Selected Computing Research Papers

Volume 6

June 2017

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

Critical Analysis of Online Transaction Verification Technologies in Financial Industries

Baboni Mmaopinkie Beleng

Abstract

The emergence of online data breaches and fraud cases has led to implementation of online verification technologies in financial banks. Many of this methods aim to achieve security of protected sensitive data, promote integrity of data and replay attacks. This paper considers four of those technologies, which are; biometric-Kerberos authentication, fingerprint recognition authentication, online verification signature and keystroke dynamic method. It also gives a recommendation on future work to improvise the approaches already presented by other scholars.

1 Introduction

Online banking has caused gigantic change in financial firms' security practices and services. Despite the advancement in online verification technologies, it is still significant that financial industries face vast growing cases in which unauthorised parties that are sophisticated enough to hack advanced levels of security have tempered with customers' accounts. Carvalho et al (2015) expressed an interesting view that statistics shows that Payment Services is the most targeted industry sector in cybercrime and Phishing attacks.

Financial industries understand and address the potential dangers of account intrusions, whereby an unapproved individual snips available assets and misuses the account to manipulate the market. This also causes financial industries to incur extra costs to lessen the situation and refund clients. Telang (2015) enunciates that firms frequently suffer some cost in distinguishing and tidying up breaches. Costante et al (2016) articulates that data breaches can affect customers' perception towards company's image by destroying it reputation especially in cases where sensitive data is breached. Fraud and digital data breach usually occurs in scenarios where account owners are shipping online, paying bills through an internet and various e-commerce activities. Kurnia et al (2010) discovered that effective fraud deterrence techniques are essential to shape the society's trust on online banking consumer oriented EC activities such as mobile banking, phone banking and internet payment system.

Nguyen and Dang (2015) proposed biometric-Kerberos authentication technique which assures shared verification between the client and server over the uncertain network thus the user and server can both confirm each other's identity. However, biometric-Kerberos is vulnerable to Password guessing attacks especially when users' passwords are simple. Ohana et al (2013) suggested a method used for online banking named as biometric fingerprint recognition for online transactions. Challenges often experienced with this technique are factors such as cost of implementation thus other financial industries might find it too expensive to install in their online banking systems. Alshanketi et al (2016) suggested Keystroke dynamic method that is mostly used for mobile biometric authentication that utilises dwell time and flight time features. Challenges often experienced with this method are lack of typo handling and inaccuracy.

Li et al (2016) proposed online verification signature that is an electronic based signature that make use of dynamic information such as velocity, pressure of writing the signature and acceleration. Though online verification signature are more reliable than handwritten ones, if professional hackers manage to collect that dynamic information about the signature then they can use it to temper against the client's private information.

This paper evaluates the online transaction verification technologies used by financial industries. Section 2 of this paper describes and summarises the proposed online transaction verification technologies, their experiments, algorithms, test results and recommendations the scholars propose to improvise the approaches already presented by other scholars. The last section present concluding remarks and summarise our future work.

2 Current online transaction verification technologies

This section will evaluate four technologies used for online transaction verification in financial industries, together with their experiments and results.

2.1 Online Verification Signature

Hanmandlu et al (2015) describes the implementation of online signature verification method with the entropy function that give probabilistic uncertainty. However, the entropy function can be used to determine possiblistic uncertainty in the measurements of online signatures to give the best results. The following is the algorithm in the possiblistic domain that mainly focuses around the fuzzy set.

$$E_H = \sum_{i=1}^{M} x_i \mu_i$$

Equation 1: Hanman-Anirban Entropy (H) (Hanmandlu et al. 2015).

Hanmandlu et al (2015) continues to clarify that µi in this equation is the membership function constructed from x. This entropy signifies uncertainty measurements values through membership. EH presents the information while xiµi symbolizes information set (set of information values). Hanmandlu et al (2015) points out each element in the fuzzy set is multiplied by the membership function to give us the information feature.

Hanmandlu et al (2015) articulates that the information set based feature given by the above equation should be employed with IPC (inner product classifier) and Support Vector Machine (SVM) classifier. This information set based feature will be used to verify the online signatures.

The execution of the employed classifiers is ascertained by means of the information set based features of test and training features. The following are results after the performance of classifiers are established using information set based features for signatures of 40 users for IPC and SVM in a polynomial kernel.

Training: Test Ratio	IPC	SVM(PR-Tools)		
		Poly1	Poly2	Poly3
15:5	100	93	94	94.5
10:5	97.5	68	63.5	63.5

Table1: Recognition Rates with Information set based features for the Signature of 40 users (Hanmandlu et al. 2015).

Hanmandlu et al (2015) explained that the results above show that IPC stretches the highest maximum recognition rate whereas SVM records lower recognition rates as compared to those of IPC. As seen from Table 1 that IPC as classifier yields a maximum recognition rate of up to 100% on the given information set based features while SVM yield 94.5% for the training ratio of 15:5 in Polynomial of degree 3.

However, the author recommends that Adaboost and Random Forest method could be used instead of SVM since SVM classifier can make a high error rate of 15.01% while random forest's error rate is 2.78% and that of Adaboost is 2.375% as articulated in Li et al (2016). Therefore, Adaboost can present satisfactory results in verification as compared to SVM.

In conclusion, Hanmandlu et al (2015) enunciates that online signature verification is endeavoured by extracting information set based features which result from representing the uncertainty in the measurement values in a fuzzy set by the entropy function.

2.2 Biometric Fingerprint Recognition

Fingerprint recognition is a biometric based authentication technique with automatic process of validating a match between two human fingerprints. Tassabehji and Kamala (2012) expressed an interesting view that a biometric banking system was established with an aim to use this method to set up every client based on public/ private key encryption protocols. The following Figure 1 depicts how the banking system make use of biometric fingerprint recognition technology to verify clients making transactions online.

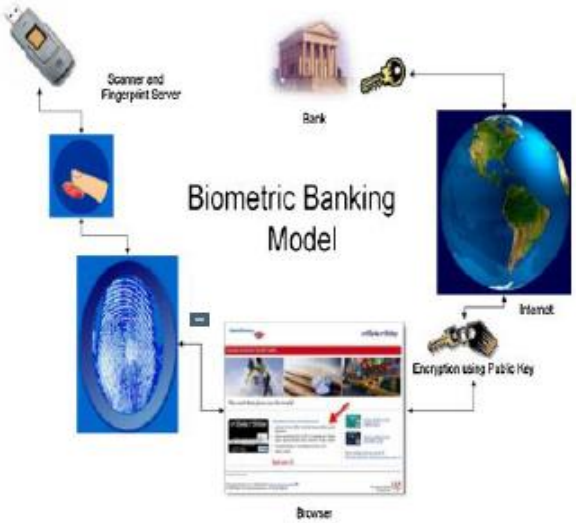

Figure 1: Biometric Fingerprint Recognition banking system (Tassabehji and Kamala 2012).

Tassabehji and Kamala (2012) explained that for client to access bank services online they are required

to place their fingers to capture its print for authentication. Once the authentication process is successful, the bank web page will be launched. The authenticated user will then be allowed to login and a secure connection to the correct bank will be established. Tassabehji and Kamala (2012) articulates that if an incorrect fingerprint is captured the system will automatically lock itself and the client is required to do re-confirmation at the bank. The following figure 2 illustrates the authentication process once the fingerprint has been captured.

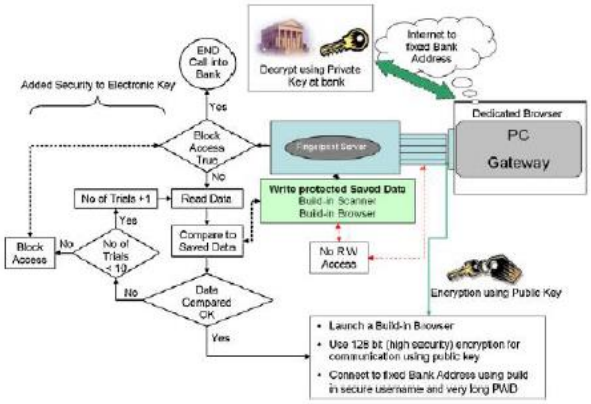

Figure 2: Control Method (Tassabehji and Kamala 2012).

Tassabehji and Kamala (2012) elucidated that a System Usability Scale was used to present the usability of wide range of biometric systems. According to the results Fingerprint recognition tends to be the most appropriate and winning technique for online verification technology out of all biometric technologies. The following diagram presents the test results in terms of suitability and practical familiarity from user's point of view as articulated by Tassabehji and Kamala (2012).

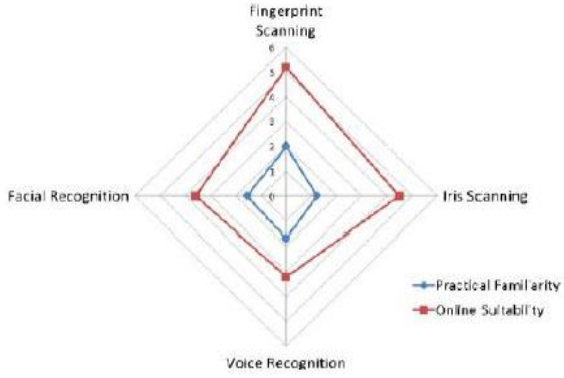

Figure 3: Fingerprint Recognition Test results from User's experience (Tassabehji and Kamala 2012).

Despite the good results from user's experience about biometric Fingerprint recognition, the authors strongly agree that enough is not made about the experience on how the adoption of fingerprint identification as a biometric measure for an online verification achieve security in electronic banking. The author therefore recommends that there should be a formal process that proves how authentication and verification in fingerprint recognition achieves security.

In conclusion, the authors designed a framework for a biometric fingerprint recognition as an online verification technology for securing banking transactions. A System Usability Scale (SUS) was used in this method to test and evaluate its usefulness from the user's perspective.

2.3 Biometric-Kerberos Authentication

The authentication protocol works by allowing the authentication procedure to execute the authentication of the user's identity to the authentication server without transferring confidential data over the insecure network. Nguyeng and Dang (2015) proposed a biometric Kerberos authentication protocol which is enhanced with fuzzy extractor scheme to protect user's biometric. The proposed protocol includes 2 phases being enrollment and authentication. The following diagram illustrates the enrollment phase.

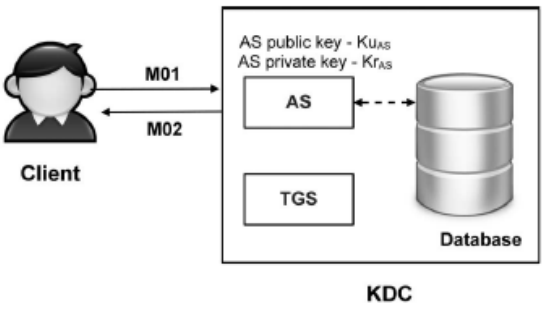

Figure 4: The enrollment Phase (Nguyeng and Dang 2015).

In the enrollment phase, the client is requested for a registration information in a message (M01) for the AS. AS's public key encrypts M01. Nguyeng and Dang (2015)

Clarified that after the M01 is received, AS will employ its private key to decrypt and checks for the duplication of mobile number, then stores the mobile number, serial number and the biometric key into the database if there is no duplication of mobile number. That is when now the AS sends M02 back to the client as a successful registration notification. This is expressed by the following equation:

$$M_{01} = Enc((\ mobile\_num, serial\_num, K_{bio}, N_{01}, timestamp, is\_enroll),\ Ku_{As})$$

$$M_{02} = Enc((\ N_{01}, timestamp, is\_success),\ Kr_{As})$$

Equation 2: Enrollment Phase Message details (Nguyeng and Dang 2015).

The Authentication phase, the first stage of this phase aims at verifying the authentication server AS. The following diagram depicts the first stage of the authentication phase.

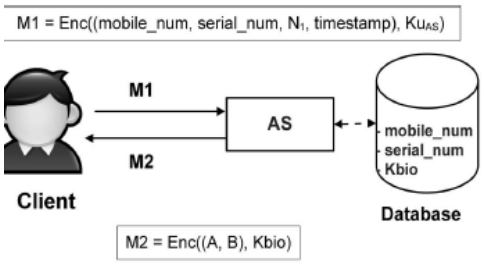

Figure 5: the first stage of the authentication phase (Nguyen and Dang 2015).

The second stage then proceed. The purpose of this stage is to establish communication between the client and the TGS.

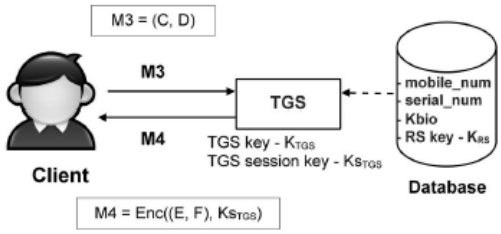

Figure 6: Second stage of authentication phase (Nguyen and Dang 2015).

The client directs message (M3) to TGS. In M3 there is a packet C which contains user's data for TGS to authenticate the client, resource service's address RS_ID and packet D being the ticket granted to the Client by AS. This is explained by the following expression.

$$M_3 = (C, D)$$
$$C = Enc((\ mobile\_num, serial\_num, N_1, timestamp, RS\_ID), Ks_{TGS})$$
$$D = M_2.B$$

Equation 3: Message details for the second stage of authentication phase (Nguyen and Dang 2015).

The Third Stage focuses on the communication between the client and resource service (RS). The following figure 7 and equation 4 explain the process from time the client directs message M5 to RS up until the RS sends the message M6 back to the client to assist the client validate the RS.

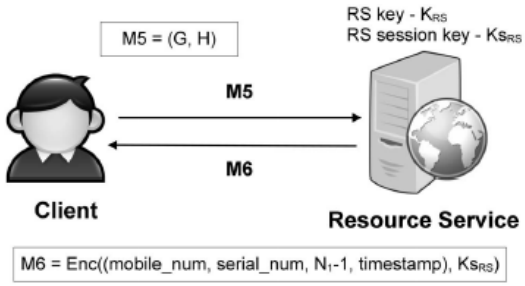

Figure 7: The third stage of authentication phase (Nguyen and Dang 2015).

$$M_5 = (G, H)$$
$$G = Enc((\ mobile\_num, serial\_num, N_1, timestamp), Ks_{RS})$$
$$H = M_4.F$$
$$M_6 = Enc((\ mobile\_num, serial\_num, N_1 - 1, timestamp), Ks_{RS})$$

Equation 4: Message details for the third stage of authentication phase (Nguyen and Dang 2015).

It would have been ideal if test results obtained from this method were presented. When carrying experiments, no tests were made, therefore the reliability of this method is not fully achieved as there were no test results presented to prove how this method verify users during online banking.

In conclusion, Nguyen and Dang (2015) proposed a biometric Kerberos based authentication protocol embedded with fuzzy extractor to fully protect user's biometric. Nguyen and Dang (2015) explained that this verification protocol guarantees the mutual authentication between the client and the server.

2.4 Keystroke Dynamic Method

This mobile-based authentication technique make use of the rhythm in which a client types characters on a keypad especially when entering login details into the bank account. Alshanketi et al (2016) carried out an experiment on keystrokes dynamics using the evaluation of 3 datasets. The author decided to use only 2 datasets named Dataset 1 and dataset 3 to explain the experiment and test results presented in Alshanketi et al (2016).

Dataset 1: this dataset provide keystroke timing data and keystroke features extracted from the data are dwell and flight times. Alshanketi et al (2016) articulated that data was collected from the 51 users, in which each user keyed the similar secret key "rhu uvinversity" in 3 sessions spread more than 3-30 days for 15-20 times. There are 985 samples in this dataset, in which each user gets a minimum of 15 samples and 21 samples as maximum.

Random Forest algorithm was deployed in the dataset 1 experiment in order to differentiate between genuine users and imposters. Alshanketi et al (2016) discovers that False rejection rate (FRR) is calculated for each user by comparing by (10%) genuine test samples (from the user) against the user's profile. The FAR and FRR of the test is obtained by averaging the Individual FRR/FAR. Alshanketi et al (2016) noted that the overall FAR/FRR are calculated by averaging the values obtained over the 10 rounds of cross validation.

The following Figure 8 and Table 2 illustrates the performance test results of keystroke dynamics.

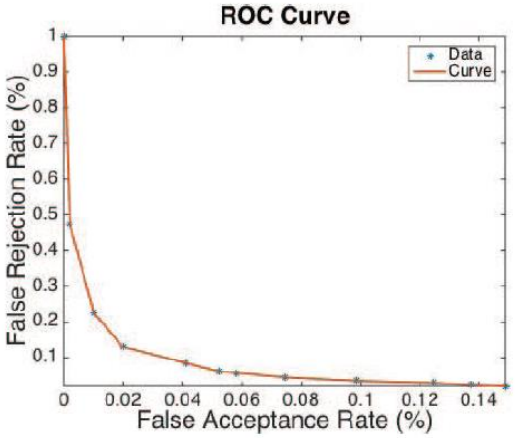

Figure 8: ROC curve for dataset 1 (Alshanketi et al. 2016).

Weight(P)	FRR %	FAR %
0	0	1
0.0001	0.2	47.4
0.0002	1	22.7
0.0003	1.99	12.97
0.0004	4.08	8.64
0.0005	5.24	6.14
0.000515	**5.8**	**5.8**
0.0006	7.44	4.55
0.0007	9.85	3.57
0.0008	12.47	2.95
0.0009	13.73	2.45

Table 2: Results Acquired for Dataset 1 (Alshanketi et al. 2016).

As seen from above, dataset 1 was analysed using cost sensitive Learning. Alshanketi et al (2016) explained that Figure 8 represents ROC curve when under sampling the negative class is not done and by changing the weight (P) for dataset 1. Table 2 illustrates the results of the performance as acquired by changing weight (P).

Dataset 3: Alshanketi et al (2016) explained that this set was obtained in the lab with the intention of learning the effect of typo handling on performance. Galaxy SIII from 10participants who typed 30 times a password string "Mohammed-63". When the user types an incorrect character the captured details are accepted, that called Typo.

Alshanketi et al (2016) proposed an algorithm to handle typing errors. Often times typing errors arise when there is a spelling mistake while typing passwords. When the user mistyped the password, there will be a need for correcting it. The user will delete each character and retype the key, by typing extra letters they will be recorded in time stamps registers. This will lead to a great difference in the flight times between the correct digits. The following diagram shows how misspelling of passwords take place.

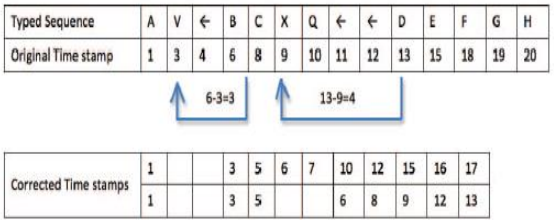

Figure 9: Mistyped Password (Alshanketi et al. 2016).

From the above figure 9, Alshanketi et al (2016) explains that the user mistakenly type 3 extra characters X, V and Q while typing the password ABCDEFGH. The user deleted this after recognizing the error then the user deletes these characters. Alshanketi et al (2016) clarifies that the typo-handling algorithm assumes that the flight time between the incorrect character and the preceding character is almost the same as the time between the correctly typed letters. Then there will be computations made on the difference on time stamps between correct character and incorrect character. Comparison were then made between handling typing errors and not handling typing errors in Dataset 3, which contains typo data. Alshanketi et al (2016) highlights that the process was repeated by performing test on the same dataset, the classifier and the typo-handling algorithm. The following figure 10 and table 4 are results obtained from this experiment when typo handling is permitted when changing the weight (P) for dataset 3. Figure 11 illustrates the results when typo handling is not allowed. Table 3 shows the results when a correct password is typed.

Password	A	B	C	D	E	F	G	H
Time Stamp	1	3	5	6	8	11	12	13

Table 3: Results of a Correct Password (Alshanketi et al. 2016).

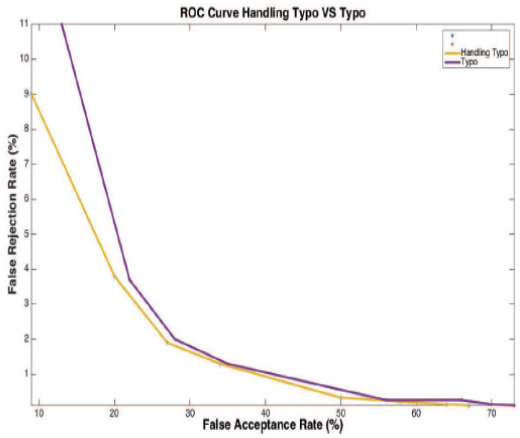

Figure 10: ROC curves results for handling typo measured against no typo handling for dataset 3 (Alshanketi et al. 2016).

