3, Main cause of errors (Antonelli et.al 2006)

Antonelli et.al (2006) concluded that their prototype scanner was successful as it achieved the same success rate as other commercial bio-metric scanners at been spoofed. This conclusion was reached because Antonelli et.al (2006) stated that "attackers were aware of the particular fake-detection technique adopted and did their best to defeat it".

Antonelli et.al (2006) states attackers could only achieve the same spoof rate against their scanner as that of other commercial scanners, but they do not mention what other scanners they compared the results against, and as such, no comparison can be made. Also the scanner seems difficult to use as even with training the volunteers achieved a high error rate due to incorrect finger movement. Because of this, their conclusions have flaws in its validity.

2.1 Perspiration Detection

The researchers Sujan et.al. (2005) carried out an investigation into how to prevent bio-metric fingerprint scanners from been spoofed by fake fingerprints made of artificial material such as Play-Doh and gelatin. This involved monitoring the perspiration on the finger to decide whether the finger is from a living person or is fake, the experiments were carried out under laboratory conditions with 33 volunteers whose ages ranged from 20 to 60, mixed gender and ethnicities. The experiments used 3 bio-metric fingerprint scanners from different manufacturers using optical, electro-optical, and solid-state scanners, because these scanners can detect the perspiration on the sensor Sujan et.al. (2005) wrote an algorithm that could detect "a distinctive spatial moisture pattern which evolves in time due to the physiological perspiration process". The time period that the user had to keep their finger on the scanner to produce the perspiration pattern ranged from 2 seconds to 5 seconds (See Figure 4 below).

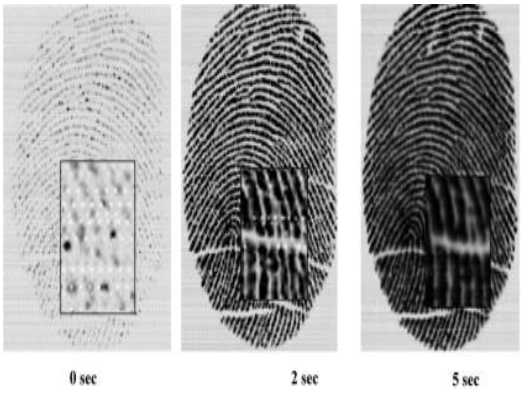


Figure 4, spatial patterns (Sujan et.al. 2005)

Espinoza and Champod (2011) did raise the issue that the 5 seconds it takes for the scanner to detect the perspiration could be too long for biometric authentication at a countries borders.

The perspiration process they looked for was around the pores and not the pores themselves and as such was a sign of perspiration from a living person. Because of the perspiration around the pores, Sujan et.al. (2005) argues that their method is harder to spoof because the fake finger would need to replicate perspiration coming from the pores and then moving over the ridges on the fingerprint.

Tan and Schuckers (2010) agreed with Sujan et.al. (2005) as they claimed that "live fingers have a distinctive perspiration pattern along the ridges, but the spoof fingers do not have". Antonelli et.al (2006) also agreed with this claim as they stated that perspiration from the pores makes the ridge lines on the finger appear darker over time.

Sujan et.al. (2005) pointed out that further research is needed as some users who have either dry or moist fingers could produce a false reading. This issue was also raised by Espinoza and Champod (2011) who stated "people presenting abnormal skin conditions could be unable to use this method". Sujan et.al. (2005) also stated that research would need to be carried out in hot and cold climates as these types of environments could affect the reading from the scanner.

Sujan et.al. (2005) explained that the experiments where based around scanners that allows the software to be updated. Because of this, the scanners do not need any further hardware added to the system which could be costly and also add extra security considerations (Sujan et.al. 2005).

With all the research carried out, Sujan et.al. (2005) claimed that cadaver and spoof fingers did not produce a perspiration pattern due to the lack of pores that gave out perspiration and as such their test returned a success rate of about 90%.

Sujan et.al. (2005) do show evidence to justify their claims that detecting perspiration can help differentiate between real and fake fingerprints with a success rate of 90%, but they do not state how many of their volunteers had skin conditions or what type of condition will affect the scanner. Sujan et.al. (2005) also failed to produce any evidence of how hot or cold climates can affect the scanner even though this was mentioned as an issue. Because of this, it is difficult to validate Sujan et.al. (2005) claims that they had a success rate of 90% without this information being supplied.

2.2 Pulse Oximetry

Reddy et.al. (2008) undertook research into liveness detection by investigating the amount of oxygenated blood in the finger by detecting the pulse oximetry while the finger was being scanned.

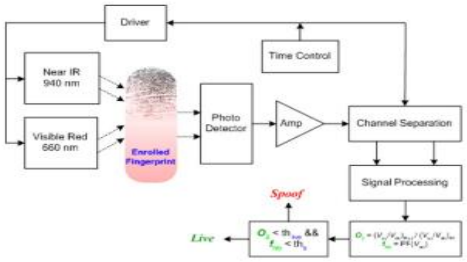


Figure 5, Block diagram of proposed anti-spoofing system (Reddy et.al. 2008)

The prototype system developed had 2 LEDs for emitting light on the wavelengths of red 660nm and near infrared 940nm. Reddy et.al. (2008) stated that "This is because the two common forms of the molecule, oxidized haemoglobin and reduced haemoglobin (Hb) have significantly different optical spectra in the wavelength range from 500 nm to 1000 nm".

To test their theory Reddy et.al. (2008) created spoof fingers made from gelatin and whole finger where made from Play-Doh. 18 volunteer's male and female took part in the experiments and had an age range between 17 and 58. Reddy et.al. (2008) stated that in the case of a real finger the presence of a pulse was detected along with the change in amplitude with the IR and Red signals, but a fake finger did not have this pulsating nature of the signal (See Figure 6 below).

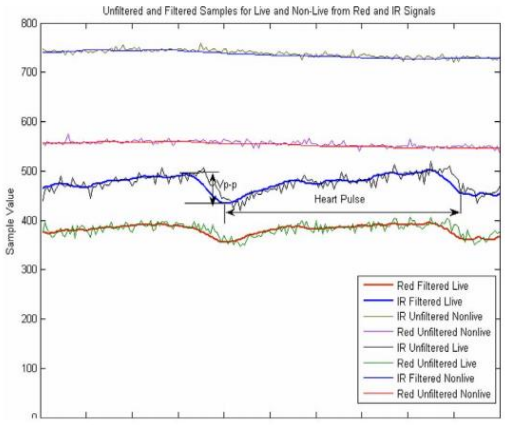


Figure 6, Unfiltered and filtered samples from red and infra-red signal using live and nonlive samples (Reddy et.al. 2008)

Reddy et.al. (2008) produced a very detailed report into monitoring the amount of oxygen in the blood while the finger is being scanned, they did conclude that issues may arise in that the health of the user could affect the results and this was also the conclusion of Peng et.al (2014) who stated that "biometric systems are seriously affected by the condition of the user's health". Further issues with the scanner were that if the user moved their finger over the scanner incorrectly, it could affect the SpO2 value.

Reddy et.al. (2008) also stated that the operational environment could affect the measured SpO2 values and as such the 100% accuracy of the sensor could not be assumed in these working conditions. The reason for this is because "the accurate measurement of the ambient light and its complete removal from the received signal is extremely difficult" (Reddy et.al. 2008).

Reddy et.al. (2008) did claim that this issue could be overcome if the scanner is kept in complete darkness with no ambient light or outside under the sun to overcome this problem.

As the prototype was an extra device added to the fingerprint sensor, it was argued by Sujan et.al (2005) that extra hardware added to the sensors can results in extra costs and if not correctly integrated with the scanner could lead to possible spoofing. This conclusion was also backed up by Tan and Schuckers (2010) who also claimed that extra hardware can be expensive, bulky and inconvenient.

Reddy et.al. (2008) gave a good explanation of how the bio-metric fingerprint scanner works at detecting the SpO2 levels and also explained the age range of the volunteers and gender as well as the experiments undertaking and issues found when using the scanner.

Although the scanner was able to tell the difference between a real and fake finger, Reddy et.al. (2008) did not produce any evidence as to how accurate the scanner was compared to other devices or how the health of the user affected the results, because of this, no comparison could be made as to the success of the scanner compared to other devices.

2.3 Finger Veins Recognition

Research by Kang et.al (2015) looked into bio-metric fingerprint scanners that analysed finger veins patterns, fingerprints and knuckles. As no online database holds data on all three, Kang et.al (2015) created their own with 1890 images.

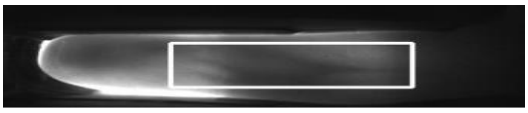


Figure 7, Finger Vein under NIR (Kang et.al (2015)

The Near Infrared (NIR) wavelength used for finger veins recognition was 850nm. Under NIR, the vein in the finger can be seen (See Figure 7 below).
The experiments carried out by Kang et.al (2015) was to "compare the recognition performance of three algorithms: the minutiae-matching method, LBP algorithm, and ORB algorithm on contactless fingerprint". For the finger vein recognition, they used the LBP algorithm but unfortunately due to low quality NIR images the vein recognition was poor.

Pflug et.al (2012) has also investigated the use of vein pattern bio-metric fingerprint recognition using NIR, Pflug et.al (2012) stated that as the "vein patterns are located underneath the skin, a vein pattern is hard to forge without the data subject's knowledge". Laboratory experiments where conducted using two vein databases with a NIR wavelength of 850nm. The datasets used where the GUC45 and the UC3M, the results for the two datasets showed that the UC3M had excellent bio-metric performance while the GUC45 had poor bio-metric performance due to the images been of low contrast.

Pflug et.al (2012) and Kang et.al (2015) have looked into bio-metric fingerprint security using veins patterns and both appear to have had problems returning accurate results due to poor image quality for NIR.

Pflug et.al (2012) and Kang et.al (2015) hypothesis behind using the vein patterns in the finger for bio-metric security is sound as it would be very difficult to spoof the vein of a user either directly from an actual finger or indirectly from a finger print taken from a glass or photo using a high dpi camera. The problems Kang et.al (2015) and Pflug et.al (2012) encountered was the poor image quality from NIR of the vein which made it difficult to achieve an accurate image and until this problem is solved the scanner cannot be used in an environment where security is paramount.

3 Conclusions

Academics have looked at different way to detect the liveness of a user when logging into a system. Reddy et.al. (2008), Kang et.al (2015) and Pflug et.al (2012) all looked into using the near infrared (NIR) wavelength to detect the liveness of the user and all had varying degrees of success. Reddy et.al. (2008) used the NIR to detect the presence of a pulse while

Kang et.al (2015) and Pflug et.al (2012) used the NIR to detect the pattern of the finger vein.

Antonelli et.al (2006) took a different approach to detecting liveness by looking at how the skin was distorted while it moved over the scanner.

Unlike Reddy et.al. (2008), Antonelli et.al (2006) approach did not have any issues with the ability to detect the health of the user which could possibly have privacy issues for bio-metric scanners used at border controls.