Weight(P)	FRR % No Typo	FAR % No Typo	FRR % (Typo)	FAR % (Typo)
0 .24	67	0.11	73	0.11
0.23	64	0.14	70	0.14
0.22	58	0.22	66	0.26
0.21	50	0.33	56	0.26
0.042	34	1.3	36	1.3
0.032	27	1.9	28	2
0.022	20	3.8	22	3.7
0.0021	**9**	**9**	**13**	**11**

Table 4: Results Attained by Handling Typo against No Typo Removal for Dataset 3 (Alshanketi et al. 2016).

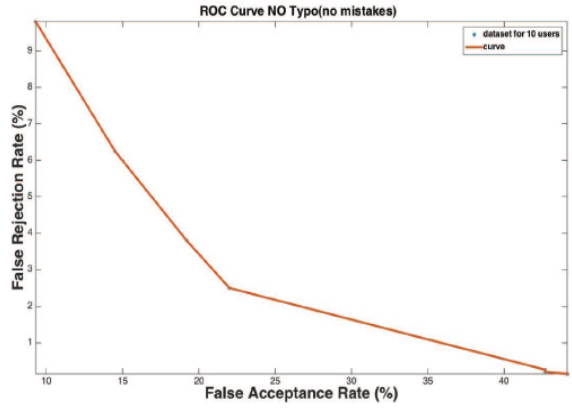

Figure 11. ROC curves without typo for dataset 3 (Alshanketi et al. 2016).

In case where typo handling is not allowed the user's attempt to login will be rejected if the typed a wrong character therefore authentication will fail.

In conclusion, there were no graphical illustrations on experiments made on Keystroke dynamics authentication. There author therefore recommends that is always better to present the carried experiment in graphical view to easily understand the work flow of the experiment since a picture is worth a thousand words. In addition, the discussion on the method tends to leave out the point of interest on how the Keystroke dynamics technique is used as a verification method and how security is achieved in deploying this method.

3 Comparison of Online Transaction verification Technologies

Biometric based online verification technologies make use of unique biometric traits that makes it hard for hackers to imitate and gain access to the sensitive data as compared to other online verification methods. It is very expensive to install fingerprint recognition technique in banking system as compared to online signature verification, which is very cheap to install. Keystroke dynamics also has low implementation and deployment costs as compared to other biometric systems. Over other verification methods, Keystroke dynamics tends to provide a continuous monitoring and authentication for online banking services. Biometric Kerberos authentication is the most credible security solution since it meets the requirements modern distributed banking systems that is highly appreciated.

However Ohana et al (2013) stresses that negative aspect to biometric identification is that biometric keys cannot be changed or altered. Therefore if the system storing biometric templates is ruptured then identity theft may occur.

4 Conclusions

In this paper, the current papers on online verification technologies for financial industries have been fully analysed. The critical evaluation of these technologies was based on the performance, reliability, experiments and test results. Out of all methods evaluated in this work, the study conducted by Hanmandlu et al (2015) is the only work that offered all the satisfactory components. It would have been ideal to use a real online banking facility in testing the techniques to ensure the credibility of the method and how online transaction verification is achieved in a realistic view.

As for biometric based verification technologies, the author strongly suggests that the biometric database server should be fully protected to secure clients' biometric data from identity theft.

Prior to the effort and contributions made by the discussed authors, these techniques can be more promising with more research and testing done even though they do not perfectly protect against data breaches and account intrusions.

References

Alshanketi, F., Traore, I., Awad E. A, 2016, 'Improving Performance and Usability in Mobile Keystroke Dynamic Biometric Authentication', *IEEE*

Symposium on Security and Privacy Works, pg.66-73.

Carvalho, R., Goldsmith, M., Creese, 2015, S., 'Applying Semantic Technologies to Fight Online Banking Fraud', *IEEE European Intelligence and Security Informatics Conference*, pg.61-68.

Costante, E., Fauri, D., Etalle, S., Hartog, J., D., Zannone, N., 2016, 'A Hybrid Framework for Data Loss Prevention and Detection', *IEEE Symposium on Security and Privacy Workshops*, pg. 324-333

Hanmandlu, M., Sayeed, F., Vasikarla, S., 'Online Signature Verification using the Entropy Function', *IEEE*, vol 15, pg.1-7

Kurnia, S., Peng, F., Liu Y., R., 2010, 'Understanding the Adoption of Electronic Banking in China', *Proceedings of the 43rd Hawaii International Conference on System Sciences*, pg.1-10

Li, N., Liu, J., Li, Q., Luo, X., Duan, J., 2016, 'Online Signature Verification Based on Biometric Features', *49th Hawaii International Conference on System Sciences*, pg5527-5534.

Tassabehji, R., Kamala, M., A., 2012. 'Evaluating Biometrics for online Banking: The Case for Usability', *International Journal of Information Management, 32*, pp. 489-494.

Nguyen, T., A., T., Dang, T., K., 2015, 'Combining Fuzzy Extractor in Biometric-Kerberos based Authentication Protocol', *International Conference on Advanced Computing and Applications*, pg. 1-6.

Ohana, D., J., Phillips, L., Chen, L., 2013, 'Preventing Cell Phone Intrusion and Theft using Biometrics', *IEEE Security and Privacy Workshops*, pg.173-180

Telang R., 2015, 'Policy Framework for Data Breaches', *Security & Privacy Economics*, Carnegie Mellon University, pg. 77-79.

Improving the Effectiveness of Network Security Training Using Experimental Programmes

John Bolam

Abstract

The focus of this research paper was to critically evaluate current computer research and provide an insight into improving the effectiveness of network security training using an experimental programme. There are many different types of research that are currently being done in measuring the effectiveness. However, this paper will only evaluate and compare two experimental programmes. Firstly, Long Lasting Memory (LLM) where two different types of exercises are combined to measure the effectiveness over time and secondly, to monitor the effectiveness of virtual classroom training, the methodologies and results are used to base some new conclusions from this evidence for future research in this area.

1 Introduction

Network security is a growing threat and many organizations are spending thousands per year on improving network security awareness. Caldwell (2016) stated that "security awareness training was reporting that the training was largely ineffective". There are now millions of devices connected to networks across the world and even though there's an increased risk of being attacked users continue to ignore any security risks involved.

Sagers, et al (2015) proposed that "The insecurity of Wi-Fi, whether due to user error or ineffective standards, represents a serious threat to personal and business security" this would give us the understanding that most security threats are from the lack of understanding or training provided to staff.

Wolfson & Kraiger (2014) stated that "perhaps there is a need for age-specific instructional formats, future researchers should explore whether and how age affects the learning process" this statement is suggesting that current training is currently ineffective and that we should look at developing new methods of teaching that is more age-focused as the brain takes information in differently the older it gets.

Sagers, et al (2015) carried out some research and reported that there was a strong correlation between age and technical knowledge, having strong evidence from the results showing that the younger generation were far more technically aware and found it much easier to get involved in today's technologies, as skills like these would have been developed from such an early age; the research also reported that older adults were less likely to have this matured technological skill as technology was not as widely available as it is today.

This survey paper will critically evaluate the research currently being done, focusing on areas such as cognitive memory training and alternatively newly implemented methods such as virtual classroom training.

2 Cognitive Training Research

In this section we will be looking at research which has proposed new and improved methods of experimental training using physical and cognitive exercises in multiple sample groups.

2.1 Long Lasting Memory

Gonzalez-Palaua, et al. (2014) carried out some research which aimed to develop an experimental training programme called Long Lasting Memory. The purpose of this experiment was to see if the proposed training programme would help stimulate and keep the brain functions active and the response would be measured through a series of cognitive tests. The LLM training program focused on combining cognitive exercises and physical activity to see if there was an improvement in memory exercises for older generation.

The method used was brought together by seeking out 50 participants randomly and putting them through a strict training regime consisting on a 1 hours' worth of physical training, together with 35 minutes of cognitive training at least 3 times a week for a total duration of 12 weeks. The results were measured to indicate if any participants showed signs of improvement throughout the study. The LLM research published promising results for

enhancing the training effectiveness in older adults using this training programme.

The recruitment programme for the experiment was done between two communities and successfully managed to get 52 volunteers for the study, this included 39 perfectly healthy adults and 11 participants that suffered from mild cognitive impairment (MCI). All participants would endure a battery of physical and cognitive tests which when complete the data could be used to measure the user's effectiveness; these being:

- The Mini Examen Conitivo
- The Digit Span Test of the Wechsler Memory Scale III
- Logical Memory Subtests of the WMS III
- The Color Trial Test 1 and 2
- The Hopkins Verbal Learning Test revised
- The Geriatric Depression Scale

researcher has stated that the tests were taken between a 2 week period before and after intervention. The researcher did not give any indication on why this method for collecting the data was used, however this choice makes the results flawed, as they have unintentionally introduced bias by making sure the participants are quickly assessed straight after the battery of tests are done. This particular choice would have increased the chances of results showing positive improvements and therefore making it look like the research was a success.

Decroix, et al. (2016) did similar research to monitor the effects over an 8-day period of intensive physical training combined with cognitive tests to see if such a combination had a negative effect on overall performance on professional female athletes. The results were conclusive and showed that five athletes showed negative results in performance whereas the research from Gonzalez-Palaua, et al. (2014) indicated positive results from older non-athletes.

	Group							
	Healthy				MCI			
	Pre-*t*-test		Post-test		Pre-*t*-test		Post-test	
Test	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MEC 35	30.91	3.05	31.84	2.50	29.61	3.534	30.44	3.823
HVLT-R Recognition	10.19	1.19	10.61	1.45	10.06	2.53	11.22	0.943
HVLT-R Free Recall	18.00	5.69	19.32	5.02	14.50	4.85	18.33	6.61
HVLT-R delay recall	4.71	2.88	6.32	2.77	3.00	3.23	6.00	2.97
CTT1 (time in seconds)	85.48	32.17	78.78	56.60	87.71	35.01	99.61	52.17
CTT 2 (time in seconds)	161.61	65.01	147.44	77.79	185.41	67.39	196.17	95.65
WMS III digit span forward	6.84	1.37	7.47	1.64	6.17	1.33	5.78	1.39
WMS III digit span backward	4.59	2.40	4.09	1.46	3.50	1.38	3.50	.98
WMS III logical memory/units recalled	26.63	10.40	29.06	10.06	20.50	10.38	22.78	12.69
WMS III logical memory/ thematic units recalled	10.97	4.50	13.03	3.56	7.89	4.17	11.28	4.62
GDS 15	5.44	5.02	2.50	2.55	5.00	3.85	2.39	2.09

Table 1 This table shows test results obtained at pre-test and post-test training intervention. Gonzalez-Palaua, et al. (2014)

The research conducted by Gonzalez-Palaua, et al. (2014) shows results of before the participants took any tests and also the results of after the received training. Their conclusion claims that the results provided do show a significant increase which lead them to believe that the training programme had been a success. However, Gonzalez-Palaua et al. (2014) came to a conclusion that more research in this area is needed, this is because the training programme was limited by time and therefore lacks the monitoring results from long term studies.

Using their results we are able to see some differences between the pre-test and post-test figures, however, based on their chosen methodology and results provided, we can come to a conclusion that in fact they are not scientifically valid and this argument can be justified because the

Karr E, et al. (2014) stated that in their research they investigated the effects of combined physical exercise and cognitive training with similar sample groups as Gonzalez-Palaua, et al. (2014) for example group A being a healthy group and group B being an impaired group. Their results show that they had very similar results that Gonzalez-Palaua, et al. (2014) found and that there were noticeable improvements in performance but only from the healthy sample group.

Büla (2016) had also did research into improving cognitive abilites by the use of increased physical activity however, they soon came to the conclusion that there were quite a few practical implications from the study. Büla (2016) had stated that during their research they found that aerobatic exercises had proven to provide the most positive results.

Phillips, et al. (2016) had also did a similar study to Büla, (2016)

3 Online Training Research

In this section we will be specifically looking and evaluating the effectiveness of training from a virtual classroom training environment.

3.1 Virtual Classroom Training

Agrawal, et al. (2016) did some research into virtual classroom training and have collected evidence on a pre- and post-intervention study.

They were specifically looking at a comparison between virtual classroom training and the standard teaching methods of today. The research aimed to monitor the effectiveness of virtual classroom training to see whether or not there were any improvements in the recipients learning. The methods used for recruitment were a total of 83 students from selected schools in Bihar, India and were assessed through their key competencies in six key areas to establish a baseline understanding for each student.

There was a 72-hour training programme that was developed and used for all students to complete which focused on key areas from the national curricula from second and third year studies. These areas incorporated training from several important roles such as case-based learning, clinical simulations, demonstrations, practice and feedback. The purpose of the training programme was to help the building of knowledge for midwifery students and help improve effectiveness in such areas. Virtual classroom training was intended to work alongside normal classroom training and to be used as an aid to help improve effectiveness for students who required it.

Comparison of pre- and post-intervention scores among the students who attended OSCE.

Name of the practice	Mean score (pre-intervention)	Mean score (post-intervention)	Mean difference	P value
1.Management of second stage of labor	2.3	9.6	7.2	<0.001
2.Active management of third stage of labor	1.8	8.8	6.9	<0.001
3.Essential newborn care	2.2	8.0	5.7	<0.001
4.Newborn resuscitation	0.6	8.8	8.2	<0.001
5.Partograph filling	1.3	7.4	6.7	<0.001
6.Infection prevention	12.8	19.0	6.2	<0.001
Total score	21.3	62.0	40.6	<0.001

Table 2 Shows a comparison of pre and post intervention scores amongst the students. Agrawal, et al. (2016)

Agrawal, et al. (2016) came to the conclusion that found all the areas monitored during the research did show an improvement in learning and claimed the study to be a success. However, Agrawal, et al. (2016) did state that even though students were showing improvements in key areas of their topic they did still lack the practical experience which is critical in their field of expertise.

Based on the methodology and results we can identify several problems, such as the project aim. The researchers initially set out to evaluate the effectiveness of virtual classroom training, however, their chosen methodology was to run the virtual classroom training alongside with their standard teaching already in place, this would mean that students are actually doing their studies twice, which would therefore increase the chances of a better learning outcome and for this reason the results from this study are scientifically invalid.

There are also severe limitations when it comes to virtual classroom training, such as cost. The researchers were using top of the range cisco provider equipment which is critical for the virtual classroom to exist. This could be a huge problem for other researchers looking to re-create this as the type of technology is not exactly specific just a quick mention of cisco equipment. Then there's also the problem of a qualified engineer setting up the equipment which could also lead to other problems arising.

Alotaibi N. & Almutairy (2012) said "virtual classrooms are the most remarkable application of using synchronous means of communication in e-learning" the research aimed at showing the effectiveness of a training programme for staff members at the university which would be used to help further develop their skills in using the new virtual classroom environment for teaching.

Their research developed a training programme that aimed to improve the effectiveness of learning from a virtual classroom environment, however, in the conclusion it had been said that the virtual classroom teaching did not provide any significant differences but did find that the use of role playing had a greater impact and believed that for future strategies research should be done in teaching through virtual classrooms through emphasizing role playing techniques to scientifically verify results.

4 Conclusions

As network security becomes a bigger threat, so does the need for training to help prevent attacks from happening. In this paper we have critically evaluated several types of current research which are looking to improve the effectiveness of training. Firstly, we took a look at the current research being done on improving the mental abilities of the brain

by the use of experimental training programmes and then lastly research on alternative methods such as virtual classroom based training. This research was critically evaluated and the evidence provided was used for our own conclusions and helped justify if the research was scientifically sound or not.

Wolfson & Kraiger (2014) recommended that research should be done in the development of an age-specific training programme as the brain functions differently the older it gets and therefore current training programmes may not be providing enough to the person. However, this research was driven by the bold claims from Van Gerven, et al. (2006) who stated in their research that there was no need for age-specific learning methods, as research done by Salthouse, (1996) who suggested that the older a person gets the harder and slower they will become in processing some tasks. Whereas a younger healthier adult could process and manage tasks more efficiently and therefore the older a person gets the more need for an age-specific training becomes.

Sagers, et al. (2015) did some research work into an experimental training programme which looked at the combination of physical and cognitive exercises to see if the user showed positive results to a battery of tests. Their results came back significantly good, for both test groups, however it became clear that the results could be susceptible to bias and therefore making the results scientifically invalid. Recommendations were brought forward which could be used to eliminate these drawbacks, otherwise the training programme developed was good.

Decroix, et al. (2016) also did some research which looked at the combination of physical and cognitive exercises to see if the professional female cyclists showed any positive results from this, however their published results showed that the combination had a negative effect and that found 5 out of 7 females showed significant signs of fatique, compared to standard training, which, in comparison to the research done by Sagers, et al. (2015) and with this comparison, we can therefore evaluate and hypothesis the question, how can athletes provide us with such a negative response compared to older adults who were showing positive improvements to the combination of training.

In this research paper we also looked at the benefits of virtual classroom training and how this new method was showing positive results from a study conducted to midwifery students in Bihar, India. The methods used for recruitment seemed valid, they were assessing students from a range of schools in the area to prevent bias from a particular school and made sure that the equipment that was provided was from cisco which is a leading manufacturer in networking technologies.

Further evaluation of this research there was a significant flaw in their project aim, which was to evaluate the effectiveness of virtual classroom training. This error is because the researchers made students undergo standard classroom training in addition to the newly implemented virtual classroom training and this causes an issue when you are trying to compare the two types of teaching methods, especially when you're effectively trying to find out if virtual classroom training is improving performances.

There were other limitations to recreate the research and see whether or not their results were scientifically valid, for example the equipment used is only briefly explained stating that they are using cisco equipment however fail to go into any specific details other than that. This presents a problem as there could be hundreds of different types of cisco equipment which could be implemented ranging from low range to top of the range, which could significantly affect the overall training experience gained, especially if the equipment was hard to install, maintain or specifically just did not work.

5 Further Work

There are several areas of research which show promising interest to improving the effectiveness network security training using experimental programmes, however, further research should be done in virtual classroom training and efforts used to design a curriculum that involves more practical role playing sessions.

References

Agrawal, N, Kumar, S, Balasubramaniam, SM, Bhargava, S, Sinha, P, Bakshi, B & Sood, B 2016, 'Effectiveness of virtual classroom training in improving the knowledge and key
maternal neonatal health skills of general nurse midwifery students in Bihar, India.' *Nurse Education Today*, vol 36, pp. 293-297.

Alotaibi N., K & Almutairy, S 2012, 'The Effect of Training Program for Staff Members to Develop Their Skills of Using Virtual Classrooms at King Saud University.' *Physology Research*, vol 2, no. 5, pp. 267-278.

Büla, C 2016, 'Physical activity and cognitive function in older persons.' *Schweizerosche Zeitschrift fur Sportmedizin & Sporttramatologie*, vol 64, pp. 14-18.

Caldwell, T 2016, 'Making Security Awareness training work.' *Computer Fraud & Security*, vol 1, pp. 8-14.

Decroix, L, Piacentini F., M, Rietjens, G and Meeusen, R 2016, 'Monitoring Physical and Cognitive Overload During a Training camp in professional Female Cyclists.' *International Journal of Sports Physiology and Performance*, vol 11, pp. 933-939.

Gonzalez-Palaua, F, Francoa, M, Panagiotis, B, Losadaa, R, Parraa, E, Papageorgioue, SG and Vivas, AB 2014, 'The effects of a computer-based cognitive and physical training program in a healthy and mildly cognitive impaired aging sample.' *Aging & Mental Health*, vol 18, no. 7, pp. 838-846.

Karr E., J, Areshenkoff N., C, Philippe, R and Garcia-Barrera A., M 2014, 'An Empirical Comparison of the Therapeutic Benefits of Physical Exercise and Cognitive Training on the Executive Functions of Older Adults.' *Neuropsychology*, vol 28, no. 6, pp. 829-845.

Phillips, Christine B, Edwards, Jerri D, Andel, Ross, Kilpatrick, Maruc 2016 'Daily physical activity and cognitive function variability in older adults.' *Journal of Aging & physical activity*, Vol 24, no. 2 pp. 256-267

Sagers, DG, Twitchell, DD, Hosack, DB, Rowley, DR, and Nagaraj, MR 2015, 'Where's the Security in WiFi? An Argument for IndustryAwareness.' *Hawaii International Conference on System Sciences*, IEEE Conference Publications, Hawaii, vol 1, pp. 5453 - 5461.

Salthouse, T. A. 1996, 'The processing-speed theory of adult age differences in cognition.' *Psychological Review*, vol 103, pp. 403–428.

Van Gerven, P. W. M., Paas, F., & Tabbers, H. K. (2006). 'Cognitive aging and computer-based instructional design: Where do we go from here?' *Educational Psychology Review*, vol 18, pp. 141–157

Wolfson, NE & Kraiger, K 2014, 'Cognitive aging and training: The role of instructional coherence and advance organizers.' *Experimental Aging Research*, vol 40, pp. 164-186

A Critical Evaluation of the Effectiveness of Animation within Education

Frances Byers

Abstract

Computer Animation is being used more frequently within education. This paper examines and evaluates how animation may benefit learners and how effective it can be. It does this by looking at research papers which use different Animation techniques and relate to educational subjects. The techniques are Computer Animation with Narration, Computer Animation with Teaching, and Computer Animation alone. The papers are evaluated and techniques compared. Finally, conclusions were reached that were mainly positive in favour of Computer Animation, especially in cases where packages were specifically made. Recommendations for further research are suggested so more progress can be made.

1 Introduction

Computer Animation is being employed more frequently within learning. It appeals to many as they find it fun and enjoyable to use. How it is presented within the field of education is of great importance, however there are mixed opinions as to its effectiveness.

It has been stated by Abbott, et. al. (2010) that for a long time animation has been seen as offering control, engagement, as well as clarity. Therefore, animation may help keep learners focused on tasks as well as improving their understanding. Another paper supports this argument stating "In its best uses, animation presents information in a more interesting, easier to understand and remembered way than static media" (Norton, & Sprague, 2001 cited in Ali, 2010). They add that having moving images to watch rather than plain images may allow learners to be more interested in the topic(s) they are learning about. Based on this, animation may therefore enable learning to be done more effectively allowing students to process information without even realizing that they are learning. A study showed "67% of respondents (of 108 students) indicated that animations and simulations in courses enabled them to grasp difficult concepts more easily." (Mtebe, J. S 2016). Brekke, et. al. (2010) state that it is a lot easier to explain how a transistor works if an animation can show the inside of it. The same has been stated by Lipeikien and Lipeika (2006) who say it can be used to demonstrate concepts that are hard to explain verbally or with pictures. It is also implied animation is useful for learning English "Animation comes into full play in English class teaching" (Sharma, P., 2015).

While there is evidence that animation is effective, not all findings agree, R. Nilforooshan, et. al. (2013) state that animations being used in education is not new though there is not any evidence to support the fact that animation is effective within education. Gero and Zoabi (2015) agree with this point saying studies that look at the long-term effect of animation based learning are quite small. Animation can be distracting, and can give unwanted information, "Animations are often too complex or too fast to be accurately perceived." (Morrison, J. B., Tversky, B., et. al. 2002).

The purpose of this research paper is to evaluate the use of animation learning techniques within educational subjects, to improve its effectiveness.

2 Animation based Learning techniques

This section reviews animation based learning techniques proposed by different researchers which are Computer based Animation with Narration, Computer Based Animation alongside traditional teaching and Computer Animation solely.

2.1 Computer Based Animation with Narration

The study undertaken by Panagiotis Ioannou et. al. (2016) compared two different teaching methods and their effects upon the understanding of students during P.E. lessons. The methods were narration only and animation with digitalized narration. The study took place during teaching of the shot-put event. Thirty primary school children from four public schools volunteered to take part in this study.

The study included children learning about two styles of shot-put, namely Glide and Spin. The children had little experience and knowledge of shot-put. They were randomly separated into two groups; A narration only group and an animation with narration group. For the pretest, students completed a subject questionnaire which involved rating how much they knew about shot-put using a one to five rating system (one being not very knowledgeable and five knowing a lot).

The study was explained to the participants. After this the first group were given a brief talk about Glide and Spin which was repeated three times. Following this, the group filled in a problem-solving questionnaire on the subject. They then completed a retention test asking them to describe the Glide and Spin techniques.

The second group viewed an animation with accompanying narration about Glide and Spin three times. Following this, they completed the same questionnaire and Retention test. Both teaching methods lasted two minutes approximately, therefore six minutes in total, taking into consideration that they were repeated.

The results from the study were generated using statistical analysis and the T-test. Figure 1 below shows the average scores and standard deviations of the two groups for both tests. It proves students in the animation with narration group performed better in both tests. The maximum score possible for the retention test was 7 and the problem solving, 6.

Group	Retention		Problem-Solving	
	M	SD	M	SD
Narration	4.16	1.17	3.27	1.04
Animation with Narration	5.98**	1.28	4.91**	1.36

**p<0.01

Figure 1 Ioannou, P et. al. (2016)

The authors thereby concluded that students can learn better when a combination of pictures and words are used rather than just words. They imply learning in this way can improve understanding and suggest that future studies should take place in an athletic environment.

The children involved in the study were on an even footing as none of them had prior knowledge of the subject and were randomly allocated to groups, therefore preventing bias. The methods used in the experiment were repeated three times for both groups. Four schools were involved which gave the study more range for data collection.

The study claims to have a problem-solving test though its content proves that this is not the case; one question simply asks about aspects of shot-put. The answers for the problem-solving test were multiple choice, which could produce false information as it would be possible to guess correct answers. One of the tests specifies that students cannot move forwards or back between questions which seems pointless. It is stated that the teaching methods went on for six minutes approximately; this included listening to the narration / watching the animation three times and seems like a very short period of time to retain new information. All tests appear to have been conducted the same day, one immediately following the other. It would probably be preferable for the retention test to have been undertaken a little later for a more accurate reflection of what had been learnt. Shot-put is a narrow field, and therefore the post tests had maximum scores of 6 and 7, respectively. This did not provide a sufficient range of data to effectively make decisions on which method was the most effective. In view of these issues, it is fair to say that the conclusions reached by the authors were not proven and therefore invalid.

2.2 Computer Based Animation with Teaching

The aim of the study undertaken by Tannu, K., (2008) was to improve the quality of science education using animation alongside teaching. The process involved finding the most effective and feasible method of teaching science using computer animation. The methods used at in this study were named A, B, and C. A involved a teacher using computer animation to teach a topic and then the students studying the topic using a computer. B had a teacher demonstrating a topic using computer animation while students observed. C entailed the teacher teaching it and a week later, the students would watch the same topic on computer.

Fifty-nine students were used for the study. They were tested using a pretest and posttest and data was collected based on this.

Data collected

No.	Pre-test marks /10	Post-test marks /10		
		Method A	method B	method C
1	9	9.5	8	7
2	7	8.5	4.5	4
3	7	3	4.5	2
4	7	7	2.5	3
5	6	6.5	6.5	2
6	6	7.5	3	5
7	6	9	2	2
8	6	2.5	0.5	3
9	6	6.5	1.5	2
10	5	7.5	3	1
11	5	2	2.5	2.5
12	5	6	2	1.5
13	5	4	2	2
14	5	2.5	2.5	2
15	4	3.5	6	2
16	4	3.5	4	3
17	4	1	5.5	2
18	3	3.5	1.5	1
19	3	3.5	2.5	1.5

Figure 2. Tannu, K., (2008)

Figure 2 is a table of the results from the test scores from both the pretest and the posttest which include the three methods.

To calculate the most effective method, a paired t test was used. The pretest was made up of simple objective type questions on the basics of science. Following the pre-test, three groups of students were arranged based on their achievement. 'Respiration and Circulation' was the topic used for the development of animations. This was taught to the three groups one by one using the same animations. Following this, a posttest for all groups was conducted simultaneously. It was based upon the students' understanding. Finally, the results of all tests were compared to show the effectiveness of the methods used.

The author of the report concluded that based on the data gathered that method A (the teacher teaching a topic using computer animation followed by students learning the same topic from a computer) was the most effective method.

School staff gave feedback on the feasibility of the methods used. The opinions of staff were divided; the majority of staff said Method A was the most effective, but it was considered the least feasible. Therefore, Method A was to be used for further research to help increase knowledge on the effectiveness of computer animation.

This study had encouraging elements; it considered teacher feedback and tests were conducted simultaneously. Tests included closed questions which were then used to generate quantitative data giving exact results rather than opinions. Practical issues with computers were considered, thereby reducing the likelihood of technical problems such as network connectivity or hacking issues. Alternative learning aids were researched and discussed.

However, the study had a few areas of concern. Firstly, the methods used in the study were too similar, i.e. Students were taught by a teacher and a computer and two of the tests involved computer animation. There is an element of bias as following the pretest, students were sorted into groups depending on their abilities. This would make the results unreliable. The closed ended questions used could produce false data as answers may be guessed without the pupils actually understanding the topics. Whilst the teacher feedback showed Method A to be the most effective, it was also considered the least feasible of the three methods, although there is no explanation as to why this is the case.

Overall, it can be said that this study is a little biased. It is uneven in places and may be open to the possibility of inaccurate data. Therefore, the conclusions reached were not valid, and a repeat of the experiment undertaken would not be recommended.

A study by Gero, A. et. al. (2015) was undertaken to monitor improvement in student achievement using animation based teaching of electronics. 40 students took part in the experiment. After initial training, students were tested. They were then split into two equal groups; an experimental and control. Both groups studied for eight hours with the same teacher. The experimental group studied using an animation method, whilst the teacher explained and pointed out important information. The control group used static diagrams to learn, along with the teacher's narration. Following this, another test took place. For the next ten weeks, students studied together to finish their course. Then, a final test was conducted. Figure 3 shows the results of the tests.

The table shows that there was a significant difference in favour of the experimental group following tests one and two.

Test	Group	Mean	SD	t	p-value
Test 0	Experiment	79.90	22.32	0.61	n.s.
	Control	83.70	15.28		
Test 1	Experiment	85.50	11.39	3.85	<0.001
	Control	60.90	25.45		
Test 2	Experiment	83.45	11.08	3.06	<0.01
	Control	66.75	21.77		

Figure 3 Gero, A. et. al. (2015)

The authors concluded that students who learnt from Animation with teaching had greater results than those taught using static images.

The study was very fair, as the content of the tests undertaken were identical. In addition, it was ensured that those marking tests were not made aware of which students belonged to which group. The same teacher was also used throughout the testing.

The study could have had a larger sample size to increase data. According to the paper, the teacher had to explain some of the points in the Animation and stress its limitations which implies that it may not have been ideal for its purpose. Overall, however the data shows greater results using Animation with teaching and therefore the author's conclusions are valid.

2.3 Computer Animation

A study undertaken by Gambari, A. I., et. al. (2014) investigated the effectiveness of using Computer Animation in Mathematics tuition in comparison to other methods. It included a pre-test, post-test and a delayed post-test so that results could be compared showing any changes / improvements in students. Sixty pupils from three different schools were involved in the study. Students were assigned to groups at random; CAP (Computer Animation Package), GIM (Geometrical Instructional Model) and TTM (Traditional Teaching Method).

CAP was designed particularly for the study as there were few commercial animation packages available, and those that were, were unsuitable. Created by a team of professionals and specialists, it was interactive, monitored progress, would revise problem areas and displayed quiz pages for self-testing. Six geometry topics were covered and both static and animated images were used. Educational and computer experts were involved in the testing process of the package prior to student use. GIM group was taught geometry using 3D plywood models. At the pre-test stage, all groups were deemed academically equivalent.

Figure 4, a table of the post-test analysis, shows that the students taught using the Computer Animation Package had greater marks than the other students.

Table 2B: Scheffe's post-hoc analysis of the groups means scores

Groups	Mean Scores	Group I (CAP)	Group II (GIM)	Group III (TTM)
Group I (CAP)	73.20		*0.000	*0.000
Group II (GIM)	64.10	*0.000		*0.000
Group III (TTM)	54.90	*0.000	*0.000	

* The mean difference is significant at the 0.05 level.

Figure 4 Gambari, A. I. et. al. (2014)

The authors concluded that students taught geometry using Computer Animation fared significantly better than those taught using other methods as it helped improve their performances and heightened their retention rates.

The study had many positive elements. Firstly, using three different strategies allowed results to be compared more effectively. The study used schools with similar characteristics making the experiment fair. CAP was tested by professionals proving it suitable for use in schools. There is a small inconsistency within the study however in terms of student numbers. The abstract states that forty students were involved but later, it is stated there are sixty.

This study was well executed despite minor flaws. The involvement of more than one school provided a varied range of data. The results do not imply any bias and provided permission is given, the experiments could be repeated within other educational environments. Its conclusions were valid, as it achieved what it set out to do through analysis, thus proving that Computer Animation is more effective in Geometry education compared to alternative methods.

Another study conducted by Nilforooshan, R., et. al. (2013) looked at the effects of computer animation on students' learning in Energy and Environmental design (LEED-ERS). Sixty-eight students were involved. It used two different methods, namely a LEED-ERS animation (specifically designed for the study) and a reading assignment. All students took part in a pretest to establish their current knowledge. Based on the test, students were split fairly into two groups. Both groups had a thirty-minute lecture and took a second test. Following this, Group A (the control group) read for 45 minutes, whilst Group B (The experimental group) watched the LEED-ERS animation for 45 minutes. After this, they both took a posttest. All tests involved different questions.

Figure 5 shows the scores of students at different stages of the study.

	Lecture Effect	Experiment Effect	Intervention Effect	Lecture Effect %	Experiment Effect %	Intervention Effect %
All Participants						
Average	7.46	2.50	9.96	13%	-----	17%
STD	12.19	14.58	10.16			
Control Group (Group A)						
Average	8.73	-0.90	7.82	15%	-1%	13%
STD	15.01	13.07	8.28			
Experimental Group (Group B)						
Average	5.97	9.40	15.37	11%	15%	28%
STD	10.80	12.50	6.72			

Figure 5. Nilforooshan, R., et. al. (2013)

The results showed that after watching the animation, knowledge was increased by 15%, whilst the reading group had a slight decrease in knowledge of -1%. The authors concluded that the use of animation by itself had greatly impacted upon students' learning of LEED.

This study was very well executed, and methods used were comparable, being different in nature.

The contents of the tests were identical but questions were different in order to test understanding. Bias was eliminated as groups were evenly matched on pretest results and the reading and animation content was virtually identical. The LEED-ERS animation was tested by experts and target users.

The study used non-interactive animation which could be somewhat limiting. In all other respects, however, conclusions drawn by the authors were valid.

3 Conclusions

In this research paper, different Computer Animation techniques within Education have been critically evaluated. Of the studies analysed, three were particularly promising. The study by Gero. A, et. al. (2015), which used Computer based Animation with teaching was successful. The package used had been specifically chosen due to its simplicity; the course it was used on is more practical than theory based and the students on it are mainly low achievers. The teacher had to point out limitations of the animations however, so it does seem it may have not been ideal for the purpose.

The two studies by Gambari et. al. (2014) and Nilforooshan, R., et. al. (2013) used packages that were specifically designed for the studies. The LEED-ERS animation uses diagrams, text, moving images and it received excellent ratings from all experts and target users. However, being non interactive, it would have limited potential. The CAP, on the other hand, uses four different software packages, has navigational and interactive features and even a quiz. CAP therefore would be recommended as it has the best functionality and positive attributes of all Animation techniques reviewed.

It would therefore seem that the best Computer Animation packages are finely tuned to the needs of the user. It would be recommended overall, that more animation packages are produced as there is a lack of suitable resources available for the purposes of the studies. Computer Animation packages require a lot of time and money to develop, so there is still much work to be done. It would be essential if studies such as these are to be undertaken in the future, that larger resources are made available. Overall, it can be said that there is definite proof within the studies that Animation is effective within education however more evidence and knowledge is required.

References

Abbott, C., Brown, D., Evett, L., Standen, P. and Wright, J., 2010, 'Learning Difference and digital technologies: a literature review of research involving children and young people using assistive technologies 2007-2010', *King's College. London,* Volume 1, pages.1-17.

Ali, A. Z.M, and Madar, A. R., 2010, 'Effects of Segmentation of Instructional Animation in facilitating learning', *Journal of Technical Education and Training*, Volume 2, page *2*

Brekke, M. and Hostad, P. H., 2010, 'New teaching methods-Using computer technology in physics, mathematics and computer science', *International Journal of Digital society (IJDS),* Volume 1, pages 17-24.

Gambari, A. I, Falode, C. O & Adegbenro, D.A, 2014, 'Effectiveness of computer animation and geometrical instructional model on mathematics achievement and retention among junior secondary school students', *European Journal of Science and Mathematics*, Volume 2, Page 2

Gero and Zoaib, W., 2015, 'Animation-based teaching of semiconductor devices: Long term improvement in students' achievements in a two-year college',*iJEP*, Volume 5, pages 42-46.

Ioannou, P. and Bakirtzoglou, P., 2016, 'Animation with concurrent Narration versus Narration In Physical Education lesson', *Phys Cult (Belgr.),* Volume 70(2), pages 135- 144

Lipeikien, J. and Lipeika, E. A., 2006, 'Animation Tools of CAS for Dynamic Exploration of Mathematics', *Informatics in education,* Volume 5, pages 87-96

Morrison, J. B., Tversky, B. and Betrancourt, M., 2002, 'Animation: Can it facilitate? ', *Int. J. Human-Computer studies*, Volume 57, pages 247-262

Mtebe, J. S., 2016, 'Developing and using animations and simulations to teach computer science courses: the case of University of Dar es Salaam', *e-Learning and e-Technologies in Education (ICEEE) 2012 International Conference*, pages 240-246

Nilforooshan, R., Adamo-Villani, N. and Dib, H., 2013, 'A study of the effects of computer animation on college students' learning of Leadership in Energy and Environmental design-LEED', *EAI Endorsed Transactions on e-Learning*, Volume 01, page 3

Sharma, H. L. and Pooja, 2015, 'Enhancing students interest in English language via multimedia presentation', *International Journal of Applied Research,* Volume 2, pages 275-281

Tannu, K., 2008, 'Computer Animations a Science Teaching aid: Contemplating an Effective Methodology', *Institute of Education Sciences,* ED500006, pages 1-23

Evaluating Current Research on the Educational Effectiveness of Augmented Reality

Michael Jopling

Abstract

This paper analyses, evaluates and compares current research aimed at solving the problem of the educational effectiveness of AR, from the bottom to the top of the educational spectrum – from preschool to junior school to elementary school to university. Researchers methods and conclusions are compared to determine the most effective method for incorporating AR into educational institutions to aid and improve pupil learning effectiveness. This paper gives recommendation for future research that can have real world implications on AR in educational institutions in the future.

1 Introduction

Augmented Reality (AR) technology continues to attract attention from educational institutions globally. A breakthrough in research on the educational effectiveness of AR could benefit educational institutions who are looking to incorporate new, effective learning methods to support pupil learning. Bower, M et. al, (2013) claims that "Augmented Reality technology is advancing so rapidly that educational research has not been able to keep pace".

Demands are getting tougher every year on students to achieve and improve grades, and for educators to maintain and improve student results. Numerous researchers such as Dunleavy, M. (2014)., Yoon, A, S. (2014)., Ludlow, B. (2015) and Bower, M et. al, (2013) have all carried out research on the educational effectiveness of AR and have all brought forward various solutions in evaluating the educational effectiveness of AR to try and move the field closer to having answers on how effective AR can be in education to support and improve learning.

The purpose of this paper is to evaluate, analyse and compare various research what is being done to solve the problem of the educational effectiveness of AR. We will evaluate current research that has been carried out on the educational effectiveness of AR in schools, university and preschool. Conclusions and comparisons will be made to see if the research carried out in this area has pushed forward the boundaries of human knowledge and helped solve the problem of educational effectiveness of AR.

2 AR in University

Chou, T. (2014) carried out research on the effectiveness of learning using AR. The primary hypothesis of the study was to measure effectiveness of learning using the AR mobile technology developed by Chou, T. (2014) among freshmen students in a university. The AR system developed by Chou, T. (2014) was a mobile touring system of the university in which the participants have just begun their studies at. 100 freshmen students volunteered to participate in the study, 37 were males and 63 were females.

All 100 students completed a questionnaire regarding their reactions and thoughts to AR being rolled out, in order to aim to help them all familiarize themselves with the campus, and based on the feedback from the questionnaire the AR prototype was redeveloped and readjusted accordingly prior to the testing process. To test Chou, T. (2014) AR Touring System students were tasked with individually navigating around the campus using the AR touring system as a guide.

Quantitative data was collected in the form of a questionnaire with the aim of measuring effectiveness of participant learning using Chou, T. (2014) AR touring system. The questionnaire was divided into three sections, the third section being effectiveness of learning. Participants were asked to rank each question between 1 and 5, 1 being strongly disagree and 5 being strongly agree. All 100 participants completed the questionnaire. The results of testing Chou, T. (2014) AR learning tool totaled a mean average of 4.35 out of 5.

Item	Strongly disagree 1	2	3	4	Strongly Agree 5	Mean	SD
17. This system provides me with correct information about buildings	0	0	6	33	61	4.55	0.61
18. The information integrated with the physical objects helps me become familiar with the environment	0	0	8	38	54	4.46	0.64
19. The use of visuals provides helpful representations for comprehension	0	0	12	37	51	4.39	0.69
20. The mobile learning system integrated with AR tags reinforces my spatial memories.	0	0	18	33	49	4.31	0.76
21. The use of visuals is intuitive.	0	1	23	27	49	4.24	0.84
22. AR tagging provides enough contextual descriptions.	0	2	11	33	54	4.39	0.76
23. The direction guidelines on the system take me to the target locations accurately.	0	1	8	35	56	4.46	0.69
24. This system provides me everything I need four touring the campus.	0	3	14	31	50	4.22	1.02
25. This campus touring system provides an efficient context-aware learning experience	0	5	19	36	40	4.11	0.89
The average score for the effectiveness of learning						4.35	0.59

Figure 1 by Chou, T. (2014) questionnaire to measure participant's effectiveness of learning.