Sujan et.al. (2005) approach to overcoming the issue of spoofing was to look for signs of sweat been secreted from the pores on the finger and then looking for the moisture pattern which would develop around the pores, but this technique did have issues with user's who had skin conditions.

If the research by Reddy et.al. (2008) into pulse oximetry could overcome the issues effecting the scanner and the research by Kang et.al (2015) and Pflug et.al (2012) into finger vein pattern recognition could also overcome the issues with poor image quality, then combined these two techniques could make a very secure bio-metric fingerprint scanner. As such further research in this field is required to produce a scanner that is both unobtrusive to the user and also able to detect if a user is alive when being authenticated.

References

Acara T, Belenkiy M, Küpçüc A, 2013, 'Single password authentication', *Computer Networks*, 57(13), p. 2597–2614.

Antonelli A, Cappelli R, Maio D, Maltoni D, 2006, 'Fake Finger Detection by Skin Distortion Analysis', *IEEE Transactions On Information Forensics And Security*, 1(3), p. 360-373.

Chingovska I, Rabello d A A, Marcel S, 2014, 'Biometrics Evaluation Under Spoofing Attacks', *IEEE Transactions On Information Forensics And Security*, 9(12), p. 2264-2276.

Espinoza M, Champod C, 2011, 'Risk evaluation for spoofing against a sensor supplied with liveness detection', *Forensic Science International*, 204(1-3), p. 162–168.

Espinoza M, Champod C, Margot P, 2011, 'Vulnerabilities of fingerprint reader to fake fingerprints attacks', *Forensic Science International*, 204(1-3), p. 41-49.

Galbally J, Alonso-Fernandez F, Fierrez J, Ortega-Garcia J, 2012, 'A high performance fingerprint liveness detection method based on quality related features', Future *Generation Computer Systems* 28(1), p.311–321.

Kang W, Chen X, Wu Q, 2015, 'The biometric recognition on contactless multi-spectrum finger images', *Infrared Physics & Technology*, 68(), p. 19–27.

Peng J, El-Latif A, Li Q, Niu X, 2014, 'Multimodal biometric authentication based on score level fusion of finger biometrics', *Optik - International Journal for Light and Electron Optics*,125(23), p. 6891–6897.

Pflug A, Hartung D, Busch C, 2012, 'Feature extraction from vein images using spatial information and chain codes', *Information Security Technical Report*, 17(1-2), p. 26–35.

Reddy P. V, Kumar A, Rahman S. M. K., Mundra T. S., 2008, 'A New Antispoofing Approach for Biometric Devices', *IEEE Transactions On Biomedical Circuits And Systems*, 2(4), p. 328-337.

Sujan T. V, Reza D, Hornak L A, Stephanie A, 2005, 'Time-Series Detection of Perspiration as a Liveness Test in Fingerprint Devices', *IEEE Transactions On Systems, Man, and Cybernetics—Part C: Applications and review*, 35(3), p. 335-343.

Tan B, Schuckers S ,2010, 'Spoofing protection for fingerprint scanner by fusing ridge signal and valley noise', *Pattern Recognition*, 43(8), p. 2845–2857.

A Critical Evaluation of Current Research into Malware Detection Using Neural-Network Classification

Tebogo Duduetsang Ramatebele

Abstract

Malware Detection is utilized by companies and individuals in order to maximize security within their computer systems. This paper is focused on the research being conducted, which demonstrates a variety of techniques for detecting novel malware attacks using Neural- Network classification. Most of these techniques come with their restrictions, and comparisons have been made in order to demonstrate the most competent neural- network technique.

1 Introduction

Modern forms of malware such as internet worms, viruses and Trojan horses pose major threats to copious computer systems linked to the internet, rendering archetypal security defenses such as antiviruses ineffective. Propagation of these threats are propelled by networked criminal hosts with the main aim of performing illegal engagements in the form of spam messages and the collection of confidential data leading towards legal repercussions and major financial losses(Azlab et. al. 2012). These threats have become a major disquiet in the world of computing as it is evident in the antiquity of malware, anti-malware reports and trend prediction attempts as criminality carries on to evolve in both scale and sophistication (Azlab et. al. 2012). Hence in protection from the proliferation of malware within the internet, anti-malware developers profoundly rely on machine learning analysis with the use of neural networks to detect novel variants of malware for defence purposes. Therefore the main focus of this current paper is to evaluate techniques utilised to detect novel malware on the internet today with the use of neural-networks in the machine learning field. According to (Kruczkowski and Niewiadomska-Szynkiewics 2014), since malware detection is a critical threat within the Internet, it is essential to perform an extensive analysis in order to improve the results of malware detection within the internet world. Furth more it is stated that malware analysis requires support by methods which are capable of heterogeneous data classification. This is evident by the research by (Ahmed et. al. 2016), as they state that the strong point of neural networks is classification of data which is essential for anomaly detection. The limitations of current research pertaining neural networks is its ability to provide identification and classifications of a network's activity on basis of limited, incomplete and data sources which are non-linear thus giving incorrect anomaly detection results (Tammi et. al. 2015). So in order to mitigate this issue merging a clustering algorithm with a classification algorithm is helpful by splitting the data into different groups and training the clustered data with a neural network classification technique.

Related studies illustrate that there are a few classification techniques within the neural networks field for effective malware detection. These classification approaches are the Multilayer Perceptron (MLP), Back Propagation (BP), Hybrid Neural Network- K means Probabilistic Neural Network (K means/PNN), Classification methods.

The Multi-layer Perceptron (MLP) Classification method proves to be significant networks in maintaining profuse large non- linear problems such as malware detection with ease. This is possible due to the structure of the multi-layer perceptron layout which consists of an assortment of neurons connected together in a network (Tammi et. al. 2015). Back Propagation (BP) Classification ensures a robust searching ability as it possess a very robust learning ability which can gain optimal solutions for data sets.

This is significant in malware detection as used by many researchers in recent years (Qian 2014).

The scope of this research paper will be fixed on the approaches of modelling neural-network classification concepts in order to asses which methods yield the best results for malware detection within the Internet world. Basis of this paper will also pay particular attention to the network models along with learning algorithms which will be utilised for training the neural network models.

The organisation of the paper is as follows; Neural-Networks are introduced in Section 2. The sub-sections which follow introduce a detailed background literature of the current malware detection techniques. The Multi-layer Perceptron (MLP), Back Propagation (BP), and the K means Probabilistic Neural Network (K means/PNN) Classification technique statistics, presenting data from experiments and evaluations will be described and discussed. Section 3 presents the comparison of techniques stipulated in the Section 2 sub-sections. In Section 4, lesson learnt from the research will be stipulated. Finally, Section 5 presents conclusions and this will be the last section of the paper.

2 Neural-Networks

Sen et. al. (2014) discovered that approaches based on the Artificial Intelligence field are more effective in detecting malware threats as compared to other approaches because in traditional approaches human efforts are required but with the use of artificial intelligence human efforts are minimal. Learning algorithms such as Neural-networks are used for malware detection. Neural-networks, as researched by (Ishitaki et. al. 2015) prove that among other rule-based computing techniques it sought to be to be the most effective and commonly used in malware detection approach, as it is are proficient in finding patterns for normal and abnormal behaviour. Furth more, studies stipulate that neural-networks are instrumented with the capability to detect normal connections along with attack connections. There are a variety of Classification techniques used within neural-networks which could be used in malware; this section will explore these techniques with discussions of work being conducted within that field.

2.1 Neural Networks and Standard Malware Detection Techniques

Neural-networks are trained in order to model specific non-linear, real world problems, this is evident through research by (Tammi et. al. 2015) to illustrating how it is the best in terms of effectiveness rather than standard malware detection techniques.

The researchers (Ozsoy et. al. 2015) study two different classification algorithms which can be used for malware detection using the University of Mannheim malware dataset with the use of samples of 1,087 malware programs from the Offensive Computing website.

Family	Train	Test-1	Val	Test-2	Total
Vundo	14	2	5	21	42
Emerleox	10	5	4	33	52
Virut	8	3	7	46	64
Sality	12	2	4	46	64
Ejik	7	6	4	101	118
Looper	10	3	6	145	164
AdRotator	14	1	2	119	136
PornDialer	11	6	4	196	217
Boaxxe	13	6	0	211	230

Table 1 The University of Mannheim malware sample of dataset with 1,087 malware programs (Ozsoy et. al. 2015).

As professed by (Ozsoy et. al. 2015) this collected data was evidently segmented into training, testing and validation as shown in Table 1, according to Microsoft's classification which identifies malware families. Machine learning used a ratio of training-test-validation set of 60%-20%-20% whereby two test sets: Test-1 containing 34 randomly selected malware programs and evaluation of features. Test- 2 on the other hand containing 918 malware was utilised for evaluation of online malware whereby the remaining malware was contained within the validation set in order to explore malware detection and training arrangements.
These malware families are used to compare Neural Network and Logistics Regression malware detection models as seen below:

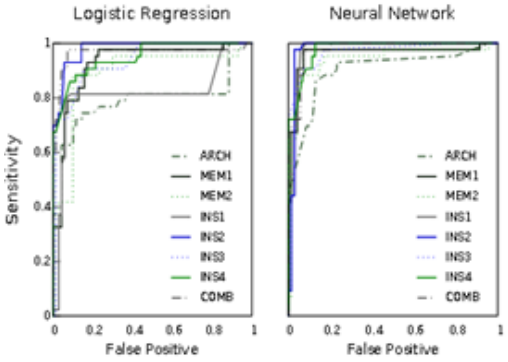


Figure 1 Comparison of Neural Network and Logistics Regression malware detection models (Ozsoy et. al. 2015) in Detection Performance of all Features in terms of Sensitivity.

The malware detection technique against which the Neural-network was compared to Logistics Regression.

As declared by Ozsoy et. al. (2015) "These features provide the highest accuracy among the set we considered: Figure1 shows that most of the instruction based features achieve nearly 100% sensitivity with around 10% false positive rate using the NN model. The LR model is less effective than NN model for all features."

Even though the researchers' claims have evidence to support their claims, their research is to some extent limited. This is because even though they were able to measure accuracy of malware detection through the measure of Sensitivity they failed to address the key attributes of the Neural-Network Model such as the type of neural network used, its input nodes, the training algorithm which was used, the number of layers used, whether or not there are hidden layers and which learning algorithm is used, hence reproduction of the results will not be possible and making modifications to the experiment will not be possible using the evidence depicted in their research.

Despite the fact that the researchers (Ozsoy et. al. 2015) state that Logistics Regression is equivalent to a single Perceptron within a Neural-network, hence they expect that NNs will give a better performance in terms of implementation complexity, comparing one traditional malware detection technique against the whole Neural-network field is far reaching as this does not give a fair overview of all traditional malware detection techniques. The reason being that an extensive research should cover experiments illustrating various traditional malware detection techniques as compared to various Neural-networks based techniques for malware detection. Performing these tasks will give a fair claim as to which type of malware detection technique is the most accurate over the other.

2.2 Multi-layer Perceptron (MLP) Classification technique

The multi-layer perceptron is the most commonly used feed forward network which is most likely to have several layers (Hegadi and Kamble 2014). In this classification technique, neurons are organized into series of layers and information of signal flows through particular networks, exclusively in one direction from the input the output layers (Tkac and Verner 2016).

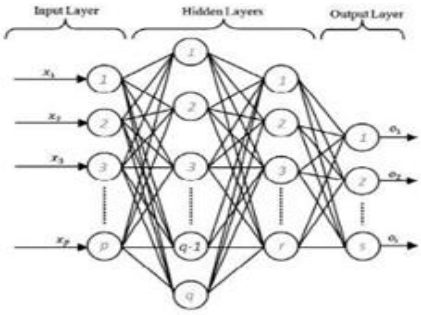


Figure 2 Architecture of Multi-layer Perceptron (MLP) neural network in wrapper approach algorithm (Hegadi and Kamble 2014).

A Multi-Layer Perceptron with two layers is fit for classification of input patterns for anomaly detection (Jadidi et. al. 2013). Within the neural-network field it is essential to train this network in order to adjust its weights and biases which may occur, hence ensuring that error is minimized.

Singh and Bansal (2013) proposed a model which evaluates various algorithms such as the Multilayer Perceptron, Logistics Regression, and Radial Basis Function with the use of data mining tool known as WEKA on a NSL KDD data set.

The framework which was designed and implemented to for algorithm evaluation was as follows:

1. Reading and Input of the Dataset – where NSL KDD dataset used for classification purposes with 25192 classified instances which consist of 41 attributes and 12 attributes chosen for analysis.
2. Selection of Data mining tool – loading data in WEKA data mining tool
3. Selection of Algorithm - upon loading of dataset, measuring of detection rate by each chosen algorithm
4. Cross Validation – the data set tested with the aid of chosen neural network independently, measuring accuracy and classifying malware attacks
5. Performance Evaluation - Evaluation and comparison with other algorithms.

Figure 3 Algorithm for proposed model (Singh and Bansal 2013).

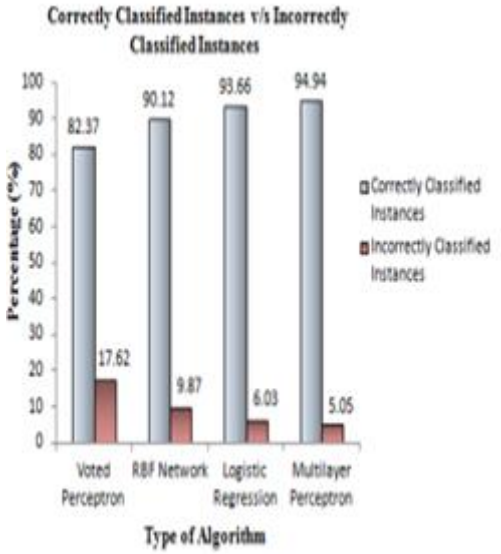


Figure 4 Algorithm accuracy (Singh and Bansal 2013).

The results demonstrated by the Figure 4 graphical representation portrays instances which are classified incorrectly by each algorithm. Subsequent to classification of the NSL KDD dataset, the researchers explain that it is clearly unmistakable that the Multilayer Perceptron Algorithm displays the highest level of accuracy as compared to the other algorithms.

Table 2 below illustrates the results of the performance of each of the four algorithms, after cross validation of the NSL KDD data set .Using this table the following graphical representations were yielded, when testing each neural network algorithm.

ALGORITHM	CCI	ICI	KAPPA STATISTIC	MAE	RMSE	RAE	RRSE	Time Taken(s)
RBF Network	90.12	9.87	0.79	0.14	0.27	29.81	54.74	3.66
Voted Perceptron	88.19	11.80	0.76	0.11	0.34	23.71	68.87	44.93
Logistic	93.66	6.33	0.87	0.09	0.22	19.85	44.42	10.4
Multilayer Perceptron	94.94	5.05	0.89	0.07	0.19	14.88	39.30	26.72

Table 2 Measure of Performance Results (Singh and Bansal 2013).

Furthermore the researchers (Singh and Bansal 2013) use the Kappa measure to assess the agreement between classifications and their classes. Subsequent to cross validation the results are illustrated in Figure 5 below. These results clearly demonstrate that the Multilayer Perceptron algorithm is the most accurate in terms of performance as it is the furthest values from zero, showing that it is performing better than chance.

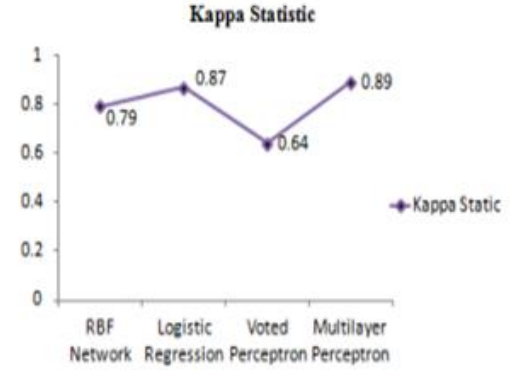


Figure 5 The Kappa Statistic (Singh and Bansal 2013).

When evaluating the mean absolute error of each of the algorithms on Figure 6, the researchers explain that on this assessment, the Voted Perceptron proved to be the highest on basis of error magnitude as compared to other algorithms.

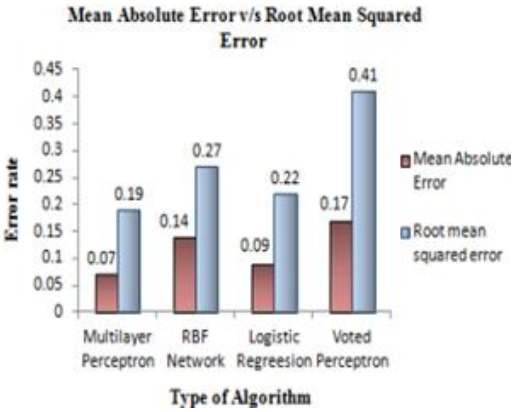


Figure 6 MAE/RMSE (Singh and Bansal 2013).

Lastly, evaluation was based on the computational time taken for each algorithm to perform testing and training on the data as seen on Figure 7. After classification it clearly shows that the Voted Perceptron algorithm took more time to build a model for training data as compared to the other algorithms.

Since the researchers are proving the claim that the multilayer perceptron displays most accuracy as compared to other neural network algorithms, Table 2 as well as Figures 4 and 5, depict that in an excellent unambiguous manner. This is because these graphical representation show clear statistics regarding how accurate the Multilayer perceptron is, as compared to other algorithms. In Table 2 the multilayer perceptron has the highest CCI of value 94.94 as compared to others and in Figure 4 the multilayer perceptron has the highest value of correctly classified instances with a value of 94.94%. Moreover figure 4 shows the highest Kappa value of 0.89.

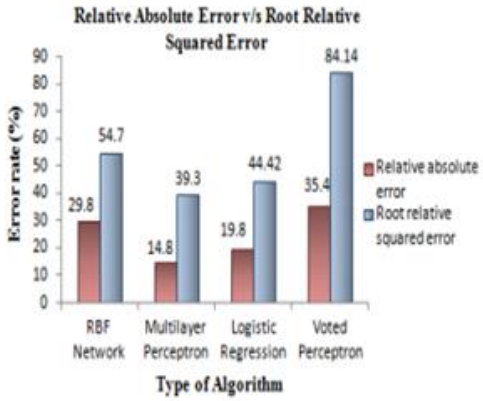


Figure 7 RAE/RRSE (Singh and Bansal 2013).

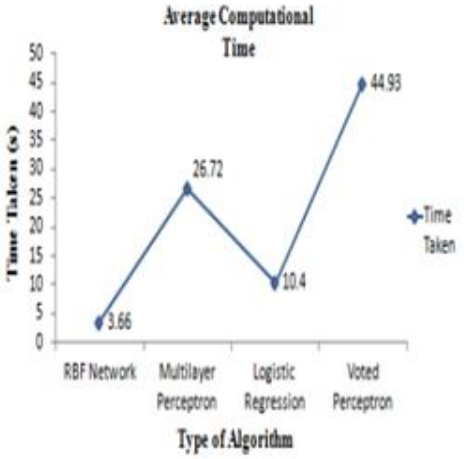


Figure 8 Average Time Taken (Singh and Bansal 2013).

However what is far reaching about this research is that even though figure 6 shows that the multilayer perceptron has the lowest mean absolute error, they fail to discuss that matter but rather they focused on discussing that the Voted perceptron has the highest mean squared error, which is irrelevant to the aim of the aim of the research. This also applies to figure 7 as the same issue raised. Figure 8 seemed to have no relevance to proving that the Multilayer perceptron provides the best accuracy as its computational time was moderate at a value of 26.72. Discussion of results

should have been only based on proving that the multilayer perceptron is the best in terms of accuracy.

2.3 Back Propagation (BP) Classification Technique

The Back Propagation Classification (BP) technique is the most commonly used approach in Neural-networks for data classification. "Different supervised techniques are used for classifying data, out of which Back Propagation neural network has been used extensively" (Sen et al. 2014).It has been used extensively for distinguishing between normal and abnormal activities (Sen et. al. 2014). Below is a figure showing Back Propagation Learning (Ghosh & Chakraborty 2012).

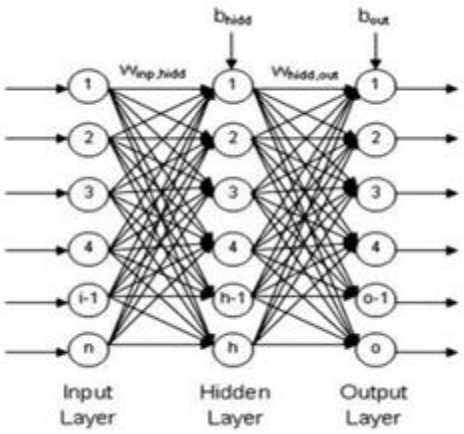


Figure 9 Back Propagation on a Multi-layered Neural-network topology. (Ghosh and Chakraborty 2012)

To further validate the superiority of BP a proposed model was presented by the researchers (Sen et. al. 2014) using 40 features and 1000 epochs for experimentation from a KDD data set. This experiment was carried out using two distinguished percentage splits of a KDD dataset where 70% of the data is for training and 30% of the data is used for testing as seen on Table 3. The second split contained 80% of the data which is used for training while the remaining 20% is utilised for testing as seen on Table 4.
After experimentation results were tabulated as shown below:

No. of hidden layer	No. of hidden nodes	Learning factor	MSE	Training			Testing			Execution time (sec)
				False Positive	Accuracy	Detection %	False Positive	Accuracy	Detection %	
4	3-4-5-3	0.015	0.016081	0.32	0.98	99.58	1.85	0.97	98.97	156.38
4	4-3-5-4	0.04	0.013247	0.38	0.99	99.30	0.14	0.97	97.84	138.24
4	7-6-8-9	0.04	0.005879	0.47	0.99	99.12	7.06	0.96	97.37	305.39
4	3-4-4-5	0.07	0.007967	0.91	0.99	98.29	8.61	0.95	97.02	124.92
4	8-9-12-15	0.035	0.006093	0.70	0.99	99.05	3.39	0.96	98.72	524.3
4	4-5-6-4	0.09	0.008846	0.97	0.99	98.18	8.97	0.96	96.77	165.89

Table 3 Malware Detection Accuracy (Sen et. al. 2014).