Chou, T. (2014) concluded that although using AR to learn was a new experience for all participants, all participants enjoyed using the AR touring system to learn about navigating round campus and participants found the AR touring system effective in learning how to navigate round the campus. Chou, T. (2014) concludes that his AR interactive learning tool could be applied in further educational settings to aid pupil learning however, the researcher further concluded that there were some limitations on this research such as the correlation of gender differences to learning effectiveness and Chou, T. (2014) stated this particularly needs to be addressed going forward.

Testing on 100 students is a broad sample size, however as concluded by Chou, T. (2014) it should have been a more balanced gender ratio as oppose to 67-37%, this would address the gender difference to learning effectiveness limitation concluded by Chou, T. (2014). The questionnaire testing method used by Chou, T. (2014) to measure learning effectiveness of participants could be argued to be a non-accurate method as all participants were made aware the AR touring system was designed by the people who would be gathering the participant post-test questionnaire data that the participants would all be completing meaning participants could feel obligated and pressured into selecting good feedback scores for learning effectiveness of the AR tool due to not wanting to offend or be negative towards the researchers as the questionnaires are not anonymous. Due to the method to gather learning effectiveness data can be argued to be considered not accurate, it is a bold statement by Chou, T. (2014) to conclude the AR touring system could be introduced into other areas of education basing a conclusion on a method considered not able to collect accurate results.

All in all, the conclusions made my Chou, T. (2014) must be rejected as the results are questionable and cannot be trusted therefore further research needs to be done in this area with a restructured data measurement method. Any valid conclusions made by the researcher such as the limitations could be overgeneralized making them no longer justified by the evidence.

3 AR vs Traditional Classroom Lesson

Furio, D. (2015) carried out research comparing the learning effectiveness of two learning methods for children to learn the 'Water Cycle'; an AR iPhone game vs a traditional classroom lesson. Furio, D. (2015) set out a primary hypothesis that 'children using AR would result in at least equivalent learning effectiveness results as the traditional classroom lesson results'.

38 children in total participated in the study, all 8-10 years old, 20 boys and 18 girls. The children were randomly split into two groups, group A and B. Both groups had the same amount of children in each, 19. Group A used the AR game first then the traditional classroom lesson and Group B took the traditional classroom lesson first then the AR game. Both learning methods, AR and the traditional classroom lesson had the exact same learning content, in order to eliminate any potential bias.

Quantitative data was gathered in the testing process, Furio, D. (2015) used questionnaires on all participating children, both pre-test and post-test. The pre-test tested the children's current knowledge on the water cycle then the post-test method would test the children on what they have learned about the water cycle.

The results displayed similar learning results between both AR and traditional classroom learning methods, the results showed that despite which learning method the children used, the children made learning improvements about the water cycle, there were no major differences in learning results between the two learning methods. The AR method learning results did however display higher than the traditional classroom method learning results, indicating Furio, D. (2015) AR game was a more effective learning method than the traditional classroom learning method. Irrespective of which learning method the children used to learn about the

water cycle, the knowledge score tests highlighted increases in scores from pre-test to post-test.

Furio, D. (2015) concluded that irrespective of which learning method the children used, there were no significant differences in end knowledge results however, the researcher did conclude the AR iPhone game achieved greater knowledge results in comparison to the classroom lesson. Furio, D. (2015) concludes the primary hypothesis of this research was successfully achieved, 'children using AR would result in at least equivalent learning effectiveness results as the traditional classroom lesson'.

A greater sample size than 38, which could be considered a small sample size, in order to potentially gather more accurate results, however the male to female ratio was adequately balanced alongside diving the group into halves, 19 children in each group to remove any potential bias. The conclusions reached by Furio, D. (2015) appear to be valid as the results are evident and conclusions were justified by evidence. Besides the limited sample size this is a good piece of research which can easily be repeated.

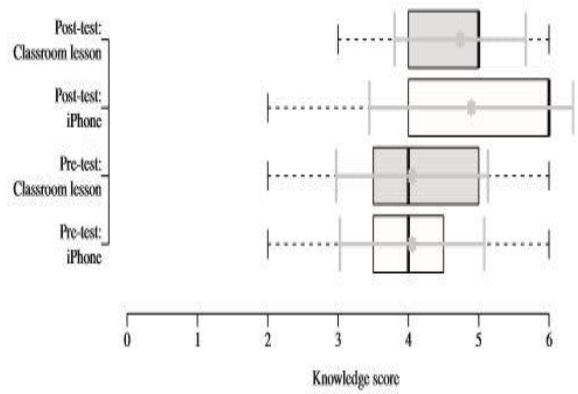

Figure 2 by Furio, D. (2015) knowledge score results. Highligting differences from pre-test to post-test.

4 AR in Elementary School

To improve low academic learning performances of elementary school student's Lu, S. (2015) developed an AR learning tool to support students learning marine education. The purpose of the study was to introduce the participants to various ecosystems of the world. Lu, S. (2015) claimed by introducing students to such graphics, audio and effects over a real world environment as oppose to textbooks, it would lead to students learning more effectively, meaning the low academic performers could improve effectively as a result.

51 students between 7-8 years old participated, 22 males and 29 females, students were randomly selected from two schools. To measure learning effectiveness of Lu, S. (2015) AR learning tool all students sat a pre-test and a post-test. Depending on what result students achieved in the pre-test, students were divided into two groups, high and low achievement groups.

The results show, the AR learning tool for learning marine education helped learners remember the content, meaning learners benefitted in the post test, by enhancing learning outcomes and resulting in greater scores in the post-test compared to the pre-test. Lu, S. (2015).

The researchers conclude, the primary hypothesis of this research by Lu, S. (2015) was to improve low academic learning performers learning outcomes and the low achievers did indeed improve their learning performance. Lu, S. (2015) further concludes, the results in the table below show there is no obvious gap between the low achievers and the high achievers, Lu, S. (2015) method does not have much of an impact on high achievers but does indeed improve the performance of the low achievers.

Using pupils from two schools showed generalizability, thus allowing less room for bias. The male to female ratio could have been balanced to 50% girls and 50% boys to give the study more balance between genders. The author for this research used 51 students from two schools, 51 students is a limited sample size compare to certain other authors who have carried out research in this area such as Chou, T. (2014).

The researcher displayed the methodology clearly throughout. The researcher made it very clear what the students had to do in the post-test to measure learning effectiveness where as it was not made clear what the students had to do pre-test, if this research wanted to be repeated, it would be difficult to repeat as this vital step in the process of measuring learning effectiveness of students is not made clear. Using multiple choice questions for students to answer as a measurement of learning effectiveness, students have the answer in front of them among the options and could just guess and possibly get the right answer, luckily. Meaning, multiple choice can be considered a non accurate measurement method as this could result in non-accurate final results. All in all, the evidence provided to justify the researcher's conclusions can not be considered accurate and reliable, therefore additional testing is required to ensure validity in this claims.

	Pretest answered correctly (%)	Posttest answered correctly (%)	χ^2
Question 1	21 (41.2%)	46 (90.2%)	−4.81*
Question 2	6 (11.8%)	51 (100%)	−6.70*
Question 3	24 (47.1%)	46 (90.2%)	−3.88*
Question 4	18 (35.3%)	46 (90.2%)	−5.29*
Question 5	48 (94.1%)	49 (96.1%)	−0.44
Question 6	42 (82.4%)	46 (90.2%)	−1.26
Question 7	31 (60.8%)	51 (100%)	−4.47*
Question 8	13 (25.5%)	46 (90.2%)	−5.74*
Question 9	12 (23.5%)	50 (98.0%)	−6.00*
Question 10	37 (72.5%)	34 (66.7%)	−0.72
Question 11	16 (31.4%)	35 (68.6%)	−3.52*

Figure 3 by Lu, S. (2015) comparison AR scores from pretest to post test.

5 AR in Preschool

Rasalingam, R. R. (2014) carried out research on the learning effectiveness of AR on preschool children in a classroom environment. The researcher claimed that because the children would be using a learning resource that is part visual and also factual, the children would become more relaxed, meaning the learning process would become more interesting and engaging, thus meaning the children will learn effectively.

The researcher tested AR on the preschool children by using an AR application called AR Flashcards, a scenario was played for each individual child displaying 26 different animals, representing each letter of the alphabet. At the end of the class, the children as a whole were asked questions about the animal alphabet. Rasalingam, R. R. (2014) measured learning effectiveness of the AR tool by pure observation and video recordings.

Rasalingam, R. R. (2014) concluded that pre-school children gain greater knowledge by using AR as a learning method as oppose to a traditional classroom lesson. Rasalingam, R. R. (2014) concluded the AR learning tool proved an effective learning tool as the children appeared to enjoy using AR to learn about animals maintaining interest and enjoying the AR throughout therefore AR will prove to be an effective learning tool in education.

No results were evident to support the researcher's conclusion that pre-school children can gain greater knowledge by using AR as oppose to a traditional classroom method but fail to provide any evidence to justify any conclusions. The researchers claim AR will be effective in education, no statistical data was provided on the learning effectiveness of the AR in this research meaning, further testing is required as we cannot accept these claims with no evidence to justify the conclusions.

6 Comparisons

Rasalingam, R. R. (2014) tested AR on a range of children however, insufficient evidence was provided to justify the claims made. Rasalingam, R. R. (2014) research can be compared to Wasko, C. (2013)., Kangdon, L. (2012)., Bressler, D, M. (2013) and Antonioli, M et al. (2014) who all also carried out research in the same area, all testing AR on a wide range of children, teenagers and adults but did not show sufficient evidence to support the claims made thus meaning the problem of educational effectiveness of AR was not solved, just like Rasalingam, R. R. (2014) meaning additional testing is required from these researchers.

In Chou, T. (2014) research, the researchers claim to of incorporated an effective AR learning tool however with results to show however, the testing method is questionable and arguably not accurate and fair meaning further testing is required to be sure of the validity of these claims, in a more accurate, restructured approach. No educational institutions would benefit from the research done by these researchers, as no evidence has been provided to support the conclusions made, the real world implications would be none.

Lu, S. (2015) and Furio, D. (2015) research is similar to one another, both researcher's methods successfully increased learning outcomes from pre-test to post-test and demonstrated educational effectiveness of AR, both using children with results displayed and both researchers achieving their primary hypothesis however, both researchers used a limited sample size meaning additional testing is required before we can be sure of the validity of these claims. All in all, additional testing is required in this area before we can be sure of the validity of the conclusions made.

7 Conclusions

In this research paper, the most recent literature on the educational effectiveness of AR have been analyzed all across academia, pre- school, school, elementary school and university. All were evaluated and compared, all with the aim of solving the problem of the educational effectiveness of AR.

Firstly, we evaluated Chou, T. (2014) research on AR in Universities, we concluded, although using a satisfactory sample size, by Chou, T. (2014) using a non balanced gender ratio it proved a0 limitation. Chou, T. (2014) conclusions had to be rejected as the validity of the results were questionable,

considered not accurate therefore, further testing is required with the aim of a restructured, accurate testing process to gather more accurate results to compare learning effectiveness.

Secondly, we then evaluated Furio, D. (2015) research on AR vs a traditional classroom method. Despite Furio, D. (2015) using a limited sample size and good male to female ratio, the testing process is easy to repeat with the aim of comparing data, Furio, D. (2015) correlated with the primary hypothesis by increasing student knowledge scores from pre-test to post-test.

Lu, S. (2015) research on AR in elementary schools managed to show increased knowledge scores from pre-test to post-test however the testing method was considered not accurate and questionable, resulting in the results being considered not valid. Lu, S. (2015) also used a limited sample size, with a non balanced gender ratio. Additional research and testing is required from Lu, S. (2015) with a restructured methodology as the current testing process is not a feasible, valid testing method.

Finally, we then evaluated Rasalingam, R. R. (2014), no results were provided to justify the conclusions therefore all conclusions by the researcher had to be rejected. We concluded further testing by Rasalingam, R. R. (2014) is required with the aim of using a different method of measuring learning effectiveness, a method where statistical data can be gathered and compared against another learning method to measure effectiveness. Other papers were also brought into the comparison section, papers which claimed to of solved the problem of educational effectiveness.

Overall, the majority of conclusions reached had to be rejected due to not being justified with evidence or the validity of the results and methods used were questionable. We set out to solve the problem of the educational effectiveness of AR, and although most research evaluated in this paper requires further research, we recommend Furio, D. (2015) research. Furio, D. (2015) method provided evidence AR can indeed be effective in learning and improving knowledge scores, additional testing would be required with a greater sample size to provide more accurate results. We can envisage AR having a huge impact on academia to aid learning and we envisage further research being carried out on the educational effectiveness of AR due to it's potential.

References

Antonioli, M, Blake, C, & Sparks, K 2014, 'Augmented Reality Applications in Education', *Journal Of Technology Studies*, 40, 2, pp. 96-107.

Bower, M, Howe, C, McCredie, N, Robinson, A, & Grover, D 2014, 'Augmented Reality in education – cases, places and potentials', *Educational Media International*, 51, 1, pp. 1-15.

Bressler, D, & Bodzin 2013, 'A mixed methods assessment of students' flow experiences during a mobile augmented reality science game', *Journal Of Computer Assisted Learning*, 29, 6, pp. 505-517.

Chou, T, & ChanLin, L 2014, 'Location-Based Learning Through Augmented Reality', *Journal Of Educational Computing Research*, 51, 3, pp. 355-368.

Dunleavy, M 2014, 'Design Principles for Augmented Reality Learning', Techtrends: *Linking Research & Practice to Improve Learning*, 58, 1, pp. 28-34.

Furió, D, Juan, M, Seguí, I, & Vivó, R 2015, 'Mobile learning vs. traditional classroom lessons: a comparative study', *Journal of Computer Assisted Learning*, 31, 3, pp. 189-201.

Kangdon, L 2012, 'Augmented Reality in Education and Training', *Techtrends: Linking Research & Practice To Improve Learning*, 56, 2, pp. 13-21.

Lu, S, & Liu, Y 2015, 'Integrating augmented reality technology to enhance children's learning in marine education', *Environmental Education Research*, 21, 4, pp. 525-541.

Ludlow, BL 2015, 'Virtual Reality: Emerging Applications and Future Directions', *Rural Special Education Quarterly*, 34, 3, pp. 3-10.

Rasalingham, R, R., 2015. 'Exploring the Application of Augmented Reality Technology in Early Childhood Classroom in Malaysia'. *University Sains*, Malaysia, volume 4, p 33-40.

Wasko, C 2013, 'What Teachers Need to Know About Augmented Reality Enhanced Learning Environments', *Techtrends: Linking Research & Practice to Improve Learning*, 57, 4, pp. 17-21.

Yoon, S, & Wang, J 2014, 'Making the Invisible Visible in Science Museums Through Augmented Reality Devices', *Techtrends: Linking Research & Practice to Improve Learning*, 58, 1, pp 49-55.

A Critical Evaluation of Current Research in DDoS Filtering Techniques within Cloud Computing Environments

Dean Richard McKinnel

Abstract

Rising popularity surrounding cloud computing, coupled with the familiar threat of Distributed Denial of Service attacks has solicited the development of adaptable filtering systems. This paper conducts a critical evaluation of current DDoS filtering systems that are applicable to cloud environments, each system is analysed to determine the viability of the testing performed, furthermore a comparison is made between each of the systems with the premise of realistic combinations of each system for real world application.

1 Introduction

Cloud computing is becoming and a more viable option for modern businesses. Marston et. al. (2011) summarizes that the current paradigm shift into cloud technologies is predominately due to the realisation of companies with regards to underutilisation against the cost of investment. The current high investment from large corporations constitutes a more crucial understanding surrounding the security of infrastructure (Chonka et. al., 2011).

Both Joshi et. al. (2012) and Vissers et. al. (2014) alluded that the most prominent threat to cloud computing is the disruption to availability via Denial of Service attacks. Dou et. al. (2013) elaborated that DDoS is exceptionally difficult to mitigate, as traditional methods of defence cannot be deployed in cloud computing environments, due to the low levels of efficiency.

This paper will analyse the research techniques employed by a small collection of researchers, make comparisons of the varied subject areas and draw informed conclusions for real world applications.

2 Current DDoS Filtering Methods used within Cloud Environments

The following section evaluates current DDoS filtering methods paying particular attention to those applicable to cloud environments. Each proposal will be critically analysed to determine the validity of each hypothesis presented.

2.1 Confidence based DDoS Filtering utilising Attribute Pairs in Cloud Environments

As the fundamental requirements for cloud computing environments differ from those of conventional networks, so must the method used in filtering. Dou et. al. (2013) proposes a confidence-based filtering method, aptly named *CBF* for use in such cloud environments. The proposed system was built to satisfy the filtering requirements needed, whilst being able to maintain the high level of efficiency required for cloud computing.

The proposed system makes use of the correlation characteristics of attribute pairs within packet headers to determine the legitimacy of packets. In order to gain a detailed insight into the occurrence of existing correlation characteristics, Dou et. al (2013) assessed the occurrences of predefined attribute pairs i.e. source IP address & destination port (the header attributes are listed in Figure 1) within different sample sections of the MAWI Traffic Archive.

Single attributes selected from IP/TCP header.

Location	Attribute
IP header	Total length Time to Live (TTL) Protocol type Source IP address
TCP header	Flag Destination port number

Figure 1 - The IP/TCP header attributes Dou et. al. (2013)

The conclusions drawn by Dou et. al (2013) from the results of this traffic analysis outlined that the occurrence of certain attribute pairs were more significant than others - examples of the evaluated pairs can be seen in figure 2; this significance showed that unique occurrences of correlation characteristics exist within packet headers and that the correlation characteristics can be used to determine legitimacy.

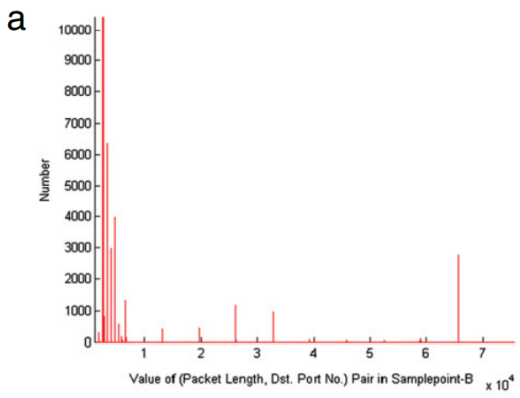

Figure 2 - Example of pair occurrence, Packet length vs. Port No. Dou et. al. (2013)

Based on this significance, Dou et. al. (2013) developed the concept of *confidence* (frequency of attribute pairs) to be used in the measurement of correlation characteristics which in turn were used to calculate the *confidence score* of network traffic. In order to gain the confidence values of each attribute pair, the assessment of the legitimate network traffic would be undertaken during a "non-attack" period. The greatest value of each pair recorded would be stored within the *nominal profile*, later to be utilised within an "attack period" to determine the legitimacy of the packets.

To test their system Dou et. al. (2013) selected a 15 minute interval from samplepoint-B within the MAWI Traffic archive. This dataset was filtered through both the confidence based system and a predecessor system: *PacketScore* that analyses single attributes. During the filtering of this dataset Dou et. al (2013) simulated an array of different DDoS attack types and recorded the false positive and false negative rates associated with each filtering system.

Although the work carried out by the authors demonstrates high levels of testability relating to the large range of different attack types, coinciding with the potential selection of different sample points within the MAWI Traffic archive, the validity regarding the accuracy of the results comes into question. The doubt surrounding the accuracy manifests itself within the variance of certain system parameters that do not conform with a controlled environment.

The first instance of variance in system parameters while testing occurred within the discarding strategy. During testing, CBF utilised a fixed discarding strategy to optimise performance during attack periods, Dou et. al (2013) later elaborates that this discarding threshold was chosen to give the best performance possible, this ethos is not reflected in the later description of the discarding strategy of the compared system however .

In contrast, the CDF based discarding strategy employed by *PacketScore*, utilised dynamic calculations to determine the discarding threshold value. This dynamic threshold may have had an adverse affect on the results gained during testing, as the level of optimisation surrounding CDF was unclear, potentially causing PacketScore to perform inefficiently which may have contributed directly to results within the testing, consequently applying direct bias to CBF.

Parameter values of CBF and PacketScore used in the experiment.

Parameter	Value	
	CBF	PacketScore
Window size (s)	5	5
minconf	0.005	0.01
Selection of single attributes	The 6 attributes in Table 2	The 6 attributes in Table 2
Discarding threshold selection strategy	Fixed	Dynamic

Figure 3 - Parameter values for each system. Dou et. al. (2013)

In addition to the variance in discarding thresholds, the *minconf* values (as seen in Figure 3) of each filtering system were set differently, thus influencing the confidence calculations. As the *minconf* is used to determine the legitimacy of rare-appearance pairs within the nominal profile, a higher minconf value leads to more traffic being listed as legitimate following the confidence calculations, consequently increasing the level of false positives recorded.