ALGORITHM	CCI	ICI	KAPPA STATISTIC	MAE	RMSE	RAE	RRSE	Time Taken(s)
RBF Network	90.12	9.87	0.79	0.14	0.27	29.81	54.74	3.66
Voted Perceptron	88.19	11.80	0.76	0.11	0.34	23.71	68.87	44.93
Logistic	93.66	6.33	0.87	0.09	0.22	19.85	44.42	10.4
Multilayer Perceptron	94.94	5.05	0.89	0.07	0.19	14.88	39.30	26.72

Table 4 Algorithm Accuracy for Malware (Sen et. al. 2014).

According to these researchers Sen et. al., (2014), the results on both tables (Table 3 and Table 4) represent high percentages between 96-99%, of malware detection, proving to have almost flawless accuracy. Different combinations of hidden nodes provide good learning factors and low Mean Square Errors. During training on both tables the false positive rates are low, the accuracy rate is almost 1, which is very high. Moreover during the Testing phase false positive rates increase slightly on both tables but the accuracy still remains high, proving that back propagation is a good technique for malware detection. Also these tables show that the more layers of a neural network, the better yielding of detection results.

However it is over reaching in the sense that comparisons were not compared against other classification techniques such as the Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF).Fair comparisons should be made using narrow-based results based on these different techniques.

Moreover another limitation is that calculations were not made based on the all main aspects of the Back Propagation (BP) topology as only hidden layer was used to determine malware detection accuracy. Further studies should include all the aspects of the Back Propagation technique topology such as the input variables to the output variables.

2.4 Hybrid *K* means Probabilistic Neural Network (*K* means/PNN) Model

Tammi et. al. (2015) proposed a model which is divided into 3 stages as seen on Figure 10. In order to utilise this model, the NSL-KDD99 data set should be split into two parts which are the training and test set, set to ratio of 70:30 respectively, clustering the training set with k- means clustering, and implementing Neural- network classification on the final stage.

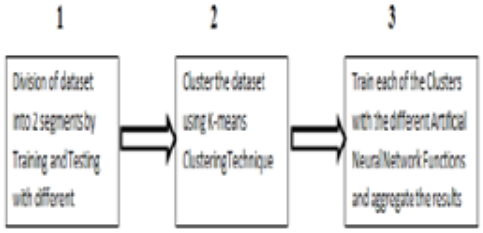


Figure 10 Proposed Model Outline. (Tammi et al.2015).

These three phases are the division of dataset into two groups, which are the a) training and b) testing datasets, training dataset with *k*- mean, five Clusters for the five different classes. Various Neural-networks trained by data sets which are clustered and aggregation of neural-networks to enhance target result on or various neural network functions.

Upon working on the dataset, the nominal data was converted to numeric value, to increase effectiveness in neural network classification calculation. The dataset included four categories of attacks namely: Dos, Probe, U2R and R2L including a normal class. This data set contained 41 attributes and after preprocessing four categories of datasets by feature selection techniques were created.

Tammi et. al. (2015) continues to explain that since *k* -Means clustering is one of the simplest, fast and robust data mining procedures, it will give the best result when given any distinct dataset which is essential for malware detection. The *k*-means algorithm may be summarized as follows:

Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centers.

1) Randomly select 'c' cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is the minimum of all the cluster centers

4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

Where, 'c_i' represents the number of data points in i number of clusters.

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3.

Figure 11 *K-Means* algorithm (Tammi et. al. 2015)

Tammi et. al. (2015) claims that, hybridising the k - Means technique with the Probabilistic Neural Network ensure the most accurate technique for malware detection, because it will combine both positive features such as high speed from PNN and robustness in anomaly detection from the k - means algorithm.

The results obtained from the proposed system, implemented on the WEKA tool with the use of different datasets are as follows:

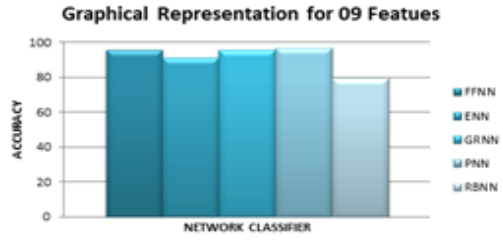


Figure 12 Accuracy of various NN using 9 features (Tammi et. al. 2015)

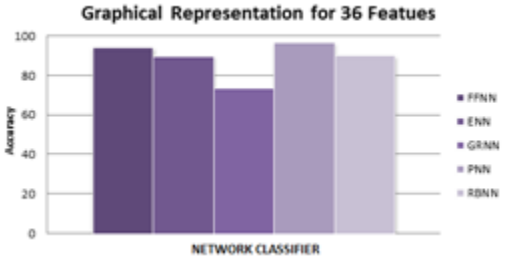


Figure 13 Accuracy of various NN using 36 features (Tammi et. al. 2015)

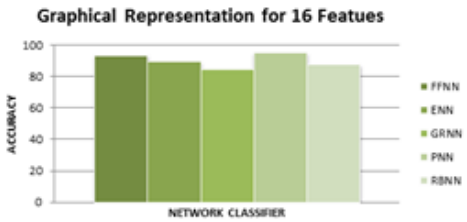


Figure 14 Accuracy of various NN using 16 features (Tammi et. al. 2015)

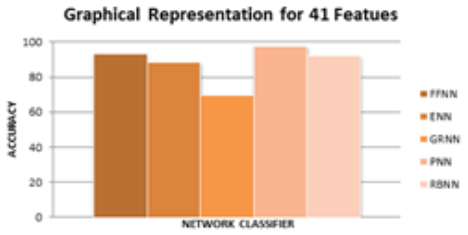
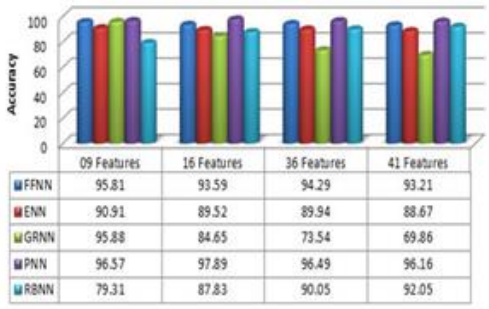


Figure 15 Accuracy of various NN using 41 features (Tammi et. al. 2015)



	09 Features	16 Features	36 Features	41 Features
FFNN	95.81	93.59	94.29	93.21
ENN	90.91	89.52	89.94	88.67
GRNN	95.88	84.65	73.54	69.86
PNN	96.57	97.89	96.49	96.16
RBNN	79.31	87.83	90.05	92.05

Figure 16 Accuracy of various NN using 9 features (Tammi et. al. 2015)

Within this study, the researchers have proved that, Probabilistic Neural Networks give the best accuracy over other Neural Networks, which is why hybridisation of PNN and k Means is necessary.

Despite the difference in datasets and number of features (Tammi et. al. 2015), they have fully backed their claims. That proves that hybrid models are more accurate as compared to MLP or BP.

3 Comparison of Techniques

Regarding all the malware techniques it is difficult to define which the most competent malware detection technique when it speed performance is concerned, Neural-networks on their own cannot achieve better because of their particular functionalities within the Internet. Hegadi and Kamble 2014) described a multi-layer perceptron (MLP) as the most accurate when compared to other algorithms such as Multilayer Perceptron, Logistics Regression, and Radial Basis Function and this was true. However in some instance multilayered perceptron did not perform optimally for malware detection. This method had its downfalls looking at performance as its ability to provide identification and classifications of a network's activity on basis of limited, incomplete and data sources which are non-linear thus giving incorrect anomaly detection results. In order to mitigate this drawback the Hybridisation is introduced as it focuses on classification based of real time attacks probabilities. The model employed a *k*-means Probability Neural Network Model.

4 Lessons Learnt

The main issue which was to be tackled is to know which technique will be the best to use for malware detection in neural networks, after research it had showed that not only is one technique important, but all the techniques are important but all of them have different features which may be combined together to create a hybrid malware detection techniques for flawless malware detection**.**

5 Conclusions

Every research paper which was evaluated implemented neural networks into malware detection. During evaluation their strengths and weaknesses where highlighted and the best solution is to combine these malware techniques to create a powerful neural network malware technique. Furth more the Probability Neural Network Model alone proves to be efficient as it shows accurate malware detection rates of 96-97%. This goes to show that the combination between the PNN and k means will be a very powerful tool in malware detection with less errors.

References

Ahmed M., Mahmood A., Hu J., 2016, 'A Survey of Network Anomaly Detection Techniques,' *Journal of Network and Computer Applications*. Vol. 60, pages. 19-31

Azalab M., Huda S., Abawazy J.,Islam R., Yearword J., Venkatraman S., Broadhurst R., 2014, 'A Hybrid Wrapper-Filter Approach for Malware Detection,' *Journal of Networks*, Vol. 9, No. 11, pages. 2878-2891

Ghosh A., Chakraborty M., 2012, 'Hybrid Optimized Back Propagation Learning Algorithm for Multi-layer

Perceptron', *International Journal of Computer Applications*, Vol. 57, pages. 1-6

Hegadi R., and Kamble P., 2014, Recognition Handwritten Numerals Using Multi-Layer Feed Forward Neural Network,' *IEEE World Congress on Computing and Communication Technologies*, pages. 21-24

Ishikati T., Elmazi D., Liu Y., Oda T., Barolli L., Uchida K., 2015, 'Application of Neural Networks for Intrusion Detection in Tor Networks,' 29th *International Conference on Advanced Information Networking Applications Workshops*, pages. 67-72

Jadidi Z., Muthukkumarasamy V., Sithirasenan E., Sheikan, M, 2013, 'Flow-Based Anomaly Detection Using Neural Network Optimized with GSA Algorithm,' *IEEE 33rd International Conference on Distributed Computing Systems Workshops,* pages. 76-81

Kruczkowski M., Niewidomska-Szynkiewics E.., 2014, 'Support Vector Machine for Malware Analysis and Classifications,' *IEEE International Conference on Web Intelligence (WI) and Intelligent Agent Technologies (IAT),* pages. 415-420

Ozsoy M., Donovick C., Garelick I, Ghazalah M., Ponomarev D., 2015, 'Malware- aware processor: A framework for Efficient Online Malware Detection, *21st IEEE International Symposium on High Performance Computer Architecture*, pages. 651-661

Qian Q., Cai J., Zhang R., 2014, 'Intrusion Detection based on Neural Networks and Artificial Bee Colony Algorithm' *IEEE Journal*, pages.1-6

Sen N., Sen R., Chattopadhyay, 2014, 'An Effective Back Propagation Neural Network Architecture for the Development of An Effective Anomaly Based Intrusion Detection Systems,' *Sixth International Conference on Computational Intelligence and Communication Networks*, pages. 1052-1056

Singh S., and Bansal M., 2013, 'Improvement of Intrusion Detection System in Data Mining using Neural Network,' *International Journal of Advanced Research in Computer Science & Software Engineering*, Vol 3., Issue 9, pages. 1124-1130

Tammi W., Biswas N., Nasim Z., Shorna K., 2015, 'Artificial Neural Network Based System for Intrusion Detection using Clustering on Different Feature Selection', *International Journal of Computer Applications,* Vol. 126,No.12 pages.21-28

Evaluating Indirect Detection of Obfuscated Malware

Benjamin Stuart Roberts

Abstract

In the ever evolving world of malware design, great steps have been taken to reduce the effectiveness of traditional detection techniques. Due to this there is an increasing rise in the amount of malware which cannot be easily detected by standard methods or can easily be modified to create new variants, thereby evading signature based scanners. This paper analyses three different methods for the indirect detection of malware, system, network, and user behaviour. An evaluation of this work is done and recommendations for further work produced. The paper concludes by analyzing key areas where the discussed techniques can be used in combination in real world scenarios. Examining the strengths and weaknesses of each method showed clear areas for combined use, as a secondary factor or in a staggered, independent manner.

1 Introduction

Malicious software is an ever growing problem in an increasingly connected world. The global rise of internet usage has presented the opportunity to easily spread malicious software to unsuspecting users. Whilst efforts are made to prevent this spread, it is an issue that is increasing year on year, causing massive financial damage (Alam, et al., 2015). Traditional detection methods use the signatures of previously discovered malware samples, however, this is becoming less effective due to the ease of implementing obfuscation in new malware varients. Whilst traditional methods are being advanced, such as the work done by Kuriakose & P. (2015) into feature ranking and the generalisation of signature detection using support vector model proposed by Wang & Wang (2015), research is being done to identify other potential methods of detection.

This paper will evaluate three classes of detection, based around the concept of detecting the actions of the malware rather than the malware itself; network, system behaviour and user interaction based detection. A critical analysis of the current state of each of these fields will be performed and potential for symbiotic use identified.

2 System Bevaviour

Ding, et al. (2014) suggest improvement upon opcode based detection methods to improve the accuracy against obfuscated malware by utilising control flow information. Opcode detection is based on static methods, giving an inaccurate representation of opcode behaviour as opcode sequences are related by proximity in the decompiled code. By utilizing control flow, the authors aimed to improve detection accuracy by following and running opcode detection for all command paths.

A control flow graph (CFG) was produced by identifying blocks of code that only had a single entry and exit point. A sliding window, n-gram approach was used to extract opcode fragments for analysis. To reduce the amount of noise, information gain and document frequency feature selection methods were used. To prevent all features being selected from the same class the top elements from each set were used for testing. Each feature was then normalised to have unit length to overcome the bias introduced by varying executable size. For training the system three methods were selected, being K-Nearest Neighbour, decision tree and Support Vector Machine.

For testing, two text based methods and the proposed method were used, and each training method used for each. Three testing criteria were used, accuracy, false positive rate and false negative rate. 650 files of each class were used for testing. Heavy use of ‘call’, ‘jmp’ and junk instructions was noted in certain malware samples, used to hinder static opcode detection. For information gain, 3 length n-grams were used, with the top 400 samples selected for further testing (Table 1). During document frequency selection, all opcode 3-grams and variable length, basic block opcode streams were sorted in descending order of frequency. For each sample set the top 200 opcode sequences were selected (Table 2).

Table 1 – Information gain distribution where $n = 3$.

	IG<1.4 (%)	1.4 ≤ IG<1.5 (%)	1.5 ≤ IG<1.8 (%)	IG ≥ 1.8 (%)
Control flow-based features ($n = 3$)	0.866	0.093	0.024	0.017
Text based features (Igor et al., 2013) ($n = 3$)	0.782	0.181	0.035	0.002

Table 23 Information gain distribution where n = 3 (Ding, et al., 2014)

Table 2 – Document frequency distribution for opcode sequences.

	Feature type	Avg. Freq. In malicious files	Avg. Freq. In benign files	Common features
Control flow-based features ($n = 3$)	Malicious	120	56	74
	Benign	112	364	
Text-feature-set1 ($n = 3$)	Malicious	360	305	118
	Benign	253	401	
Text-feature-set2	Malicious	62	38	30
	Benign	42	73	

Table 24 Document frequency distribution for opcode sequences (Ding, et al., 2014)

Testing showed that increasing the number of features did not significantly improve the results. Testing the information gain set showed that control flow based methods have 0.9%-1.7% higher accuracy and 2%-3% lower false positive rates than text based methods. In the document frequency testing it was calculated that control flow methods had 0.5%-2.3% higher accuracy and 1.5-7% lower false positives. Both tests also showed decision tree as the best classification method.

The work done by Ding, et al. (2014) shows excellent promise in improving opcode detection rates, showing a statistically significant advantage over text based methods. By using multiple classification methods they verified the accuracy of their results by showing that it gives an improvement using each method. They made a good effort to eliminate bias at every level of testing, helping to ensure the accuracy of the tests.
Whilst they chose a good selection of malicious software, the same cannot be said of their benign file selection, all obtained from the System32 folder of a Windows XP machine. This lack of variety of files may have caused the feature detection to be skewed. To give an accurate representation, the top downloaded, freely available executables should be used, as this would give a greater diversity of features. This paper leaves several areas open for further research. Further research should be done on a larger scale using a more varied set of files for each class. Additionally the effectiveness of this method can be tested using file types other than those found on Windows, demonstrating the effectiveness on a wider variety of devices.

They state that further research should be done to combine their method with existing system behaviour detection methods, such as the work done by Elhadi, et al. (2014) and Ghiasi, et al. (2015), which respectively cover API call and register based detection. Elhadi, et al. (2014) detail a method to simplify the computational complexity of call graph matching by doing approximate matching on subgraphs. By combining the results of these two methods, it may be possible to use the API call behaviour to better discriminate relevant features. Ghiasi, et al. (2015) expand on this by introducing register values as an additional method of detection. By monitoring registry values, in relation to API calls, they found that highly accurate detection was possible. As the opcode work by Ding, et al. (2014) only considers the opcode being used, this work could be combined to provide more vectors for analysis, improving detection rates.

3 Network Analysis

An additional vector for malware detection is by monitoring network traffic, allowing the prevention of communication with command and control servers, and the identification of infected hosts. Malware producers however are implementing more advanced systems to combat this. Lee & Lee (2014) focused on the detection of malicious DNS queries. They achieved this by developing Graph-based Malware Activity Detection (GMAD). GMAD was created to counter the temporal and spatial evasion techniques used by malware during DNS queries. They began by producing a graph of the DNS queries, which they then split into domain clusters using cutting edges at various values of client sharing ration (CSR). Existing domain blacklists were then compared with the graph and domain clusters, with those containing a blacklisted address classified as malicious (Figure 1).

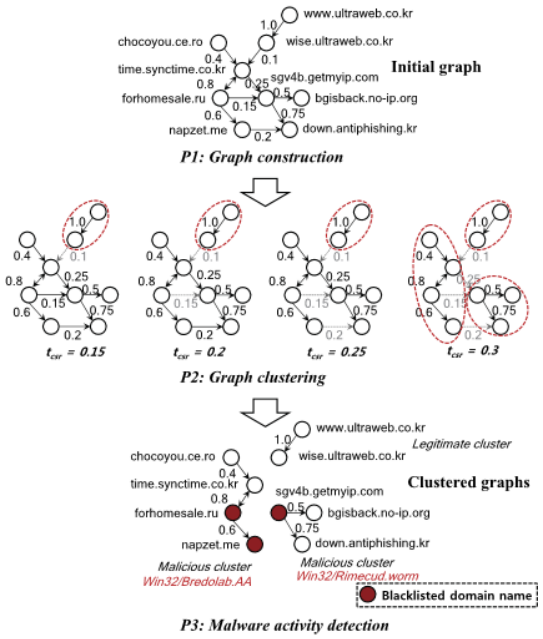


Figure 17 A real-case example of detection using GMAD process. (Lee & Lee, 2014)

For testing, a large dataset of DNS data captured from two separate ISPs was used. Their results showed a 84-95% level of precision and a 0.05-0.26% false positive rate. As analysis of the false positive data showed that many were incorrectly identified due to a low client count for that domain, an additional test was done replacing the CSR based edge cutting with client set size (CSS). This showed that as CSS was increased the number of detected domains decreased, however the amount of true positive detections increased. Comparison between this and previous testing showed that CSR and CSS methods were better at detection under different circumstances, suggesting that both should be used in tandem.

The research details an effective way to detect previously unknown malicious domains by relating them to known malicious domains. They made a good effort to verify the validity of their results by doing manual analysis of identified malicious domains. This however could have been improved upon by making an attempt to calculate false negative rates. Whilst the dataset was too large to do complete manual analysis, with around 2 million domains detected as not malicious, a smaller dataset could have been selected to gather preliminary data. This could have been accomplished by choosing a set of clients with known infections, either by random selection or by classifying their importance based on number of DNS queries or malicious queries. Additionally, attempts could have been made to automate this process by obtaining site statistic data about popular sites, and using this to identify truly benign domains. This would reduce the data that would need to be processed manually, allowing for a more indepth analysis.

In the research done by Menten, et al. (2011) they applied their method in such a way that it could be applied on intermediary network devices, allowing for wide scale detection across the entire network. It would be useful to implement this technique in a GMAD based detection system. Depending on the specifics of network design, port access transition (PAT) may be configured in the network. As PAT connections all use the same IP address, only differentiating by port number, detection of the infected host would include the comparison of GMAD and PAT logs. To prevent this situation GMAD could be run independantly on each network segment, allowing for quick and accurate identification of infected hosts. In large networks this would also be an advantage to reduce the work each GMAD device would need to do.

Shabtai, et al. (2014) use the comparison of the network usage of an application running on a device to a known good version obtained from the software source. This work could be used to give additional data points to use for classification of malicious domains, as if network usage changes in a malicious way, the domains used can be classified as malicious. This would add to the blacklists used in GMAD classification introducing more rules for accurate detection.

4 User Interaction

The work done by Elish, et al. (2015) identifies sensitive API calls and compares them to user actions. A TriggerMetric, showing the likelihood that this action is caused by user interaction, is created and can be used to validate the actions of an unknown program. Through testing they showed that this method had a low false negative rate (2.1%) and a low false positive rate (2.0%), on a sample of 2684 benign and 1433 malicious applications, also detecting many samples not found by existing scanners.

A data dependence graph is created, pruned by reachability analysis, before using backward depth-first search to identify valid dependence paths. These paths show the connection between a user action and a sensitive API call, which is used with the total calls to the operation to calculate the percentage of valid calls. These are then used to create a normalised distribution of valid calls (DPVC). Initial classification of a sample is done by comparing the percentage of user triggered actions to a pre-set

variable, 75%, which gave the best balance between false positives and false negatives. The DPVC is then compared to the average DPVC of known malware samples using weighted cosine similarity, and compared with a threshold (0.8). If both of these tests identify the sample as benign it is removed from further testing.

The initial test identified 92.5% of the malicious files, with the second identifying an additional 5.4%. Testing on the benign set showed 8.94% of the files as being malicious. These were subsequently scanned using available scanning tools which returned results for 137 out of 240 samples. Of the remaining samples 21 were selected for manual analysis, revealing 11 to contain malicious code.