Following these variances in testing parameters, considerations should then be brought to the original hypothesis, which focused on the significance of the correlation characteristics regarding attribute pairs in comparison to single attributes. The differences in both the discarding strategy and the minconf values may have adversely affected the results gained from testing the systems and thus made the results highly inaccurate, making the significance of using of attribute pairs less significant that perceived.

It should be noted however, that although these influences may have caused inaccurate results, the indirectly biased data did not allow CBF to outperform PacketScore in some instances. This is highlighted later by the authors, stating "CBF does not have a strictly high accuracy compared to the previous researchers". The results in which PacketScore perform better than CBF may be evidence of the researchers honesty, as the researchers could have easily changed these variables to sufficiently ensure that CBF dominated in all attack types.

Conclusively, the need for a controlled environment is required to ensure the validity of the results

recorded. Further testing should be performed in which acting variables, like that of the discarding strategy and the minconf variable should remain the same (or a close equivalent) in both systems to ensure that the hypothesis can be justified respectively.

2.2 Confidence based DDoS Filtering utilising characteristic deviation in Cloud Environments

Packet analysis is a common methodology with regards to DDoS mitigation; Shamsolmoali et. al (2014) devised a neural network technique that encompasses data mining with the objective of packet filtering in cloud environments.

The system being discussed is a multistage system; stage one consists of a pre-processing method, tasked with the extraction of several header fields within each packet. Stage two involves a multistage analysis; firstly the TTL (Time to Live) attribute extracted in pre-processing aids in detection through comparison with the hop-count attribute (which cannot be spoofed), any inconsistencies will be treated as attack traffic. Following this analysis the traffic is then checked for anomalies, anomalies that do not following the pattern of legitimate traffic profiles collated during the pre-processing stages.

The final stages of packet analysis make use of divergence to distinguish traffic legitimacy, namely the concept of Jensen-Shannon Divergence. If the divergence value is less than that of the learned profile it will still be treated as attack traffic. If deemed as attack traffic the divergence value is then compared to the value of 0.0, if equal to this value the system will flag the traffic as a potential flood attack, if this is the case a frequency count threshold has been assigned to the system to determine if a flood attack is taking place (the process above is detailed in figure 4).

Algorithm.2. Packet Anomaly Detection Algorithm

Input: packet Header Attributes
Output: drop illegitimate packet and Identify Attacker IP
Begin
For each sample *(t)*
 If learning period
  Define probabilities of each value for header Attributes for every *IP* and store them;
 Else
  Define probabilities of each value for the header Attributes for every IP;
  Define the D_i *for* IP_i;
 If $Di \approx 0.0$
  Possible for flooding attack
  Check for flooding using frequency counter;
 If flooding attack (*frequency counter* > *threshold*)
  DDOS attack detected
  Drop matching packets;
  Else
  Forward to destination;
  End If
  Else
  Add the Attacker *IP* to Blacklist;
  End If
  End If
 End For
End

Figure 4 - *Sudo*-divergence algorithm for anomaly detection. Shamsolmoali et. al. 2014

Shamsolmoali et. al. (2014) evaluated their proposed method within a lab environment, containing a multitude of virtual machines; some creating legitimate traffic with others using the *Newtag* tool to generate common DDoS attacks.

During testing, the false positive rate was measured in conjunction with the actual detection rate. A prerequisite test was carried out using the authors generated traffic contained within their lab environment. Figure 5 shows the prerequisite test yielded results comprising of a detection accuracy of 95% and a false positive rate of 4.5%. Shamsolmoali et. al. (2014) extrapolated from the initial test that a detection rate of 100% could be achieved but increasing the detection threshold would have an adverse affect on false positive rate. The secondary test was performed with the commonly used KDD CUP dataset, this dataset was used so that direct comparisons could be made with current mitigation systems. The results of the comparison testing are listed in Figure 6.

Figure 5 - Prerequisite lab testing results. Shamsolmoali et. al. 2014

Classifiers	Correctly Classified Instances	Incorrectly Classified Instances
Decision Tree	92.2	7.8
Bagging	94.5	5.5
Multilayer perceptron	95.3	4.7
PART	98.9	1.1
SVM	97.3	2.7
C_2DF	97.2	2.8

Figure 6 - *T*esting results from KDD CUP dataset. Shamsolmoali et. al. 2014

When evaluating the testing performed by the authors it should be noted that the Shamsolmoali et. al. (2014) carried out a prerequisite test using their own developed dataset, the results gained from this testing were similar to the results of the comparison testing using the KDD CUP dataset, thus this reproducibility reinforces the validity and reliability surrounding the results provided by the authors. In addition to this correlation the application of the KDD CUP dataset provides a good foundation from which to test from. Many researchers including Osanaiye et. al (2016) and Grace et. al (2016) make use of this dataset in their testing to provide a standardised set of results which are comparable with other research ventures. The wide application of this dataset by other researchers makes one believe that the results gained from its usage to be significant through comparison with others.

Upon completion of the system testing the authors accepted that their system did not outperform some of the other systems it was compared to with regards to accuracy and false positive rate. The authors elaborated that changes must be made to ensure that the detection accuracy is viable for application. The authors acceptance of this underperformance merits their research with an element of validity, as the authors could have easily facilitated the injection of bias into their own results to ensure the values presented by their system outperformed others.

2.3 DDoS Filtering utilising Multivariate Correlation Analysis in Cloud Environments

There are many ways in which packet headers can be interpreted for use in DDoS mitigation. One of the more complex instances presented to the research community is that of Tan et. al. (2014), whom developed a system based on *Multivariate correlation analysis*. This type of analysis involves the categorisation of traffic behaviours in addition to anomaly detection to mitigate DDoS attacks. The specific MCA developed by Tan et. al (2014) utilises a triangle area technique (TAT) to increase the efficiency of detection.

The proposed system consists of 3 sections; following the trend of conventional filtering systems the first stage is in place to form legitimate, individual traffic records for later comparison. Step 2 begins the process of the multi-correlation analysis, in the MCA system utilised by Tan et. al (2014) a triangle map generation process is used to extract correlation characteristics from the registered traffic records collated in stage 1.

For each of the traffic characteristics extracted from the individual records the triangle map generation creates TAMs (Triangle Area Maps), the triangle area map functions using a symmetrical matrix in which the upper triangle and lower triangle are compared for symmetry, the occurring symmetry of TAMs allows the system to determine the significance of the characteristics and therefore categorise traffic (examples of TAMs can be seen in Figure 7). The significant characteristics extracted are used in turn to replace individual records in stage 1, thus developing a more precise database for differentiation of legitimate traffic.

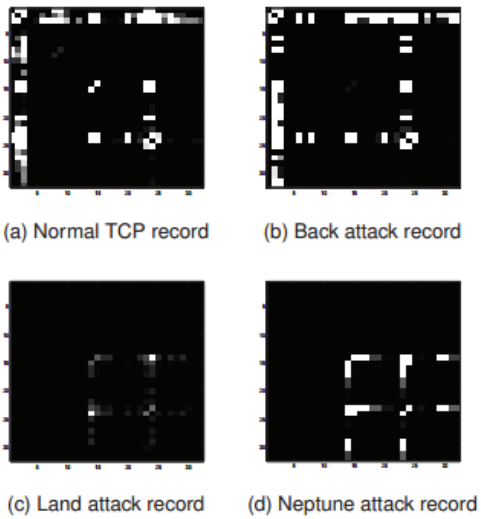

Figure 7 - Triangle Area Maps formed from the original dataset. Tan et. al. (2014)

Stage 3 is split into 2 phases, a training phase and a test phase. When active, the training phase facilitates the creation of the normal profiles which are stored within a database. The test phase consists of building profiles for singular traffic records which can again be compared with normal profiles to determine traffic legitimacy (A flow chart of the above process is shown in Figure 8).

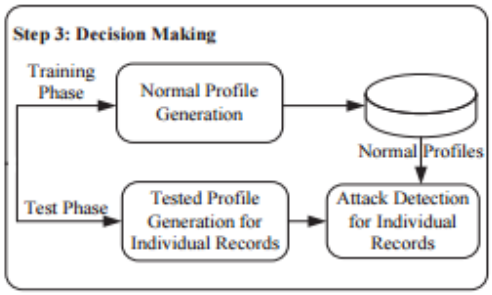

Figure 8 - The decision making process of the proposed system. Tan et. al. 2014

To evaluate the proposed system Tan et. al (2014) engaged their filtering system with the commonly used KDD CUP 99 data records. The authors took 10% of the data sample from the data set, containing legitimate traffic comprising of TCP, UDP and IMCP packets in addition to an array of attack types e.g. Land, Neptune and Back attacks.

Many aspects of the proposed system are evaluated during testing, initially the triangle area based MCA is tested to assess the competency in network characterisation. Following this testing, an evaluation of the detection performance is undertaken in which the full KDD CUP dataset is used. Finally an overall evaluation is performed with the following 4 evaluation criteria: detection accuracy, detection rate, true negative rate (TNR) and the false positive rate (FPR). The results of this overall evaluation is presented in Figure 9 and Figure 10.

Type of records	Threshold				
	1σ	1.5σ	2σ	2.5σ	3σ
Normal	98.74%	99.03%	99.23%	99.35%	99.47%
Teardrop	71.50%	63.92%	57.93%	52.81%	48.45%
Smurf	100.00%	100.00%	100.00%	100.00%	100.00%
Pod	100.00%	100.00%	100.00%	100.00%	100.00%
Neptune	82.44%	61.79%	57.00%	54.84%	52.96%
Land	0.00%	0.00%	0.00%	0.00%	0.00%
Back	99.96%	99.82%	99.58%	99.44%	99.31%

Figure 9 - Detection performance for each of the attacks. Tan et. al. 2014

	Threshold				
	1σ	1.5σ	2σ	2.5σ	3σ
FPR	1.26%	0.97%	0.77%	0.65%	0.53%
DR	95.11%	89.44%	88.11%	87.51%	86.98%
Accuracy	95.20%	89.67%	88.38%	87.79%	87.28%

Figure 10 - Average False Positive, Detection rate and accuracy. Tan et. al. 2014

It is apparent from the results in Figure 9 that the detection performance shows promise with exceptions to certain attack vectors and degradation over others. Furthermore Figure 10 affirms this with regards to the measured metrics.

The authors later stipulated that the cause of the degradation was a direct result of the dataset used not being subject to data normalisation. As the original dataset had not been normalised, the data itself contained null values for traffic patterns that had not been learned during the training phase of the normal profile creation. These null values confused the MCA and hence caused false characteristics to be extracted, thus causing the degradation of the results.

To overcome this the authors made use of the made use of the normalisation technique outlined by Wang et. al. (2009). Following the normalisation the authors carried out an additional 10-fold cross validation test the results of which are detailed in Figures 11 and 12.

Type of records	Threshold				
	1σ	1.5σ	2σ	2.5σ	3σ
Normal	97.36%	97.97%	98.32%	98.56%	98.75%
Teardrop	100.00%	100.00%	100.00%	100.00%	100.00%
Smurf	100.00%	100.00%	100.00%	100.00%	100.00%
Pod	100.00%	100.00%	100.00%	100.00%	100.00%
Neptune	100.00%	100.00%	100.00%	100.00%	100.00%
Land	100.00%	100.00%	100.00%	100.00%	100.00%
Back	99.32%	98.96%	94.09%	93.79%	93.56%

Figure 11 - Detection performance for each of the attacks following normalization. Tan et. al. 2014

	Threshold				
	1σ	1.5σ	2σ	2.5σ	3σ
FPR	2.64%	2.03%	1.68%	1.44%	1.25%
DR	100.00%	99.99%	99.97%	99.97%	99.96%
Accuracy	99.93%	99.95%	99.93%	99.93%	99.93%

Figure 12 - Average False Positive, Detection rate and accuracy following normalization. Tan et. al. 2014

Compared to other research in the field, the quantity of the testing performed, by the authors was extensive. As described, proceeding the first tests performed the realisation that an influential variable (non-normalised data) had affected the results and the rectification of this added a significant level of validity to the secondary testing carried out, as it eliminated bias from the results. A potential downside to system evaluation is the use of the outdated KDD Cup 99 dataset, Tan et. al. (2014) do make note at the beginning of their evaluation that these are redundant records, but are widely accessible leading to their popularity and wide application in testing. A more recent dataset may provide a more accurate and reliable set of results.

3 Comparison

A problem and potential area of development that is resonated within the conclusions of each of the research areas evaluated, is the performance of the detection rate not being able to achieve 100% without having an adverse affect on the false positive rate.

Following the evaluations of each of the filtering systems, the most promising filtering system based on the evaluations is that proposed by Tan et. al. (2014) in the form of their multivariate correlation analysis. The aforementioned system achieved a consistent average detection rate of almost 99.98% across the different attack types that it was pitted against while maintaining a low average false positive rate of below 1.8% (lowering to 1.25% with an increased threshold).

Comparatively speaking it difficult to make assumptions as to detection rate of the work carried out by Dou et. al (2013) as this metric was not measured, but rather both the false positive and false negative rate in which the authors system achieved a very inconsistent array of results with regards the false positive rate, ranging from 0.51% to 7.7%. It should be noted however that the results gained from Dou et. al. (2013) may be sufficiently more accurate due to the use of the more recent dataset. The systems developed by Tan et. al (2014) and Shamsolmoali et. al (2014) are both evaluated using the antiquated KDD CUP dataset originating from 1999. Being nearly 15 years old this dataset would not offer the level of reliability comparable to the MAWI Traffic Archive employed by Dou et. al. (2013).

4 Conclusions

This paper has critically evaluated 3 DDoS filtering systems that are applicable to cloud environments. It should be noted however that the research evaluated in this paper accounts for only a small percentage of the research in this field, but gives insight into the array of different filtering techniques developed.

The structural similarities of the evaluated systems makes possible combinations perfectly viable. To achieve a filtering system with both an acceptable efficiency and a high filtering rate, an amalgamation of the filtering tactics should be considered. Dou et. al (2013) concluded that their system had an "acceptable" accuracy and a high speed with regards to DDoS mitigation. Positive results may be reaped from the application of the deviation calculations employed by Shamsolmoali et. al. (2014) in their system to improve the filtering accuracy of the system proposed by Dou et. al. (2013) without any detrimental decrease in the speed of the filtering system which is highly rated in Cloud technologies. When considering the alleviation of false positives the systems of Dou et. al. (2013) and Shamsolmoali (2014) could make use of the data normalization undertaken by Tan et. al. (2014) in hope of reducing their higher false positive rates.

Other potential ventures may include combining the evaluated systems of the cloud shuffling technique proposed by Jia et. al. (2014), in which cloud infrastructure is dynamically shuffled to allow for attack containment allowing the evaluated systems to contain more elaborate and time consuming detection techniques as users will not notice any latency in this respect.

It is without question that more research is needed by each of the filtering systems to develop a suitable filtering strategy with regards to cloud computing. Promising basic research by Sharma et. al. (2016) combining the use of 3 filtering systems in an attempt at maximizing system functionality with application towards cloud computing. Unfortunately this research is hypothetical, as no testing has been completed. Future research like this may provide the solution required by the cloud computing.

References

Chonka, A., Xiang, Y., Zhou, W. and Bonti, A., 2011, 'Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks', *Journal of Network and Computer Applications*, *34*(4), pp.1097-1107.

Dou, W., Chen, Q. and Chen, J., 2013, 'A confidence-based filtering method for DDoS attack defense in cloud environment' *Future Generation Computer Systems*, *29*(7), pp.1838-1850.

Grace, C.J.C., Karthika, P. and Gomathi, S., 2016, 'A System for Distributed Denial-of-Service Attacks Detection Based on Multivariate Correlation Analysis', International Journal of Advanced Research in Biology, Engineering, Science and Technology, pp.23-28.

Jia, Q., Wang, H., Fleck, D., Li, F., Stavrou, A. and Powell, W., 2014, 'Catch me if you can: A cloud-enabled ddos defense', In *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (pp. 264-275). IEEE.

Joshi, B., Vijayan, A.S. and Joshi, B.K., 2012, 'Securing cloud computing environment against DDoS attacks', In *Computer Communication and Informatics (ICCCI), 2012 International Conference on* (pp. 1-5). IEEE.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J. and Ghalsasi, A., 2011, 'Cloud computing—The business perspective', *Decision support systems*, *51*(1), pp.176-189.

Osanaiye, O., Cai, H., Choo, K.K.R., Dehghantanha, A., Xu, Z. and Dlodlo, M., 2016, 'Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing', *EURASIP Journal on*

Wireless Communications and Networking, *2016*(1), p.1.

Shamsolmoali, P., Alam, M.A. and Biswas, R., 2014, 'C2DF: High Rate DDOS filtering method in Cloud Computing', *International Journal of Computer Network and Information Security*, *6*(9), p.43.

Sharma, N., Singh, M. and Misra, A., 2016, March. Prevention against DDOS attack on cloud systems using triple filter: An algorithmic approach. In *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on* (pp. 560-565). IEEE.

Tan, Z., Jamdagni, A., He, X., Nanda, P. and Liu, R.P., 2014, 'A system for denial-of-service attack detection based on multivariate correlation analysis', *IEEE transactions on parallel and distributed systems*, *25*(2), pp.447-456.

Vissers, T., Somasundaram, T.S., Pieters, L., Govindarajan, K. and Hellinckx, P., 2014, 'DDoS defense system for web services in a cloud environment', *Future Generation Computer Systems*, *37*, pp.37-45.

Wang, W., Zhang, X., Gombault, S. and Knapskog, S.J., 2009, December. Attribute normalization in network intrusion detection. In *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks* (pp. 448-453). IEEE.

An Evaluation of Security Strategies Aimed At Improving Cloud Computing

Gofaone Oatile

Abstract

Cloud computing is considered one of the most growing fundamentals in the field of information technology. However, the biggest problem and challenge faced by cloud computing technology is the issue of both data security and data protection. This paper describes different security strategies implemented and different methods proposed in recent years. It focuses on network encryption technique, third-party auditor, key-aggregate searchable encryption (KASE) and privacy-preserving using secret sharing scheme. This methods and experimental results are then tabulated and assessed in order to determine the most effective method for cloud security and endorsements are made concerning the best method. In conclusion, the author suggest ways of improving the experiments conducted based on the issues identified.

1 Introduction

One of its key benefits to companies and big organisations is reduced cost, time and a flexible infrastructure, as it is better than developing and operating your own infrastructure. Yunchuan et.al (2014) points out that data security in cloud computing is more complicated than in traditional information systems. Kire (2016) further asserts that cloud computing runs on a network infrastructure that is prone to different types of security threats. Some of this security threats are unwanted access, data remanence, data location, data segregation and vendor lock-in. Dushyant and Vijay (2015) points out that confidentiality, integrity and availability are the most important parameters that are considered for security. Lie (2011) also agrees with this by pointing out that data privacy and service availability are the key security problem.

Cheng-kang et.al (2014) proposed a scheme aimed at improving data sharing and securing data. The scheme uses a public-key cryptosystem that produces a stable size ciphertexts. Nevertheless, the limitation of this scheme is that in cloud storage the amount of ciphertext increases quickly so they is need for expansion of the public-key. Ovunc and Tolga (2015) also proposed a scheme that removes data privacy concerns in public cloud by using the fully homomorphic encryption, the limitation of this scheme is the performance disadvantage.

Some techniques have been proposed by different authors to offer an improved performance and security on data saved in cloud storage system. Baojiang et al. (2016) proposed a scheme in which a single aggregate key is sent to the operator for distributing larger amount of files in the cloud then the operator has to present a single trapdoor. Syed and Abdul (2015) also proposed an auditing method, which ensures that integrity, insider threats and malicious activities are minimised with the use of a time-released session and service level agreement. Alexandru and Nicolae (2014) proposed a secret sharing scheme for providing a secure storage of data in the cloud. In this scheme, data is divided into chunks and each chunk carries a minimum information content. Chang (2015) proposed a strategy, which uses double encryption that helps to guarantee that the user's information transported safe and effective.

The main aim of this paper is to evaluate current security strategies used in cloud computing which are network encryption technique, third-party auditor, key-aggregate searchable encryption (KASE) and privacy-preserving secret sharing scheme. The whole of this paper is organised as follows: first is a detail discussion of proposed security techniques by different authors and critical evaluation of this techniques in Section 2. Then lastly the conclusions.

2 Evaluation of Security Strategies Used In Cloud Computing

2.1 Third-party Auditor

Syed and Abdul (2015) proposed an auditing method which ensures the integrity of the third-party auditor and it helps minimise threats and malicious activities. This approach is considered cost effective because, the scheme is responsible for managing both the encryption key and checking the integrity of the data. The third-party auditor guarantees transparency, efficiency, reliability, data confidentiality as well and integrity.

Each user in the cloud community is a member of the community using different services provided by the cloud and the cloud provider is expected to produce the necessary data computation. Both the user and the provider verify whether the third-party auditor performs the assigned tasks within the given time frame and time-released session (Syed and Abdul 2015). The auditor is also responsible for checking the integrity of the third-party auditor using the time-released session keys. The cloud provider will then ensure the integrity of the third-party auditor sing the time-bounded session key. Time-bounded session key simply means checking the integrity of a third-party auditor without having access to the data.