Their testing showed that malicious samples had a lower percentage of valid API calls, with 479 out of 1433 having 0%. It was found that their method did not detect phishing applications as these mimic valid applications, giving a high percentage of valid calls. Additionally it was found that when malware is repackaged into an existing application, false negatives could be caused due to the high amount of benign behaviour within the application. They suggest that more extensive testing could be done by using call graph analysis and clustering to identify loosely connected components that should be tested separately, and showed promising initial results on 4 samples.

The authors of this paper have shown that their method can accurately detect malware based on how actions relate to user interaction. This method has great significance, as shown by its ability to detect malicious samples undetected by conventional scanners. In their testing they took great care to remove bias by automating the process, and documented thoroughly allowing for the tests to be repeated. A good sample size was used for testing including a wide variety of benign and malicious samples from multiple different authors, further ensuring their reliability.

Whilst this research has the greatest significance in the mobile malware detection, research should be done to see if the method can be used on other operating systems. As many more programs in a desktop environment use background system calls it may be necessary to first classify programs into function groups, representing its main function, allowing for comparisons to be made with that group, giving a more accurate estimation of the trustworthiness of the system calls. This additional technique would also benefit mobile malware detection for the same reason. More in-depth detection could be achieved by analysing each call individually, for example an API call to send a text message is likely to be user triggered in benign software and not in malicious. This could potentially identify the relevance of each API call, which could be weighed in the calculations allowing for a lower false positive and false negative rate.

To improve on the speed of this method, permission based detection, such as the work of Talha, et al. (2015), could be used to refine API searches to be limited to the scope of those accessible with certain permissions. Both methods could be used in tandem to produce more accurate results. This would be achieved by using the work of Talha, et al. (2015) to produce the likelihood that a certain permission is used by malware, then using each permission as an individual group which would be fed into the method designed by Elish, et al. (2015). When producing an overall maliciousness score the resulting group scores would be weighted by the maliciousness score of the group, before being averaged into a single result. Sheen, et al. (2015) found that when permissions and API behaviour were used with collaborative decision fusion, precision was raised from 83~95% to 98.31~98.91%.

5 Conclusions

As advanced malware uses many techniques to prevent detection through conventional means, new methods must be developed. This paper looked at three areas of indirect detection; system, network, and user behaviour. Analysis of the papers show that the new methods have a statistical improvement over conventional methods, as well as overcoming antidetection techniques. Future research should be done to identify the potential to use these techniques in tandem. As each method has its own pitfalls, such as code analysis being impossible if the malware cannot be decompiled, research should be done into what overlap between the detection rates of different malware variants and features with the goal of identifying areas where the results could be combined to produce a more accurate method. This would have the additional benefit of reducing the pitfalls associated with each technique by introducing more detection vectors.

In the real world, malware detection must be a quick process. As detection should be done before a file is executed, the detection method should not introduce an extended waiting period before the opening of a file. This all but excludes opcode based detection from real time scanners. However it is possible for it to be implemented in a delayed fashion. Standard real time detection could be used for a first pass of a file, after

which it could be allowed to open. During this time opcode based analysis could be performed and the application shut down if malicious behaviour is detected. This is not ideal due to the potential for the malware to interfere with the detection process, but it would allow for a more in-depth analysis at run time.

Analysis shows that GMAD detection was not resource intensive and worked in a fast manner, even for a large data set. This makes it a perfect candidate for real time protection and embedding on network devices. This could be implemented with a control server within the network that could gather information and find infected hosts, including times and order of detection. This information could be used with other sources to help identify the source of the infection, greatly reducing investigation time.

The work of Elish, et al. (2015) is more limited in its scope due to its reliance on the features of sandboxed mobile environments. However the techniques used could be used on other platforms with minimal changes. Web browsers for example run JavaScript in a sandbox, and this technique could be used directly for web browser based detection. Additionally they are similar behaviours that can be seen in Windows malware to the API calls in mobile malware. An obvious example would be attempting to access Ring 0, but other milder actions could be found. To facilitate speed of detection, a server based model should be used for this technique allowing for a client to only calculate the checksum of a unknown of the file, and compare it to the server in most cases. As the client would not run the unknown file before classification there is reduced chance of an infected file interfering with detection operations.

References

Alam, S., Horspool, R., Traore, I. & Sogukpinar, I., 2015. 'A framework for metamorphic malware analysis and real-time detection'. *Computers & Security,* Volume 48, pp. 212-233.

Ding, Y., Dai, W.., Yan, S.& Zhang, Y., 2014. 'Control flow-based opcode behavior analysis for Malware detection'. *Computers & Security,* Volume 44, pp. 65-74.

Elhadi, A. A. E., Maarof, M. A., Barry, B. I. & Hamza, H., 2014. 'Enhancing the detection of metamorphic malware using call graphs'. *Computers & Security,* Volume 46, pp. 62-78.

Elish, K. O., Shu, X., Yao, D., Ryder, B. G. & Jiang, X., 2015. 'Profiling user-trigger dependence for Android malware detection.'. *Computers & Security,* Volume 49, pp. 255-273.

Ghiasi, M., Sami, A. & Salehi, Z., 2015. 'Dynamic VSA: a framework for malware detection based on register contents'. *Engineering Applications of Artificial Intelligence,* Volume 44, pp. 111-122.

Kuriakose, J. & Vinod, P., 2015. 'Unknown Metamorphic Malware Detection: Modelling with Fewer Relevant Features and Robust Feature Selection Techniques.'. *IAENG International Journal of Computer Science,* 42(2), pp. 81-93.

Lee, J. & Lee, H., 2014. 'GMAD: Graph-based Malware Activity Detection by DNS traffic analysis'. *Computer Communications,* Volume 49, pp. 33-47.

Menten, L. E., Chen, A. & Stiliadis, D., 2011. 'NoBot: Embedded malware detection for endpoint devices'. *Bell Labs Technical Journal,* 16(1), pp. 155-170.

Shabtai, A., Tenenboim-Chekina, L., Mimran, D., Rokach, L., Shapira, B. & Elovici, Y., 2014. 'Mobile malware detection through analysis of deviations in application network behavior'. *Computers & Security,* Volume 43, pp. 1-18.

Sheen, S., Anitha, R. & Natarajan, V., 2015. 'Android based malware detection using a multifeature collaborative decision fusion approach.'. *Neurocomputing,* 151(2), pp. 905-912.

Talha, K. A., Alper, D. I. & Aydin, C., 2015. 'APK Auditor: Permission-based Android malware detection system.'. *Digital Investigation,* Volume 13, pp. 1-14.

Wang, P. & Wang, Y.-S., 2015. 'Malware behavioural detection and vaccine development by using a support vector model classifier.'. *Journal of Computer & System Sciences,* 81(6), pp. 1012-1026.

Evaluation of Current Security Techniques for Online Banking Transactions

Annah Vickerman

Abstract

Online banking is one of the technologies that are important in supporting financial services. Various technologies have been developed to determine improved performance and security for online banking transactions. This research paper analyses, compares and evaluates techniques that are used to counter fraud in online banking transactions. It presents the use of biometric fingerprint recognition technique, Identity Based mediated RSA (IB-mRSA) technique, Neuro-Fuzzy technique and personalised security mechanism in e- banking. Upon the evaluation of these techniques, recommendations are drawn with respect to the most current suitable technique to secure online banking transactions.

1 Introduction

Online banking is one of the most recent conventional methods of e-commerce which has brought many benefits, but due to its infant stages of development, has been prone to various disadvantages in the form of security breaches and fraud. Sharma and Lenka (2013) indicated that, banks have made it a priority to have authentication control over online banking transactions so as to reduce risk of infiltration and to offer reassurance and safety to its customers. Online banking has advanced to an extent that financial transactions worth millions can be processed in seconds, without having to visit a bank branch (Jassal and Sehgal 2013).Since the introduction of this service, it has seen an exponential influx of clientele due to its convenience. Usman and Shah (2013) stresses that, though more people are embracing this service and are exposed to the flow of information from the internet, privacy and security go hand in hand and they are crucial to the growth of electronic transactions. Consequently, there has been a dire need for the protection of these services as they have proved to have flaws. It has been observed by Dharmendra Chahar and Niranjanamurthy (2013) that the greatest obstacle facing online systems is that the developers are focused on fraud identification and not fraud prevention. A major concern for financial service organisations is to have a safe and secure environment for their computer technology (Nwogu 2014).

Research has been done to address the above mentioned security concerns. Tassabehji and Kamala (2012) proposes a technique that uses biometric fingerprint recognition for online banking transactions. Darwish and Hassan (2012) proposes a modified technique claimed to authenticate clients, increase security and preventing attacks in e-banking transactions. Barraclough et al. (2013) proposes a scheme that detects phishing sites and provide high accuracy in real-time. Hamidi et al. (2013) carried out research on a method referred to as personalised security mechanism in e-banking, as a way of improving security level. The above mentioned proposed solutions look into these concerns regarding the performance and security by ensuring privacy, accuracy, robustness and effectiveness of the e-banking transactions.

This research paper consists of two sections: the first section will discuss and evaluate current security techniques which have been proposed by various researchers. Lastly; the conclusions, which will provide a summary of the major themes of the paper.

2 Current Security Techniques

This section will look into the current security techniques used for online banking transactions proposed by different researchers. A variety of research papers have been analysed and evaluated, with particular interest as to how the method operates, how testing was conducted and validity of the experiments.

2.1 Application of Biometric Fingerprint Recognition for Online Banking Transactions

Tassabehji and Kamala (2012) conducted a research on biometrics for online banking which focuses on security concerns. Tassabehji and Kamala (2012) articulated that the biometric banking system was implemented focusing on the use of biometric fingerprint recognition to establish each user based on public/private key encryption protocols. The study shows that users considered fingerprint biometrics to be the uncomplicated to use, most secure and most preferred as compared to other biometric technologies.

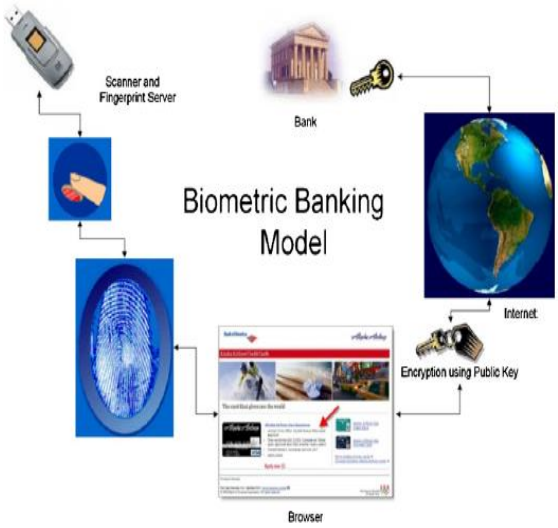


Figure 1 A Schematic diagram of proposed biometric banking system Tassabehji and Kamala (2012)

The diagram illustrates the representations of b-banking system to measure the performance of the fingerprint approach. Tassabehji and Kamala (2012) indicated that, with users accessing their bank online, they place their fingers to be scanned for authentication. A web browser in the computer will be launched once the authentication is recognized. However, such browser cannot accept URL's and this hinders any potential interfering with web addresses that may forward the internet connection to a different address. The key will log-in the authenticated user and will then organize a secured connection with the correct bank. This is to show that whenever a wrong fingerprint is placed, the key will automatically lock itself and users will have to go back to the bank for re-confirmation (Tassebehji and Kamala 2012).