According to Syed and Abdul (2015), the proposed method was tested on two different scenarios using a privacy-preserving model. In addition, Syed and Abdul (2015) also point out that the model was programmed in C++ and tested on a GreenCloud simulator and a default Ubuntu 12.04 operating system. Experiment was run on a computer 2.8GHz Pentium Dual Core CUP and 5GB RAM, 64 bits' version with Windows 8.

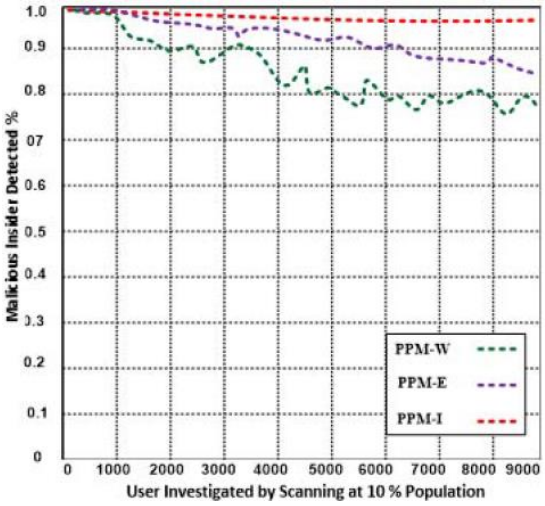

Figure 1 Malicious Insider Detection versus Users (Syed and Abdul 2015)

Figure 1 shows a malicious insider versus the number of users.

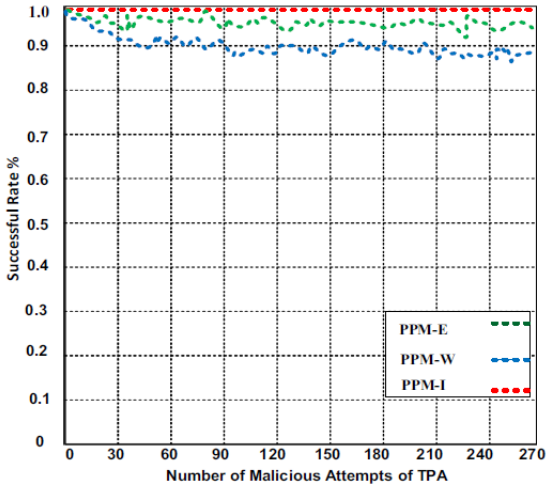

Figure 2 Malicious Attempts versus Successful rate (Syed and Abdul 2015)

Figure 2 shows a malicious attempt of the third-party auditor with different scenarios; ideal, expected and worst. Given these facts as the number of users increase the overall detection rate decrease and the expected outcome turns to be acceptable according to Syed and Abdul (2015). An ideal case also provided a successful rate of 100%, expected provided 88.3% and 96.6% for worst.

With this in mind, it is clear that Syed and Abdul (2015) did not carry out any experiment on the security part of the method especially the storage correctness and different auditing methods as the auditing methods make sure that data stored in the cloud is secure. In addition, the proposed method did not show any security results of their method instead, they presented different equations on privacy-preserving. Therefore experiments needs to be carried out for the untested parts of the proposed method. More experiments also need to be conducted to test the authors proposed method against other methods in order to get a solid and fair conclusion.

The experimental setup for testing the scheme was well presented and the outcome showed the method proposed.

2.2 Key-aggregate Searchable Encryption (KASE)

Baojiang et al. (2016) proposed a key-aggregate searchable encryption technique, in which the data owner for sharing documents sends a single key to user. The user will then present a single corresponding keyword to the cloud. Baojiang et al. (2016) considers the scheme secure and efficient.

Every searchable encryption key is linked with an index of the document and there are seven algorithms created to make this scheme efficient.

These algorithms are setup, KeyGen, encrypt, extract, trapdoor, adjust and test.

According to Baojiang et al. (2016), the owner of the data to initialize system parameters executes a setup algorithm, and then a key is generated using the KeyGen algorithm. Thereafter the data owner uses encrypt to encrypt data and decide on which ciphertext class is associated with the data. The extract algorithm is used to produce an aggregate searchable encryption key, and then the user uses the trapdoor to execute a keyword search. A right trapdoor is produced by the cloud server using the adjust algorithm, and the cloud server will assess the algorithm to finish the keyword search.

The experiment studies whether the cryptographic processes based on pairing computation can be effectively executed by means of both computers and mobile devices. Two libraries were used the jpbc library which is used to apply the cryptographic operations in mobile smartphones and the pbs in computers. Baojiang et al. (2016) points out that the evaluation will be done in two different platforms one being the Samsung G3503U phone with Android OS 4.2 and the other is in C++ on a computer of Intel(R) Core™ i5-3337U CPU 1.80GHZ with Windows 7 OS.

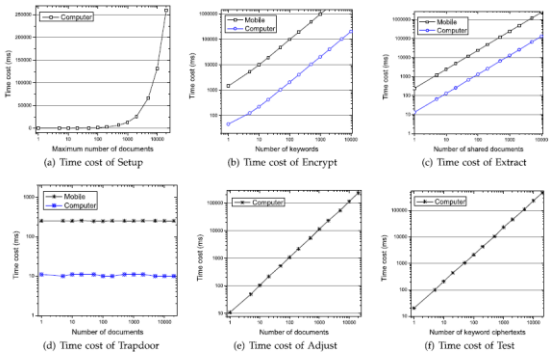

Figure 3 Time cost of Key-aggregate algorithms (Baojing et al. 2016)

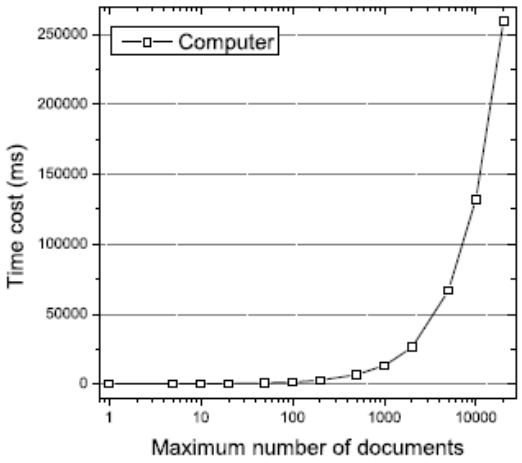

Figure 3.1 Time cost of setup (Baojing et al. 2016)

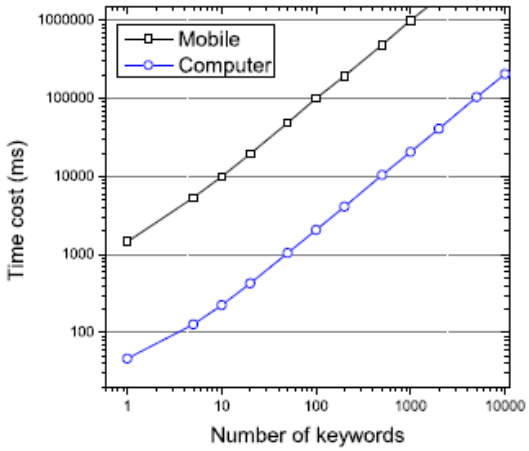

Figure 3.2 Time cost of encryption (Baojing et al. 2016)

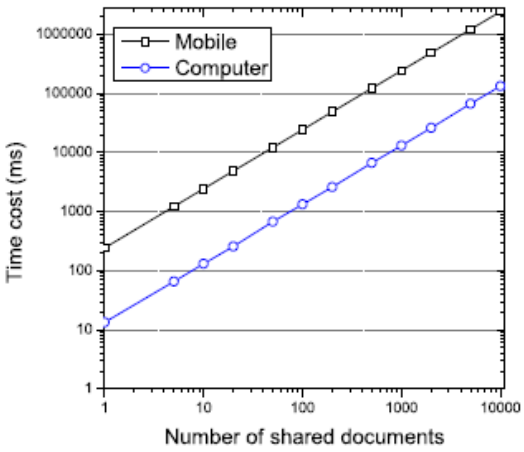

Figure 3.3 Time cost of extract (Baojing et al. 2016)

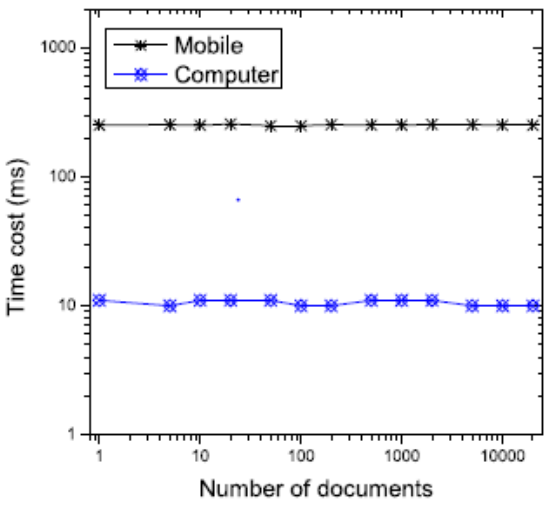

Figure 3.4 Time cost of trapdoor (Baojing et al. 2016)

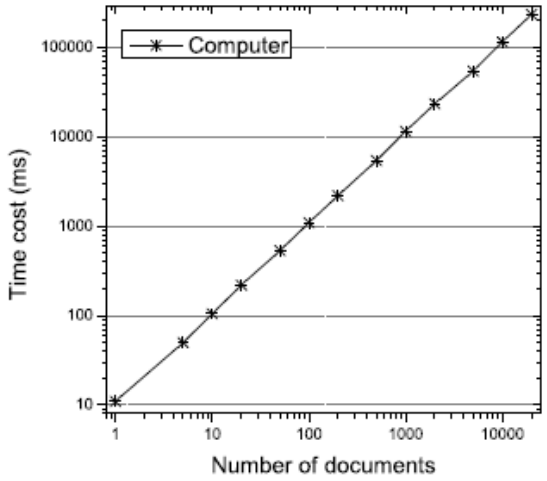

Figure 3.5 Time cost of adjust (Baojing et al. 2016)

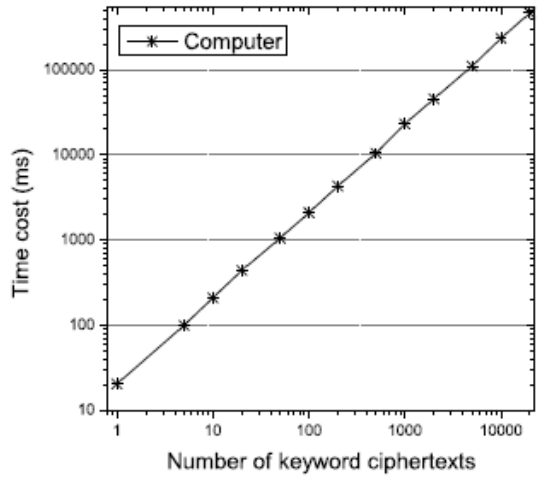

Figure 3.6 Time cost of test (Baojing et al. 2016)

The execution times were tested in a computer. As shown in Figure 3.1 to figure 3.6, Baojiang et al. (2016) points out that once the quantity of documents raises to 20000 the setup procedure will need 259 second. Once the keywords increase to 10000 the encoded algorithm needs 206 second in computers and 10018 in mobile phones. When the documents shared increases up to 10000, the extract algorithm requires 132 seconds and 2430 on mobile phones. The trapdoor is constant and adjust is linear. When the number increases up to 20000 in the test algorithm it will take 467 seconds.

Baojiang et al. (2016) was able to test all the components of the proposed technique, and the results of each algorithm were presented. The author tested both computers and mobile devices because users can retrieve data by these devices. This simply means that the aspect of generalizability is met since the outcomes do not only influence one device but two, in addition, the aspect of biasness will be regulated.

The author did not demonstrate a sense of comparison, as they did not compare their results of the algorithms with other proposed schemes. Furthermore, it is also not settled in the author's paper on how many epochs the experiment was repeated to guarantee reliability. To conclude a critical evaluation of the research confirms that the technique proposed cannot be applied in the case of federated clouds.

2.3 Network Encryption Technique

Chang (2015) proposed a network based encryption strategy that helps keep the transmitted data in the network secure. The proposed strategy uses double encryption and it ensures that the user's data is transmitted securely by utilizing the check message technology. Chang (2015) continues to advocate that the proposed algorithm will help in verifying the integrity of a message.

The cloud server will calculate the message abstract when the user sends information to the cloud server. This is defined as a message authentication code (Chang 2015). For this message abstract to be generated, the DES (Data Encryption Standard) key is combined with a hash function, and then the abstract will encrypt the ciphertext and the Data Encryption Standard algorithm. The server will then decrypt the Data Encryption Standard key then apply the hash procedure to compare the MAC addresses. Chang (2015) explains that the HMAC algorithm is also used to define the sent time and order for sending the message.

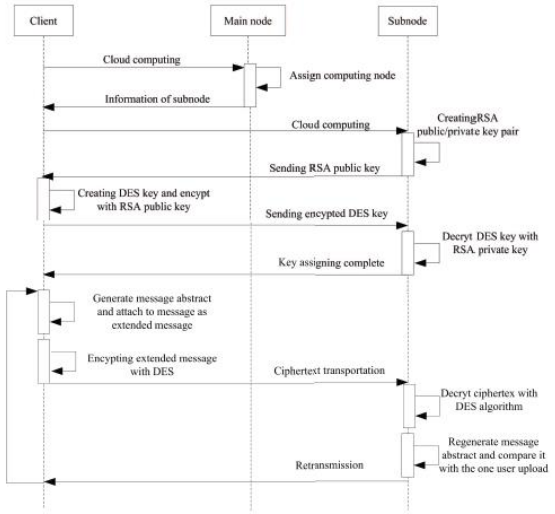

Figure 4 Secure transmission process of cloud computing data (Chang 2015)

Figure 4 shows the process of secure transmission of data in the cloud. The security transmission, which uses the double encryption, guarantee a secure key distribution. Chang (2015) illustrates that in the security transmission process as shown in Figure 4, the client applies for a secure cloud service and then the server will obtain the node based on the client's information file in the system. In return, the client will obtain a computational node message. Chang (2015) points out that the sub node will then generate the RSA (Rivest Shamir Adleman) pair of private and public keys and send them to the client. Which will then generate the Data Encryption Standard(DES) key. Chang (2015) points out that the server will receive the encrypted DES key and the stores it, which means the client can now use the DES key to encrypt data.

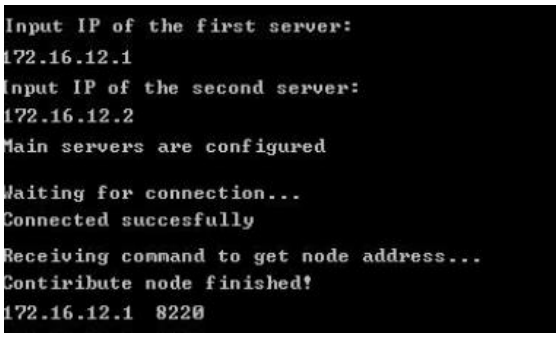

Figure 5 System operation, which tests the server (Chang 2015)

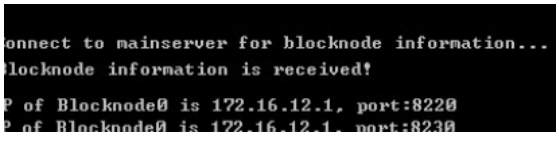

Figure 6 The sub node operation (Chang 2015)

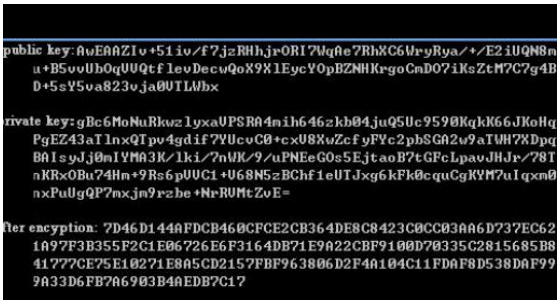

Figure 7 RSA public and private key generation (Chang 2015)

The figures below show the experimental results of the test carried out in the system (Chang 2015). In Figure 5, 172.16.12.1 and 172.16.12.2 are IP addresses, which are distributed by the computing node. Figure6 shows the sub node operation which is waiting for the user end connection. After the connection the sub node will then generate a private and a public, this is shown in Figure 7. According to Chang (2015), the DES key is generated by the client then the public key will encrypt DES key and store it.

Having evaluated the strategy, it is observed that the author did not mention the type or name of the server on which the experiment was carried on, the author did not mention the tools he used. It is also observed that the room for biasness is massive due to the fact that the author only tested the proposed strategy only on one server and the experimental test were carried out only once which means they is a limited chance of precision and reliability on the results.

Besides the limited experimental tests, this is a very good research because the author showed an element of reproducibility other researchers and repeat it since the author described all the inputs and outputs. Furthermore, the experimental results of the proposed strategy demonstrated security and effectiveness of the strategy.

2.4 Privacy-preserving Using Secret Sharing Scheme

Alexandru and Nicolae (2014) proposed a secret sharing scheme. According to Alexandru and Nicolae (2014), the proposed scheme uses the splitting heuristic and the chunk distribution strategy. The proposed scheme tackles the setback of data preservation in a private-public cloud.
When the user updates a file containing confidential material into a public cloud, the information is fragmented into several smaller file chunks. The chunks are then stored in a way that the likelihood of recreating the original file is minimum. Alexandru and Nicolae (2014) continue to point out that every chunk uploaded is documented in a Chunk Distribution Dictionary, which is a dictionary that contains the index of the original file. The chunk

is warehoused and the hash sum of the chunk is stored for integrity check. The chunk is then sent to the public cloud but not as a similar manner as they are attained from the source file.

According to Alexandru and Nicolae (2014) when accessing the file, the file chunk is reclaimed using the hash sum and an integrity check is performed for security of the information in each chunk. The index field from the Chunk Distribution Dictionary is used to reorder the file into the original file.

1: *compute* $I(f)$
2: *while* $f \neq \emptyset$
3: c_i = *GetBytes*$(f, MINCHUNK)$
4: $f = f - c_i$
5: *compute* $I_i(f, c_i)$
6: $maxkl = I_i(f, c_i)$, $countdecline = 0$
7: *while* $I_i(f, c_i) \leq I(f)$
and $countdecline \leq STEPSAHEAD$
8: b_k = *GetBytes*$(f, MINCHUNK)$
9: $c_{i+1} = c_i \cup b_k$
10: $f = f - b_k$
11: *compute* $I_i(f, c_{i+1})$
12: *if* $I_i(f, c_{i+1}) > maxkl$ *then*
13: $maxkl = I_i(f, c_{i+1})$
14: $candidatechunk = c_{i+1}$
15: $countdecline = 0$
16: *else* $countdecline = countdecline + 1$

Figure 8 Algorithm for splitting a file chunk (Alexandru and Nicolae 2014)

Figure 8 shows an algorithm implemented by Alexandru and Nicolae (2014) for splitting a file into secure chunks in the cloud. The first line computes the entire file entropy, f is the entire information content and I is the information carried by the chunk file. According to Alexandru and Nicolae (2014), GetBytes extracts the next bytes in order. Line 6 to 16 seeks an ideal file chunk by carrying the new byte to the entropy. The byte sequence is calculated until the relative entropy is greater than the whole file entropy. Maxkl collects the uppermost computed value and the countdecline variable rises when the relative entropy frequently decreases. According to Alexandru and Nicolae (2014), the search space of an ideal chunk is called the stepsahead, and the ideal chunk is shown by the last located candidatechunk.

An experiment was ran on files that were extracted from an e-learning data platform, which encloses confidential data and 3800 user accounts details. The results presented are obtained were obtained from the using the algorithm presented by Alexandru and Nicolae (2014).

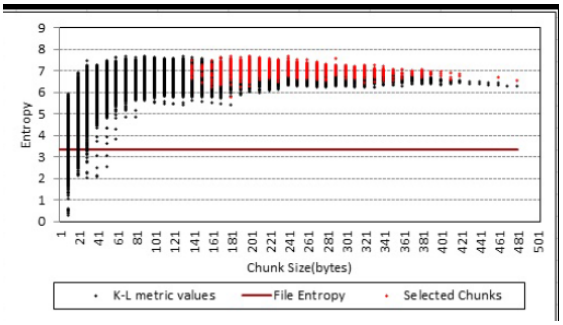

Figure 9 chunk computation Alexandru and Nicolae (2014)

According to Alexandru and Nicolae (2014) the minchuck was set to 10bytes, stepshead were set to 10. The experiment produced 1462 chunks and the maximum relative entropy value of 7.699355 or a chunk size of 20bytes. The total entropy was 3.3477474 and the region for optimal chunk is marked red. It is observed that the algorithm produces chunk value of its maximum value sustaining a stable trend (Alexandru and Nicolae 2014).

The author executed their investigation and research excellently as they stated out the purpose of their investigation. The only problem is that the author did not mention how they got the sample size for the experiment. The author should have mention the sample size to make the outcome of the certain and valid because failure to do this can arise issues like uncertainty, unreliability and invalid results. Apart from this the experiments carried out were on point, the conclusion matched the experiments and methodology used which makes this a good research.

3 Conclusions

The most current security strategies were evaluated in this paper. Both security and performance of different methods were looked into. Different algorithms on different schemes were presented to illustrate how data can be protected in cloud computing even though most research were more theoretical. More experiments on all evaluated papers need to conducted to prove claims proposed by authors.

Baojiang et al. (2016) is the only researcher who presented all the aspects or components of the scheme. The experiment covered all their algorithms used and their results were presented. The network based encryption techniques proposed by Chang (2015) clearly demonstrated security and effeciveness which makes it a better method compared to other proposed methods. Even on the performance aspect the sheme illustrated on how the components of the scheme perform. On reseacrh done by Syed and Abdul (2015), the auditing

proposed method illustrated transparency, efficiency, reliability, data confidentiality and integrity this makes it desirable scheme. On research prepared by Alexandru and Nicolae (2014), it is problematic to draw a conclusion because the proposed method was not tested in contrast to other methods for comparison. Thus, the outcomes are unreliable because the tests outcomes are required of the recommended schemes alongside other schemes.

Some researches may have produced biased results due to poor experimentation for example Alexandru and Nicolae (2014) did not specify how they got the sample size, and this can lead to a null and void conclusion. It can be suggested that when researchers develop a scheme a poll of researchers should be involved and very large sample. The validation of results needs more attention to gain trust and confidence of the user. In future several comparisons with different approaches and results to show effectiveness of proposed schemes.

References

Butoi, A. and Tomai, N., 2014, December. Secret sharing scheme for data confidentiality preserving in a public-private hybrid cloud storage approach. In-*Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on* (pp. 992-997).

Chu, C.K., Chow, S.S., Tzeng, W.G., Zhou, J. and Deng, R.H., 2014. Key-aggregate cryptosystem for scalable data sharing in cloud storage. *IEEE Transactions on Parallel and Distributed Systems*, *25*(2), (pp.468-477).

Cui, B., Liu, Z. and Wang, L., 2015. 'Key-aggregate searchable encryption (KASE) for group data sharing via cloud storage.' *IEEE Transactions on Computers*(pp.2374-2385).

Jakimoski, K., 2016. 'Security Techniques for Data Protection in Cloud Computing.' *International Journal of Grid and Distributed Computing*, *9*(1), pp.49-56.

Kocabas, O. and Soyata, T., 2015, June. 'Utilizing homomorphic encryption to implement secure and private medical cloud computing.' In *2015 IEEE 8th International Conference on Cloud Computing* (pp. 540-547).

Liu, W., 2012, April. 'Research on cloud computing security problem and strategy.' IEEE *2nd International Conference on* Consumer *Electronics, Communications and Networks (CECNet),* (pp. 1216-1219).

Rizvi, S., Razaque, A. and Cover, K., 2015, November. 'Third-Party Auditor (TPA): A Potential Solution for Securing a Cloud Environment.' *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing (CSCloud)* (pp. 31-36).

Sisode, D.B. and Verma, V.K., 2015. 'Maintaining Data Confidentiality and Security over Cloud: An Overview.' *International Journal of Advance Research in Computer Science and Management Studies*(pp.173-177).

Sun, Y., Zhang, J., Xiong, Y. and Zhu, G., 2014. 'Data security and privacy in cloud computing.' *International Journal of Distributed Sensor Networks*(pp.1-9).

Xue-Zhou, C., 2015, June. 'Network Data Encryption Strategy for Cloud Computing.' In *2015 Seventh International Conference on Measuring Technology and Mechatronics Automation* (pp. 693-697).

An Evaluation of Current Research into the Potential Negative Impact from Violent Video Games on Teenagers' Aggression

Christopher Riddell

Abstract

The Game Industry is one of the most rapidly growing industries ever seen, but the side effects of the video games it produces go largely unnoticed. This paper evaluates current research in the area of violent video games and their impact on aggression to demonstrate problem areas, and suggest potential mitigation techniques.

1 Introduction

The Video game industry has grown exponentially since its rise in the 1980's, and can be seen advertised anywhere and everywhere, as Wei (2007) demonstrates with "The number of Internet users soared to 100 million in June 2005, from 10 million in 2000" in China alone. There are a wide variety of games available, with anything from Football Simulators to First-Person shooters being available in video game form.

The peak age for playing video games comes in the adolescent period, as well as that being the period when the aggression trait develops and increases the most (Moller and krahe, 2009). Anderson and Bushman (2001) further emphasise this point, with "The shooters were students who habitually played Violent Video Games". This implies that the violence of the video games the students played, had an impact upon them leading to them committing a violent crime, which further emphasizes the importance of researching the effects on teenagers to mitigate further violence. Furthermore, Lenhart et. al. (2008) states that 97% of Teenagers play video games of some sort, with 50% having played 'Yesterday', clearly demonstrating just how widespread the video game market is, and how influential it can be.

There is a lot of disagreement whether video games do or do not have an effect on teenagers, and "Despite over 30 studies, researchers cannot agree if violent content video games have an effect on aggression" (Sherry 2001). Bartholow (2002) also proposes that "Evidence of the effects of playing violent video games on subsequent aggression has been mixed".

This paper critically evaluate research by various authors regarding the potential aggression increase or decrease caused by violent video games. This is followed by a conclusions section that will come to scientific conclusions based on the evaluated research, as well as contain mitigation technique suggestions.

2 Current Research

This section of the paper will introduce and discuss different research and evaluate the usefulness in tackling the potential negative impact of violent video games on teenagers' aggression.

2.1 Teenager Aggression

Willoughby et al. (2011) conducted longitudinal research that evaluated extensive violent video game play over the high school years of a group of teenagers. The duration of the test was 3 years, in order to produce the required amount of data to comply with the researches aim, to evaluate an association between Violent Video Games and Aggression in Adolescents.

The Participants were selected from eight different high school, which is good science as it provides a much wider range of backgrounds and participants than if the study was conducted in just a single high school. This study was also part of a Larger study that 'examined youth lifestyle across the high school years' Willoughby et. al. (2011). The study used 1492 participants out of 1771, as the study only included those who completed the survey twice over the duration of the study. Furthermore, the study lists all of the nationalities of the participants, which has a good mix to test most backgrounds against the study. However, the Study does mention and quote frequently that the participants were 50.8% female, which leads to the conclusion that some participants must have been turned down because of their gender and the common stereotyping of that video games are predominantly played by males. Because of this, the validity of the Study is slightly dampened because of the participant selection process.

The experiment of this Study consisted of a questionnaire, which was conducted throughout grades 9-11 on the participants, and asked various

questions targeting specific answers and to get measureable data. The areas in which the Questionnaire got data are as follows; Demographic factors, direct aggression. Violent video game play, non-violent video game play, overall video game play, depressive symptoms, delay of gratification, peer deviance, sports involvement, friendship quality, parent-adolescent relationship quality, parental control and school culture. The questionnaire clearly targeted a wide range of areas about the participants, which clearly gathered a lot of data for evaluation. The study also includes all of the questions asked by the questionnaire, which makes the study much more reliable, as it shows exactly how the data was gathered, and therefore can be repeated if necessary by other researchers to validate its claims.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Frequency of overall VG play 9	—																	
2. Frequency of overall VG play 10	.61	—																
3. Frequency of overall VG play 11	.57	.64	—															
4. Frequency of overall VG play 12	.52	.62	.69	—														
5. Aggression 9	.21	.13	.18	.17	—													
6. Aggression 10	.27	.25	.23	.21	.51	—												
7. Aggression 11	.16	.17	.17	.15	.49	.56	—											
8. Aggression 12	.21	.30	.29	.28	.50	.52	.72	—										
9. Sustained nonviolent VG play	.17	.20	.22	.16	−.09	−.06	−.13	−.11	—									
10. Sustained violent VG play	.55	.59	.62	.58	.21	.29	.23	.34	.13	—								
11. Nonviolent VG play 9	.12	.06	.04	.03	−.09	−.06	−.10	−.13	.70	.00	—							
12. Nonviolent VG play 10	.08	.17	.10	.07	−.06	−.05	−.11	−.09	.75	.04	.38	—						
13. Nonviolent VG play 11	.20	.23	.32	.24	−.05	−.01	−.06	−.04	.67	.18	.27	.43	—					
14. Nonviolent VG play 12	.21	.26	.33	.38	.03	.05	−.01	.05	.61	.26	.26	.39	.51	—				
15. Violent VG play 9	.48	.43	.44	.43	.19	.27	.20	.26	.05	.81	.02	−.01	.08	.16	—			
16. Violent VG play 10	.47	.57	.51	.50	.16	.24	.18	.26	.11	.83	.01	.09	.14	.22	.56	—		
17. Violent VG play 11	.49	.56	.61	.54	.16	.27	.23	.31	.13	.81	−.02	.04	.27	.27	.54	.62	—	
18. Violent VG play 12	.49	.54	.59	.66	.18	.20	.14	.26	.17	.79	.04	.09	.21	.43	.54	.65	.69	—

Figure 1 - The correlation table displaying the Results of the questionnaire Willoughby et . al.(2011)

The results of the study were that there was a correlation of .20 between violent video games and aggression, and mostly under a .10 correlation between non-violent video games. The study also concluded that most violent video games play was seen in males. Furthermore, the Study also went to conduct further research about the frequency and sustained play of such games, which indicated a slight rise in aggression when the participants played violent video games for a sustained time. The study also found that by playing violent video games, aggression went up over time. Therefore, the study concluded that through its 3 different evaluations it had found support for the hypothesis that sustained 'violent video game play leads to increased aggressive behavior over time' Willoughby et. al. (2011).

The Study did display good science in the participation background, yet the participation selection seemed fixed to the extent of the gender selection, as it mentions numerous times the 50.8% of females participating in the Study. However, the initial experiment was conducted well and in a good environment, with experienced research staff monitoring the questionnaire area. The results of the experiment did seem fair and well presented, therefore leading to the conclusion that it was good science. Moreover, the further 2 conducted analysis of the collected data did display an increase of aggression over time, but it could be argued that the initial test found that the correlation of aggression in violent video games was only .20, with the study claiming that .10 correlation of the non-violent games to be insignificant. This would suggest that yes there was a positive aggression increase over time, but the problem was not such a significant one that the Study has much valid claim to aggression increased significantly enough over time to cause a problem. The 2 following tests did use good models for testing an aggression increase, however, so there is still good science to be seen in the Study, with all experiments being conducted professionally.

2.2 Virtual Reality Role

Konijn et. al. (2007) carried out research into the role of wishful thinking, or the idea that the Player wants to act like the virtual characters, to identify if there is any significant correlation between violent video games and aggression. This area of research is of importance in this research paper as it is one that had not been researched before, but that had been commonly linked to aggressive behavior and violent video games, that the player wishes to replicate the virtual character's actions.

The Method of experimentation was a questionnaire 2 weeks before the test was to be conducted, identifying violent video games players and collecting data about the level of aggression before the test was conducted. In the experiment, 12 games were preselected as games to be used through different categories, from violent fantasy to non-violent fantasy, and violent realistic games and non-violent realistic games. These games were selected by a separate group of boys to avoid any potential fixing of the experiment.

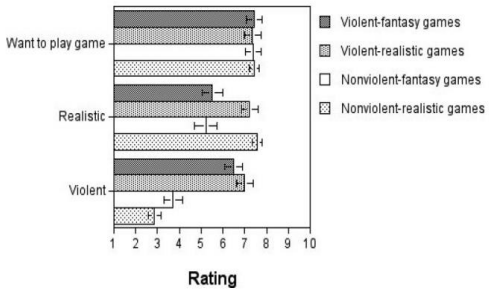

Figure 2 - The results of the preliminary questionnaire Konijn et. al. (2007)

The experiment consisted of a group of 99 boys playing a randomly selected game, and then had to be paired off to press a button faster than their opponent to release a sound that became higher

every time, therefore more uncomfortable for their opponent. The group was told beforehand that any sound in the last 3 levels could cause ear damage.

The experiment's results demonstrated that violent video games and wishful thinking had a very strong correlation to aggression, as did the violent video games. Those participants who played violent video games in the test also registered a higher level of sound level applied to their partner's ears. Furthermore, realism did not affect aggressive behavior, but it did influence wishful thinking. Konijn et. al. (2007) then stated that their results went further than any previous research, as it identified a correlation between violence and low educational ability. However, the experiment did not prove all that was once thought by the Authors. The research indicated no significant correlation between aggressive participants before the experiment, and violent video game characters. The Authors also stated that the chance of repeating the experiment and finding the same result would be 99%.

The research conducted by Konijn et al. (2007) was very fair and contained good science in the sense that all stages of the experiment were designed to contain as little bias as possible, with the author's removing participants that did not understand the test or follow instructions and selected the games to be used with another group of boys. Konijn et. al. (2007) proved exactly what they intended to, and found a link to aggression from wishful thinking.

Through the other areas they tested, such as realism and immersion, they provided other useful results that could be compared and contrasted to wishful thinking, which was good science on their behalf. The researchers also state exactly what games were used and the conditions for their experiments, which gives the research good re-usability and validation as it can be clearly tested again. In conclusion, the research done by Konijn et. al. (2007) is a very good piece of science that can be compared to other useful areas of violent video games to get a broader image of the potential problem.

2.3 Mental Health

Ferguson and Olson (2014) conducted research into the impact of violent video games on bullying within children with Clinically Elevated Depression and Attention Deficit Symptoms. This area of research is of high significance to my research, as it addresses two of the problems associated with violence, which are bullying and delinquency or criminality. Ferguson and Olsen used participants with mental health deficits in order to conduct an experiment that would show results of aggression more significantly if any was detected, as it was thought that the children would be more susceptible to aggression.

Ferguson and Olsen (2014) hypothesised that both the participants with clinically elevated Depression and clinically elevated attention deficit symptoms would exhibit a correlation between violent video games and bullying and delinquency.

The method consisted of 377 children participants, 182 with attention deficit symptoms, and 284 with depression. The mean age of the group was just under 13, which fits well into our research into the effects upon teenagers, and children were chosen from urban and suburban schools with a range of ethnicity. Aggression was measured through a series of questions aimed at retrieving answers that clearly point either to aggressive thoughts or away from it.

In the experiment, Ferguson and Olsen (2014) requested the participants to name 5 games they had played recently, to gain an understanding of the participant's exposure to violent video games through the Entertainment Software Ratings Board (ESRB). The rating is mostly based on violence, which is suitable for a valid result to be obtained. A questionnaire was used to obtain data on the participant's engagement in delinquency in the last 9 months, which was then recorded and analysed. The same process was repeated for bullying, but with different questions aimed at gathering data upon participant bullying. Ferguson and Olsen (2014) state that all research carried out abided by ethical and standards set, and with full parental consent.

The results of the test went completely against the hypothesis of Ferguson and Olsen. They found that the vast majority of participants played video games in the last 6 months, which was to be foreseen, with boys having substantially more exposure to violent video games than girls, as seen in both the researches previously presented in this Paper. However, in both the participants with clinical depression and clinical attention deficit symptoms, only stress and trait aggression affected the predicted outcomes.

Variable	β	95 % Confidence interval	*t* test	Significance
Gender	.06		0.71	.48
Parental involvement	.06		0.70	.49
Stress	.32	(.18, .44)	4.21	.001*
Family/peer support	−.15		−1.69	.10
Trait aggression	.38	*(.25, .50)*	4.23	*.001**
VGV	.04		0.45	.65
VGV × trait aggression	.03		0.39	.70

VGV = exposure to video game violence

Figure 3 - Delinquency regression of participants with clinical elevated attention deficit symptoms Ferguson and Olsen (2014)

Table 2 Bullying regression: beta weights and significance of entered variables for adolescents with clinical elevated depressive symptoms

Variable	β	95 % confidence interval	*t* test	Significance
Gender	−.11		−1.74	.14
Parental involvement	−.01		−0.09	.92
Stress	.23	(.12, .34)	3.24	.001*
Family/peer support	−.05		−0.67	.50
Trait aggression	.28	(.17, .38)	3.74	.001*
VGV	−.07		−0.95	.34
VGV × trait aggression	−.02		−0.23	.82

VGV exposure to video game violence

Figure 4 - Bullying regression of participants with clinical depression Ferguson and Olsen (2014)

Exposure to video game violence in all senses of the research had no significant effect upon the predictive outcome of both delinquent outcomes, or bullying outcomes. Ferguson and Olsen (2014)'s results provided support for the conception that video game violence does not have an effect on children with mental health problems, such as seen in the research conducted. The results are especially important as they demonstrate clear evidence against violent video games impact on violence.

Ferguson and Olsen (2014) present a professional and ethically aware piece of research that clearly demonstrates good, clear results with strong sources of validation. For instance, the study complied with APA standards for ethical human research, as well as a revised bully/victim questionnaire being used to assess the participants previous bullying.

The research conducted is also completely aware of the implications such a result will have, as it expresses that although the results went against the hypothesis, it does not mean that it can undeniably prove that aggression is aggravated by violent video games. The children were especially picked to assess whether adolescent people with mental health issues were more impressionable by violent video games, to which the research provided results against. The research also displays example questions that were used, and the full result data, which can then be easily repeated and validated.

The only element of this research that could be argued to be somewhat bad science is that the mean age of 12.93 is potentially quite low. The mean age does still have importance in my research, however a higher mean age of around 15 or 16 would have been preferred, as then adolescent teens are more aware of the concepts of right and wrong, and about bullying and delinquency. However, the lower mean age of this research does provide a very interesting insight into the potential problem of older adolescent teenagers being affected more by violent video games.

3 Conclusions

As video games become undoubtedly more popular in the near future, it is of the utmost importance that the potential problem of aggression caused by violent video games is addressed promptly. Through evaluating three pieces of research, it is clearly evident age has a part to play, as well as gender and length of exposure to violent video games.

The age problem is clearly seen in Ferguson and Olsen (2014), as their research implies through the mean age of the participants, that violent video games don't have such a negative impact on young teenagers, such as seen here. However, both Willoughby et. al. (2011) and Konijn et. al. (2007) demonstrate in their research a negative impact of violent video games, but both use a mean average of at least 15, 2 years older than Ferguson and Olsen (2014). Therefore, a mitigation technique could certainly be to address the impact of violent video games at a younger age, perhaps 13 or younger based upon the evaluation of Ferguson and Olsen (2014).

Gender also has a very important part to play in mitigation against violent behaviour based on violent video games. Ferguson and Olsen (2014) use a predominantly female participant group, at 62.5%, whereas Willoughby (2011) uses 50.8% females and Konijn et. al. (2007) uses 100% males. However, Willoughby has potential bias in her participation selection process, because of the constant need to prove half of the participants were indeed female, which leads to the conclusion that her test is not completely valid, although the result will be taken into account for arguments sake. Farrar et. al. (2006) finds in her results that "Only gender was significant, with males being more verbally aggressive than females", which only further emphasises my conclusion. Based on evaluation of these papers, it would certainly appear that females are far less prone to aggression, and therefore are less affected by violent video games than what males are. Therefore, another mitigation technique would be to target males more than females to reduce the total aggression seen in teenagers.

Through the evaluation of Willoughby (2011)'s research, you can see the clear aggression increase over time through the last 2 segments of analysis. The fact that my evaluation argues that the impact was not of enough significance in the first analysis to state such a claim, it is irrelevant in this point as there is a clear undeniable increase in the amount of participant aggression seen over time. This is

supported by Anderson et. al. (2008) as seen in their longitudinal study's results, that habitually playing video games increases physical aggression in time. Wallenius (2008) also finds in his longitudinal research that there is a direct link between game violence and aggression, when time is applied as a factor. Therefore, prolonged exposure to violence in video games can have a serious negative impact of aggression not just on the individual, but on a wider group such as a friendship group, for example. Quite clearly this is a serious problem that could be addressed through the early intervention and prevention of violent video games at a young adolescent age, and possibly even tightening on violent video game age restrictions.