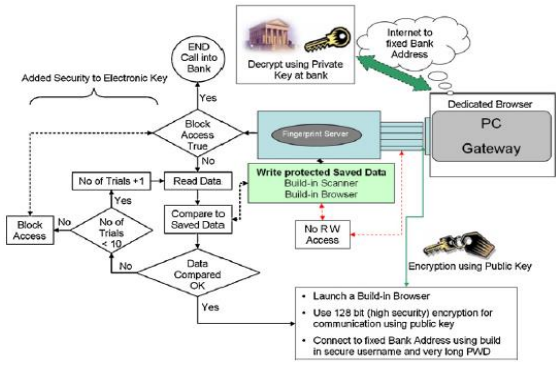


Figure 2 Displays a proposed control method Tassabehji and Kamala (2012)

An experiment was conducted to test the usability of the proposed system. Tassabehji and Kamala (2012) delivered that the System Usability Scale (SUS) was established since it was considered to be a reasonable, effective and very good tool for assessing the usability of the entire range of systems. Tassabehji and Kamala (2012) produced a representation that aims at interpreting a combination measure of the overall system usability that is being studied for each of the statements and the mean was calculated. Tassabehji and Kamala (2012) indicated that, the SUS scores were calculated according to Brooke's instructions. Alongside the score, the authors also produced an adjective rating scale that associates with the score, hence it adds qualitative feature to it. According to Tassabehji and Kamala (2012), results show that fingerprint recognition is considered to be the most suitable and favoured biometric technology for online use as represented in figure 3 below.

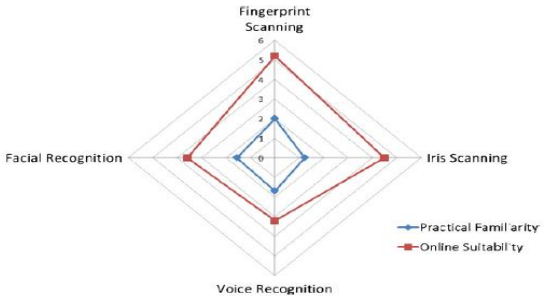


Figure 3 Shows user experience of biometrics Tassabehji and Kamala (2012)

In conclusion, the authors developed a biometric system for verifying online banking and used the proved System Usability Scale (SUS) to evaluate its usefulness from the user's viewpoint. Tassabehji and Kamala (2012) disclosed that upon the demonstration, users seemed pleased about a biometric system. It is clear that the authors of the proposed method did not consider different components when evaluating the system. They did not show the formal process that helps verification and authentication. The security method is still left questionable as experiments need to be conducted to get full prove of whether this method can really provide security as said by the authors. Experiments of all biometric technologies need to be conducted, together with their results so that when they claim that the fingerprint technology is considered to be the most suitable, they also provide evidence for comparison. This is not enough as it still leaves one in doubt. According to what is presented by the authors, they claimed that the scale is fast and easy to administer score. However, they mentioned that if the numeric

score is incomplete, it is hard to read qualitatively sometimes. This means that the possibilities of obtaining wrong results are high because lack of accuracy is clearly shown.

2.2 Identity Based mediated Rivest-Shamir-Adleman (IB-mRSA)

Darwish and Hassan (2012) proposed a method known as an Identity Based mediated RSA (IB-mRSA) in combination with one-time ID technique which aims at authenticating clients for internet banking transactions. The model emphasises on expanding security and blocking attacks. Darwish and Hassan (2012) articulated that, private keys are split among the client and the trusted server, there are no cheats in between as one-time ID can only be used once and both parties must always be involved in each signature. Therefore, neither can decode message without other's presence, since they depend on each other. The scheme depends on three (3) modules that guarantee that DoS attacks are avoided and leakage of user's identity is obviated, namely Certificate Authority (CA) module, Security Mediator (SEM) module and One-time ID module (Darwish and Hassan 2012).

Certificate Authority (CA) module, a component that involves certificate issuance and revocation interface. A client can only use their private key only if they got an exact token message from the server after decrypting a message. The CA server is kept away from the internet in order to avoid unauthorized entries (Darwish and Hassan 2012). Security Mediator (SEM) module involves verification and partial signature processes. One-time ID module, allows the use of one-time ID only once and making it complicated for the attacker to find who is interacting even when he overhears the ID (Darwish and Hassan 2012).

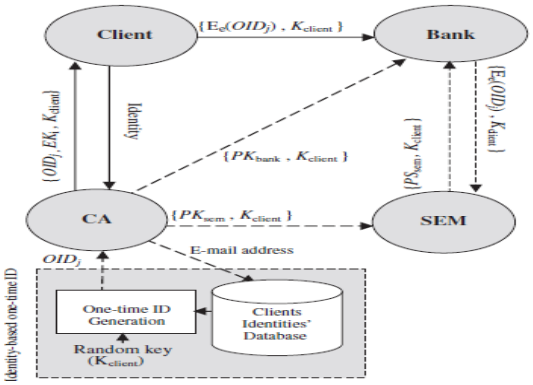


Figure 4 Displays a proposed online transaction security model Darwish and Hassan (2012)

By ensuring that the proposed scheme attains what it is designed for, three (3) algorithms were discovered, namely key generation, signing and lastly verifying.

Key generation takes account of all key set up and generates a different set for each client. Signing, CA produces a list of clients' identity and stores it in encrypted database for security purposes and same identity is used to create each client one-time ID on a later stage (Darwish and Hassan 2012). Verifying, SEM ensures that client's identity is valid as requested message is signed with its private key to produce a partial signature and provides feedback to the bank. However, if the client's unauthorized, SEM displays an error.

To understand how exactly the model is operating, experiment was conducted on key generation, signing and verifying algorithms. Darwish and Hassan (2012) stated that Identity Based mediated RSA process was set up on a workstation that had an Intel Pentium 4 processors with a speed of 1.2GHz, 1GB of RAM. The authors also indicated that, all test machines executed Windows NT version with over a 100Mbps Ethernet LAN in a lab environment. The author's coding on this model was component-based. To further describe experimental environment, the authors stated that, roundtrip latency between two machines was measured at 5 milliseconds (ms) and maximum constant throughput of network connection at 7Mbps.

	SAS model	Proposed model
Client's computation	7.34	3.3
Signing computation	6.9	37.5
Verifying computation	7.8	15.4

Table 1 Experimental results showing comparison of two models (ms) Darwish and Hassan (2012)

Table 1 shows that "the proposed model offers a substantial computational advantage over SAS with respect to client's computation" (Darwish and Hassan 2012).

To conclude this, upon performance of the model, the authors provided every detail in relation to resources they used to carry out their test. The results were also presented well, even though they tested their model against only one model being SAS model, this may not be convincing enough. There are many other models they could have used to provide reasonable results and

proving that indeed their model has better client's computation like they claim. Comparing with one model makes it difficult to compare again with other models as one cannot tell how the model performs. Therefore, more experiments need to be done on other platforms in terms of performance and security so that a concrete conclusion can be drawn.

2.3 Application of Neuro-Fuzzy for Online Banking Transactions

Barraclough et al. (2013) proposed Neuro-Fuzzy scheme alongside five (5) inputs which aims at detecting phishing sites with high accuracy in real-time, as well as protecting clients when performing online transactions. Neuro-Fuzzy is a combination of Fuzzy Logic and Neural Network. Neuro-Fuzzy has broad approximations to perform Fuzzy IF…THEN rules, and Fuzzy Logic includes reasoning on an advanced level where linguistic information from experts department is used, whereas Neural Network works properly when dealing with unprocessed data (Barraclough et al. 2013). The scheme has five (5) inputs that are exclusively ideal of phishing techniques and strategies, namely legitimate site rules, user-behavior profile, PhishTank, user-specific sites and lastly Pop-up from Emails.

Legitimate site rules, a summary of law covering phishing threats. User-behavior profile, a record of users' behavior when communicating with phishing and legitimate sites (Barraclough et al. 2013). PhishTank, a community site conducted by Open Domain Names where community experts ensure that suspected websites are confirmed and voted as phish. User-specific sites, mandatory requirements between the user and online transaction service providers (Barraclough et al. 2013). Pop-ups from emails, phrases that phishers regularly use, that show up on screen.

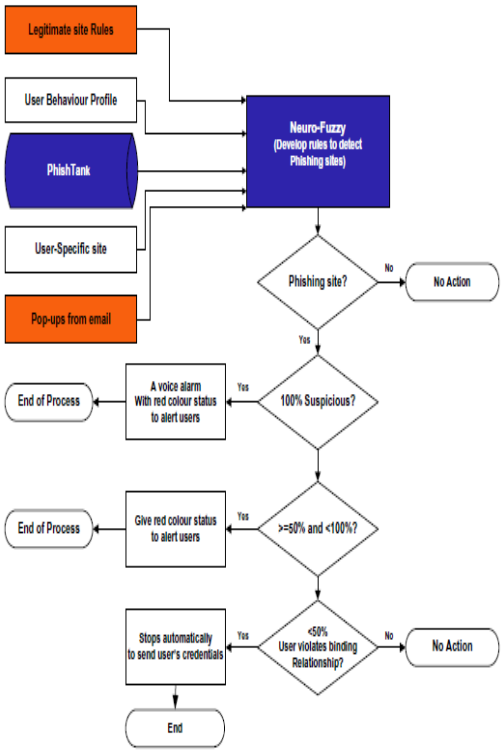


Figure 5 A diagram of intelligent phishing detection system Barraclough et al. (2013)

To understand how the scheme is performing, experiment was conducted using 2-fold-cross validation method for training and testing the proposed scheme regarding the accuracy and robustness. Barraclough et al. (2013) stated that, this method involves breaking down data set into training and testing due to its efficiency for regular datasets. Table following shows the method that was used for the experiment:

Results summary Five inputs	Training error	Testing error	Test error%	Test average error%	Test accuracy
Legitimate site rules	0.011569 0.012057	0.012018 0.012057	1.2%	1.5%	98.5%
User behavior profile	0.01755 0.011736	0.014552 0.01939	1.7%		
PhishTank	0.016882 0.016882	0.016879 0.016879	1.7%		
User-specific site	0.01134 0.011168	0.011346 0.011151	1.1%		
Pop-Ups from Email	0.020452 0.0122	0.016142 0.023276	2.0%		
Average error	0.0141836	0.015429			

Table 2 Displays 2 Fold cross-validation method Barraclough et al. (2013)

Barraclough et al. (2013) articulated that each of the five (5) inputs datasets are divided into training and validation data, whereby training data is discovered through the input layer in no particular order, which is when training starts. The back-propagation corrects

any errors that may be encountered. The process endures neural network and the inference engine decides on which is attained by reasoning together with rules given in the rule base. It then reaches the output layer. This process is repeated in such a way that each data set is used only once. In completion of training, the same process is applied to testing. The accurate measurement processed at verification stage is the final output where systems can be distinguished. This process is done again for each input so that data set for testing is used once. To obtain the average error rates used regarding performance, total results obtained from all inputs are calculated then divided by number of inputs (Barraclough et al. 2013).

The authors developed a fuzzy inference structure that comprises of five (5) functional components, namely layer 1 being the input, layer 2 the fuzzification, layer 3 the rule layer, layer 4 the normalization and lastly layer 5 which is the defuzzification. An algorithm that gets data from one file, where it is read and used to create the fuzzy IF…THEN rules was also developed (Barraclough et al. 2013).

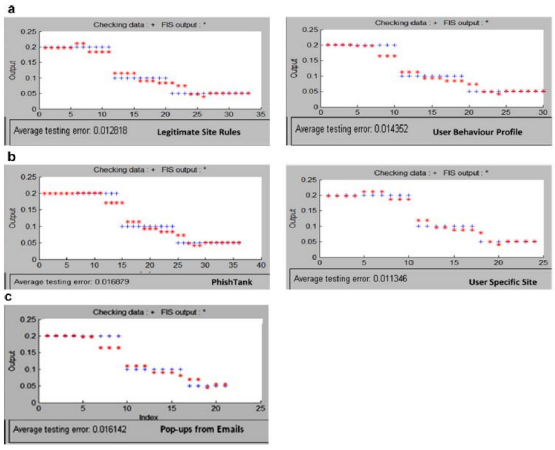


Figure 6 Displays test error results for each of the five inputs Barraclough et al. (2013)

Concluding this, more experiments need to be done showing the performance of this scheme against other existing proposed schemes. Since the authors claimed that their scheme is more accurate, they need to also show results that indeed this approach establish a significant improvement, like Darwish and Hassan (2012) did. Even though they compared against one method which is still not enough. It is not convincing enough for the authors to merely mention that experiment was repeated, however they did not specify how many times the test was repeated, to ensure accuracy and reliability as far as security is concerned.

2.4 Personalised security

Hamidi et al. (2013) proposed a method referred to as personalised security in which online banking transactions are secured and prevented against fraud over a cloud environment. The purpose of the scheme is to provide a safe environment for applying user-defined policy for individuals, companies and government organizations, by focusing on how to boost security level. Hamidi et al. (2013) articulated that, large companies that deal with large income need to protect their accounts with best security. The process of card security was used to show transactions of the system. The use of phone numbers as part of transaction confirmation was also considered (Hamidi et al. 2013).

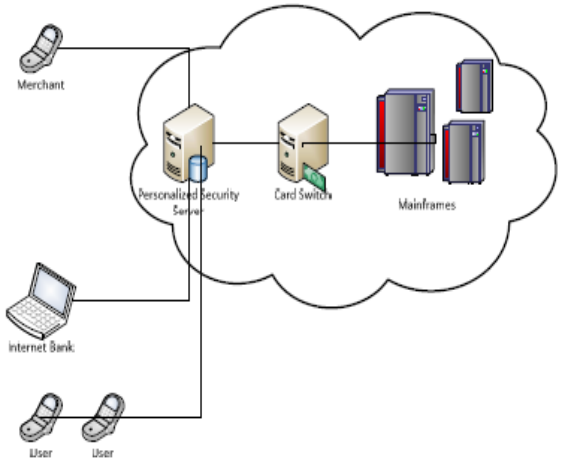


Figure 7 Shows an overview of the proposed scheme Hamidi et al. (2013)

An experiment was conducted to test the performance of the proposed mechanism. Hamidi et al. (2013) stated that, Flask Architecture was employed as a result for card transactions in cloud environment. This method deals with improving security and applying essential user access control. Figures 8 and 9 below illustrate registration process of the system and placing of transactions respectively.

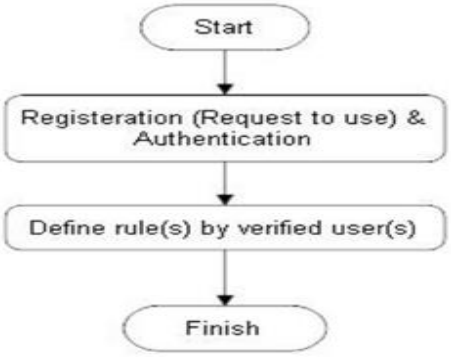


Figure 8 Displays registration in the system Hamidi et al. (2013)

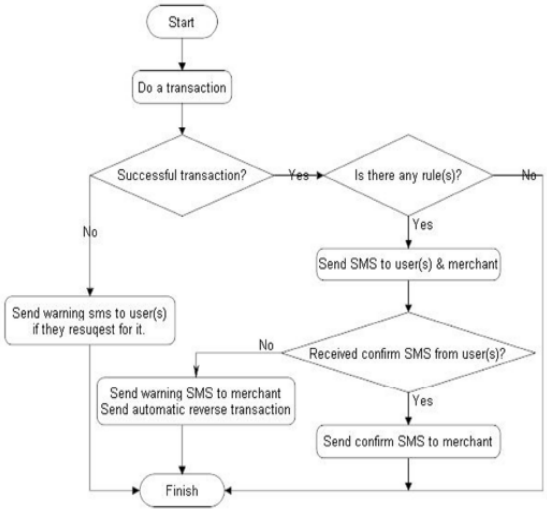


Figure 9 Displays the placing of transactions Hamidi et al. (2013)

The authors also provided policies which were used to checkup different types of transactions; the approved, rejected and the fraud detection transactions (Hamidi et al. 2013). Figures 10 and 11 below illustrate 50% and 80% of effectiveness of the personalised security schema respectively.

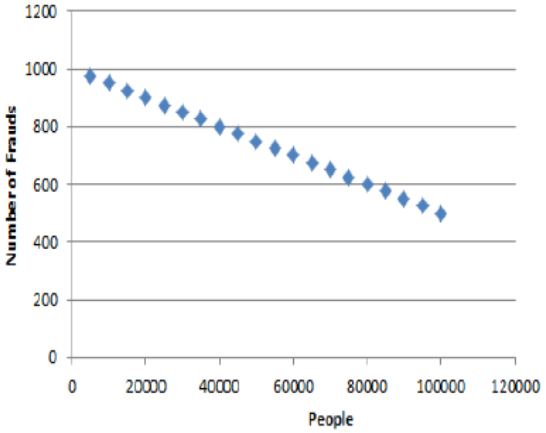


Figure 10 Showing 50% effectiveness Hamidi et al. (2013)

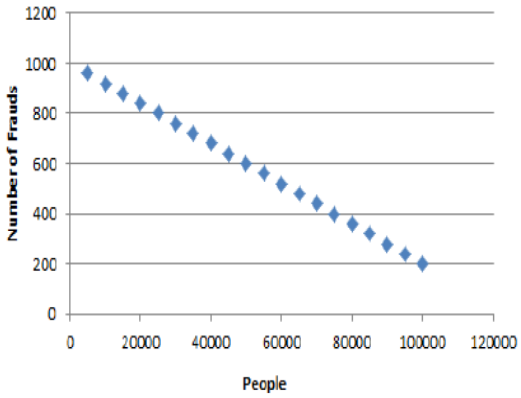


Figure 11 Showing 80% effectiveness Hamidi et al. (2013)

The authors claimed to provide a safe environment in large institutions. Therefore, experiments need to be conducted and results proving that indeed transactions are secured against fraud. The scheme did not show any provision of best security for the accounts as claimed by authors. Therefore, more experiments should be done to show how best of security can be provided. The authors also mentioned that clients should wait for an SMS or Email confirmation. The client cannot rely on an SMS or Email because if the network is down, they will not be alerted on time. The authors did not compare their proposed scheme with other schemes. This is not enough since there is no prove that indeed fraud is reduced.

3 Conclusions

In this research paper, the most current papers on security techniques for online banking transactions have

been thoroughly analyzed and evaluated. Evaluation was based on the performance and security of the techniques. The algorithms were presented well by the schemes and illustrations were made on how the schemes perform, even though the security aspect of the research was somehow forgotten. With the evaluation of this paper, more experiments need to be carried out regarding the application of security algorithms in order to establish what the authors have said. With regards to the performance of the techniques evaluated in this paper, research done by Barraclough et al. (2013) is the only one that presented the results of all components of the scheme. IB-mRSA model by Darwish and Hassan (2012) shows that the model has better client's computation time as compared to the SAS model. For the research done by Barraclough et al. (2013), it is hard to come up with a concrete conclusion since the scheme is not tested against other schemes. With that being said, experiments and results of the proposed scheme against other schemes is needed to show comparison and prove reliability of the scheme.

On the research evaluation of this paper, performance tests have been conducted with their resources and versions used. Algorithms were also created to show how different schemes can be secured, but still it is not enough. With no experiments showing better security, it is difficult to conclude the preferred research. However, with the presented different schemes, research by Barraclough et al. (2013) and Darwish and Hassan (2012) could be combined and come up with a strong security as well as performance needed for online banking transactions.

References

Ahmad Kabir Usman and Mahmood Hussain Shah, 2013. 'Critical success factors for preventing e-banking fraud.' *Journal of internet Banking and Commerce* , 18(2), pp. 1-9.

Anand Sharma and Lenka S. K., 2013. 'Authentication in online Banking Systems through Quantum Cryptography.' *International Journal of Engineering and Technology,* 5(3), pp. 2696- 2700.

Barraclough P. A., Hossain M. A., Tahir M. A., Sexton G., Aslam N., 2013. 'Intelligent Phishing Detection and Protection Scheme for Online Transactions.' *International Journal of Experts Systems with Applications,* 40, pp. 4697-4706.

Dharmendra Chahar and Niranjanamurthy M., 2013. 'The Study of E-commerce Security Issues and Solutions.' *International Journal of Advanced Research in Computer and Communication Engineering,* 2(7), pp. 1-12.

Emeka Reginald Nwogu, 2014. 'Improving the security of the Internet Banking System using Three-Level Security Implementation.' *International Journal of Computer Science and Information Technology & Security,* 4(6), pp. 167-174.

Nayer A. Hamidi, Mahdi Rahimi G. K., Alireza Nafarieh, Ali Hamidi, Bill Robertson, 2013. 'Personalized Security Approaches in e-Banking Employing Flash Architecture over Cloud Environment.' *The 4th International Conference on Emerging Ubiquitous Systems and Pervasive Networks,* 21, pp. 18-24.

Rajpreet Kaur Jassal and Ravinder Kumar Sehgal, 2013. 'Online Banking Security Flaws: A Study.' *International Journal of Advanced Research in Computer Science and Software Engineering* , 3(8), pp. 1016-1020.

Rana Tassabehji and Mumtaz A. Kamala, 2012. 'Evaluating Biometrics for online Banking: The Case for Usability.' *International Journal of Information Management* , 32, pp. 489-494.

Saad M. Darwish and Ahmed M. Hassan, 2012. 'A Model to Authenticate Requests for Online Banking Transactions.' *Alexandria Engineering Journal,* 51, pp. 185-191.