References

Anderson, C.A. and Bushman, B.J., 2001, 'Effects of Violent Video Games on aggressive behaviour, aggressive Cognition, aggressive affect, physiological arousal, and Presocial behavior: A Meta-Analytic review of the scientific literature', *Psychologial Science*, 12(5): pp. 353-359

Anderson, C.A., Sakamoto, A., Gentile, D.A., Ihori, N., Shibuya, A., Yukawa, S., Naito, M., Kobayashi, K., 2008, 'Longitudinal Effects of Violent Video Games on Aggression in Japan and the United States', *Pediatrics*, Vol.122, p1067-1072

Bartholow, B.D., Anderson, C.A., 2002, 'Effects of Violent Video Games on Aggressive Behaviour: Potential Sex Differences', *Journal of Experimental Social Psychology*, Vol.38, p283-290

Farrar, K.M., Krcmar, M., Nowak, K.L., 2006, 'Contextual Features of Violent Video Games, Mental Models, and Aggression', *Journal of Communication*, Vol.56, p387-405

Ferguson, C.J., Olson, C.K., 2014, 'Video Game Violence Use among "Vulnerable" Populations: The Impact of Violent Games on Delinquency and Bullying Among Children with Clinically Elevated Depression or Attention Deficit Symptoms', *J Youth Adolescence*, Vol 43, p127-136

Konijn, E.A., Bijvank, M.N., Bushman B.J., 2007, 'I Wish I Were a Warrior: The Role of Wishful Identification in the Effects of Violent Video Games on Aggression in Adolescent Boys', *Developmental Psychology*, Vol.43, No.4, p1038-1044

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A.R., Evans, C., Vitak, J., 2008, 'Teens' gaming experiences are diverse and include significant social interaction and civic engagement', *Pew Internet & American Life Project*

Moller, I., Krahe, B., 2009, 'Exposure to Violent Video Games and Aggression in German Adolescents: A Longitudinal Analysis', *Aggressive behaviour*, Vol.35, p75-89

Sherry, J.L., 2001, 'The Effects of Violent Video Games on Aggression: a Meta-Analysis', *Human Communication Research*, Vol.27, p409-431

Wallenius, M., Punamaki, R., 2008, 'Digital Game Violence and direct aggression in adolescence: a longitudinal study of the roles of sex, age, and parent-child communication', *Journal of Applied Developmental Psychology*, Vol.29, p286-294

Wei, R., 2007, 'Effects of Playing Violent Videogames on Chinese Adolescents' Pro-Violence Attitudes, Attitudes Toward Others, and Aggressive Behaviour', *CyberPsychology & Behaviour*, Vol.10, No.3

Willoughby, T., Adachi, P.J.C., Good, M., 2011, 'A Longitudinal Study of the association between violent video game play and aggression among adolescents', *Developmental Psychology*, Vol.48, No.4, p1044-1057

Evaluation of Current Computing Research Aimed at Improving Fingerprint Recognition Systems

Shaun Nkgasapane

Abstract

Fingerprint recognition systems have become one of the most used biometrics for authentication on computerized systems. This paper provides a critical evaluation of current research aimed at improving fingerprint recognition systems, three main methods are evaluated wavelet based approach to ridge thinning in fingerprint images, image descriptor method for fingerprint liveness detection and minutia tensor matrix algorithm as well as the fingerprint recognition model. Comparison of methods is carried out to determine the most effective method for improving fingerprint recognition systems and recommendations are made regarding the most suitable method to use

1 Introduction

Most fingerprint recognition systems focus on minutiae (microscopic feature of a fingerprint) as attributes for comparing fingerprints, nonetheless challenges arise during the extraction stage of minutia fingerprint segmentation. This stage partitions a fingerprint picture into forefront and background, two kinds of errors can emerge: "forefront can be labeled as background and attributes like minutiae can be lost or background can be categorized as forefront and spurious attributes may be introduced" Gottschlich et al. (2012).For this reason they mentioned that it is necessary to have a method that handles both errors. Several issues related to the vulnerability of fingerprint recognition systems attacks have been stressed in the biometrics literature. Marasco and Ross (2014) explained that such vulnerability entails the use of artificial fingers, where materials such as silicone, gelatin and play-doh are inscribed with fingerprint ridges. Consequently fingerprint recognition systems can be manipulated when these artificial fingers are placed on the sensor, the system will then process subsequent fingerprint images, thereby allowing an invader to forge the fingerprints of another individual. Research is being carried out to improve fingerprint recognition systems to make them more secure by detecting the liveness of fingerprint using Wavelet-Markov local descriptor method, Gragnaniello et al. (2014).

A small number of microscopic features of a fingerprint that is partially showing is still a challenge in matching fingerprints, as minutiae based algorithm will detect few minutiae .As a result the recognition rate of the fingerprint identification would diminish thereby failing to identify fingerprints of a particular individual. Partial fingerprints can occur as a result of an individual incompletely touching the fingerprint scanner and this yields to a fingerprint matching challenge, Chen et al. (2014).Additional research based on improving partial fingerprint matching is being carried out using fusion scheme, based on enhanced support vector machine Girgis et al. (2015).

Li et al. (2012) articulated that a fingerprint template protection scheme called Fingerprint Fuzzy Vault is prone to cross matching attacks. In the sense that the exact finger can be registered for different applications .To address this problem research is currently done to provide a more secure scheme called Cancellable and Fuzzy Fingerprint. It is clear that indeed fingerprint recognition systems need to be improved, Hu and Wang (2012).

Therefore the purpose of this paper is to critically evaluate current computing research papers specifically on schemes, techniques or methods and models used to improve fingerprint recognition systems and reach useful scientific conclusions. Different schemes/techniques and models that have been used to improve fingerprint recognition systems are evaluated, such as Wavelet based Approach to Ridge Thinning in Fingerprint Images, Image Descriptor for Fingerprint Liveness Detection, Minutiae Tensor Matrix, a newly found method for matching fingerprints and Standardized fingerprint model utilized in improving fingerprint recognition.

2 Wavelet-based Approach to Ridge Thinning in Fingerprint Images

Flaws in ridges of fingerprints yield to errors in resolving the position coordinates of true microscopic features of a fingerprint together with

their corresponding arrangement in the image. As a result of flaws in thinned ridges of the fingerprint, methods used to extract microscopic features are bound to omit some true minutiae while selecting false points Fang et al. (2005). Subsequently they presented a method for thinning ridges centered on the local minima of wavelet transform moduli. According to their claims, thinning algorithm enhances the ridges of the fingerprint's quality and structure, making thinning outcomes good for withdrawal of minutiae.

Prior to the experiment, they mathematically analyzed the wavelet minima to ensure that the desirable features are especially conducive to detect and process the fundamental skeletons of the fingerprint image. To test the algorithm they used a procedure called multiscale algorithm which eliminates unnecessary features of the image in the background. For every single image they unsystematically selected a unit measure of the wavelet method and obtained the matching frame of the original ridge and processing the wavelet minima. Furthermore they compared the execution time of the method with three other methods namely; Zou's method, Zhang and Suen's method as well as the principal curve oriented method. In comparison of these methods they tested images that were unsystematically chosen from the information repository. The matching minutiae sets were calculated based on thinned ridges extracted from the three mentioned methods.

Claims were made that the wavelet based algorithm is applicable to grey level images but many previous methods failed to do ridge thinning in grey images. They also articulated that the method they proposed improves thinned ridges structure and makes thinning outcomes conducive for withdrawing minutiae.

A well detailed procedure of the proposed thinning algorithm is presented, step by step. It is very clear what they aimed to achieve and did a meticulous work of mathematically analyzing the wavelet minima before testing the algorithm. They backed up their discussions with diagrams and equations and this aided in understanding their explanations better. Fang et al. (2005) mentioned that they compared their proposed method with other three methods of different authors in relation to the execution time (how long it takes for the method to get executed).Nonetheless they did not mention the performance of other factors, for instance how well the quality of fingerprint image is, as compared to this three algorithms, potential problems that may arise when applying the proposed algorithm and how this problems can be mitigated. It could have been beneficial if this factors were covered hence this would have aided in having a fully justified claim Therefore we can conclude that the wavelet based algorithm is able to do ridge thinning in grey images which many traditional methods filed to do and dispute the claim that it enhances quality of the thinned ridge structure than other methods since there were no results shown pertaining the improved quality.

3 Image Descriptor Method for Fingerprint Liveness Detection

Kim and Jang (2016) proposed a unique image descriptor method for fingerprint liveness detection. This method distinguishes a true fingerprint from a fake one. The inspiration to propose such method emerged from the fact that user verification via fingerprints is prone to spoofing attacks for example, fabrication of fingerprints utilizing substances like gelatin and silicone. The main focus of this method is to obtain the local textural patterns processed from the accumulated smoothing space as their attributes.

Their method was employed by building the smoothing space from the initial captured fingerprint pictures by edge-aware filtering. To achieve this, they used a rapid global smoothing technique because it prominently retains the directional structure throughout the smoothing process and executes fast as compared to other methods. To stress the dissimilarity between live and fake fingerprints they proposed to gather multiple outputs computed from various smoothing parameters. Figure 1 illustrates the difference between live and fake fingerprints on proposed smoothing space.

Figure 1-Difference between live and fake fingerprint Kim and Jang

i) The image labeled (a) illustrates the original captured image (left: live, right: fake).

ii) The image labeled (b) illustrates proposed smoothing space for fingerprint images (left: live, right: fake)

iii) The image labeled (c) illustrates magnified versions for white rectangle regions in image (a) and (b)

The testing was based on two representative datasets, LivDet 2009 and LivDet 2011, in evaluating the quantitative performance for fingerprint liveness

detection, they directly employed the data structure for training. Furthermore to demonstrate the efficiency and robustness of the method, they compared their Local Accumulated Smoothing Patterns (LASP) descriptor with three representative local descriptors widely utilized for liveness detection namely, Local Binary Pattern, Local Ternary Pattern and Local Speed Pattern. Results of their testing showed that their testing yielded better performance as compared to other methods presented in two datasets. They claimed that their proposed method can be applied to prevent malicious spoofing attacks launched by using spurious fingerprints. Table 1 and 2 demonstrates results performance of the proposed method (image descriptor method for fingerprint liveness detection) against other methods in two datasets.

Methods	LBP [2] (%)	LTP [3] (%)	LSP [4] (%)	Proposed (%)
Biometrika	74.74	83.00	71.32	90.01
CrossMatch	88.05	83.03	87.80	87.72
Identix	83.58	89.29	85.42	87.76
Average	82.12	85.11	81.51	88.49

Table 1-Detection performance in the LivDet 2009 database Kim and Jang (2016)

Methods	LBP [2] (%)	LTP [3] (%)	LSP [4] (%)	Proposed (%)
Biometrika	70.95	65.40	69.60	77.40
Digital	72.80	71.70	75.35	72.90
ItalData	77.80	71.85	64.85	82.40
Sagem	80.99	79.76	77.46	82.42
Average	75.64	72.18	71.82	78.78

Table 2-Detection performance in the LivDet 2011 Kim and Jang (2016)

Considerate effort was devoted in conducting research and benchmarked datasets were used when comparing image descriptor method for fingerprint liveness detection with other methods. The research was based on multiple research papers which were relevant to the subject at hand (detecting a fake fingerprint from a live one).Formulas used were comprehensive and well explained. Nonetheless it is not clear how the experiment was carried out They just mentioned that the two datasets were used in comparing methods and how many images each dataset had. Immediately after that performance results were outlined in tabular form. A brief description should have been given on how they managed to attain results because repeatability factor in experiments is very crucial, as it ensures that results are not biased. For instance one may want to test the same method but if it is not clear how the experiment was done, a claim made seems to be invalid. Therefore a conclusion can be made that this claim is not fully justified as it has to be clear how the experiment was really done.

4 Minutia Tensor Matrix

Feng and Fu (2015) aimed to establish correlations among microscopic sets of fingerprints, hence fingerprint matching is a challenge due to various factors nonlinear distortion, noise and partial overlap. They proposed a new algorithm for matching fingerprints using the idea of minutiae tensor matrix (MTM).They stated that within MTM, crosswise or slanting components depict resemblance of minutia pairs. Furthermore they mentioned that right minutiae pairs are bound to generate considerable correspondences and compatibilities, consequently they form a dense block .Figure 2 below shows the correspondence graph on the left and the tensor matrix on the right.

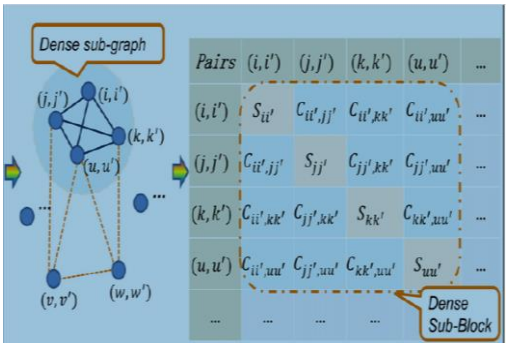

Figure 2 Feng and Fu (2015)

To test their fingerprint matching algorithm they designed two distinct kinds of minutiae tensor matrix, local MTM and global MTM. Regarding the local matching level, local MTM was generated and a unique local correlation computation was advanced. As for the global matching level, global MTM was established to compute correlations of the whole minutiae sets. When comparing results they used Fingerprint Verification Competition database (FVC2004-DB1) so as to test the radius of the Local Minutiae Topologic Structures (LMTS).They varied the radius size from 30 pixels to 160 pixels. Thereafter they observed the relationship between Equal Error Rate (ERR), average time and the LMTS radius. Figure 3 demonstrates results obtained.

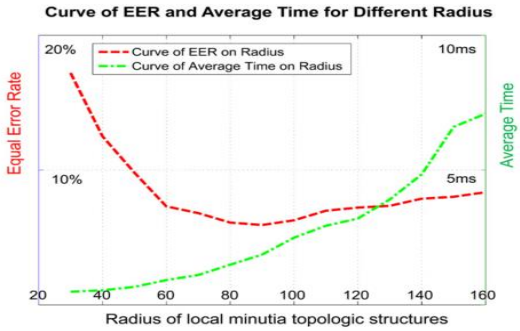

Figure 3 Feng and Fung (2015)

Figure 3 shows that average time grows with the length of the radius and the MTM is influenced by LMTS radius. Initially ERR drops and then escalates together with the radius as the fingerprints shoe rigidity in local region and non-linearity in global region.

They claimed that their new proposed algorithm has stronger description aptitude and enhanced robustness to noise and non-linearity. To back up their claim they stated that results of the experiment they carried on Fingerprint Verification Competition database (FVC 2002 and FVC 2004) shows effectiveness and efficiency.

Research that was carried out by Feng and Fu (2015) was based on considerable insights of other researchers. The method used for the experiment was scientifically sound and logical as relevant equations were used and thoroughly explained. It was also good that they used bench marked datasets to conduct their experiment. Though they claimed that their algorithm has a better description ability and enhanced robustness to noise and nonlinearity it is not clear how that is the case. It could have been beneficial if both results were presented, hence figure 3 only shows results of their experiment. For instance they could have also presented results of what they were testing against so that the distinction between the two becomes clearer. For this reason results seem not to be fairly reported, therefore we can conclude that these claims are not fully justified by the evidence.

5 Standardized Fingerprint Model

Tam and Thai (2010) proposed a standardized fingerprint model to enhance the precision of fingerprint recognition systems. This model integrates fingerprints that that represent every fingerprint template stowed in the information repository when matching.

The model they proposed involved: preprocessing of a fingerprint image where for every single image, thinned ridge lines and extraction of minutiae are identified. Secondly parameter sets are discovered and modified thereafter a fingerprint with the largest area is selected as a mean image then fingerprints are integrated by recalculating parameters. Finally the post processing step assist in eliminating noise and the synthesizing stage is executed.

To test the efficiency of this model they conducted an experiment using the DB4 FVC2004 database hence most of the fingerprint images in this databases are of poor quality. The database contained eight hundred fingerprints that belonged to hundred fingers. So the estimation of the scope of data was grounded on the experiments of the first hundred pairs of fingerprints in the dataset. Seventy nine thousand matchings among varying pairs we done to calculate the delivery of fake matchings where for every single image, fingerprints obtained from various fingers were unsystematically picked from the database. Experiment results were compared to another results based on the approach of Xiping Luo, 2000.Table 3 and 4 demonstrates the results obtained.

Comparison of their approach and Luo's

	Our approach		Xiping Luo, 2000	
	GAR	FAR	GAR	FAR
1	100%	7.14%	100%	39%
2	98.99%	7.12%	99%	32%
3	97.98%	3.57%	97%	25%
4	96%	3.57%	94%	25%
5	95%	3.57%	90%	14.28%

Table 3 Tam and Thai (2010)

Top 10 matchings of their approach and Lou's

#	Our approach	Xiping Luo, 2000
1	98.3%	92.3%
2	98.9%	93.6%
3	99.1%	94.1%
4	99.1%	94.1%
5	99.1%	94.1%
6	99.2%	94.6%
7	99.3%	94.9%
8	99.4%	95.2%
9	99.4%	95.2%
10	99.4%	95.2%

Table 4 Tam and Thai (2010)

Tam and Thai (2010) claimed that the proposed fingerprint model improves the accuracy of fingerprint recognition systems, can correctly recognize fingerprint images of poor quality.

The data used to test the model was considerable hence they used the database that contained 800 fingerprints that belonged to 100 fingers. The fact that fingerprints obtained from various fingers were randomly chosen from the database, displays a good

aspect of a sound science. The methodology they used to carry out the experiment was remarkable as in every processing stage equations and illustrations used to back up their claim were thorough, well detailed and concise. Furthermore results of the experiment are fairly reported as there is solid evidence that backs up their claim. Moreover, comparison of results made are significantly visible and proceeds to concur with the claim made that indeed the proposed fingerprint model enhances the accuracy of fingerprint recognition systems, making it possible to even recognize images of poor quality. Therefore we can conclude that the claim made by Tam and Thai (2010) is valid.

6 Conclusions

This research paper provides thorough analysis of current methods, algorithms and models to improve fingerprint recognition systems. Fingerprint recognition systems are faced with many challenges that sometimes makes it difficult to recognize fingerprints due to various factors like flaws of the ridge map, non-liner distortion, noise and partial overlap of fingerprints. Furthermore the art of attacking this systems has gradually grown over the years this includes the use of substances like gelatin and silicone to bypass fingerprint recognition systems.

Methods and algorithms to alleviate such problems are evaluated in this paper. The wavelet based approach to thin ridges of fingerprints for suitable extraction of minutia is vigorous and a detailed procedure for fingerprint is presented the method succeeded in thinning fingerprints though it could not improve the fingerprint quality. Whereas the image descriptor method aimed at detecting liveness of a fingerprint to distinguish between a fake and a real finger. Though the results of the experiment illustrated that the method was lucrative it was not clear how the experiments was done to arrive at such results. The minutiae tensor matrix method was proposed with the aim of establishing correspondences between minutia sets by dealing with nonlinear distortion, noise and partial overlap of fingerprints. This algorithm was claimed to have stronger description efficiency and robustness to noise and non-linearity. Nevertheless there was no comparison of results made. On the other hand the fingerprint recognition model was proposed and it was claimed improve accuracy of fingerprint recognition systems by correctly identifying fingerprints of poor quality. The claim made is valid as there is concrete evidence in their experiments, results from both sides were presented and the results were fairly and accurately presented. Therefore fingerprint recognition model is more effective in correctly identifying fingerprints of poor quality as compared to some methods. It is appropriate to utilize this model to improve fingerprint recognition systems.

References

Chen, F., Li, M. & Zhang, Y., 2013. 'A Fusion Method for Partial Fingerprint Recognition'. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(6), pp. 5-13.

Fang, B., He, Z., Tang, Y. Y. & You, X., 2005. 'A Wavelet-based Approach to Ridge Thinning in Fingerprint Images'. *International Journal of Pattern Recognition and Artificial Intelligence*, 19(5), pp. 631-645.

Fu, X. & Feng, J., 2015. 'Minutia Tensor Matrix: A new Strategy for Fingerprint Matching'. *PLOS ONE*, 10(5), pp. 89-100.

Girgis, M., Mansour, R. & Sewisy, A., 2015. 'A Robust Method for Partially deformed fingerprints Verification using genetic algorithm'. *Expert Systems Applied*, 36(2), pp. 200-230.

Gottschlich, C., Huckemann, S. & Thai, D. H., 2012. 'Filter Design and Performance Evaluation for Fingerprint Image Segmentation'. *PLOS ONE*, 14(10), pp. 154-160.

Hu, J. & Wang, S., 2012. 'Alignment-Free Cancellable Fingerprint Template Design'. *A densely Infinite-to-one Maping Approach*, 3(12), pp. 50-62.

Kim, W. & Jung, C., 2016. 'Local accumulated Smoothing Patterns for Fingerprint Liveness Detection'. *Institution of Engineering and Technology*, 52(23), pp. 1912-1914.

Li, C., Yang, W. & Xi, K., 2012. 'Engineering and Information Technology'. *A Cancellable and Fuzy Fingerprint Scheme for Mobile Computing*, 10(7), pp. 149-160.

Marasco, E. & Ross, A., 2014. 'A Survey on Antispoofing Schemes for Fingerprint Recognition Systems'. *ACM Computing Surveys*, 47(2), pp. 80-95.

Poggi, G., Sansone, C., Verdoliva, L. & Gragnaniello, D., 2014. 'Wavelet-Markov Local Descriptor for Detecting Fake Fingerprints'. *Electronic Letters*, 50(6), pp. 439-441.

Tam, H. N. & Thai, L. H., 2010. 'Fingerprint Recognition using Standardized Fingerprint model'. *International Journal of Computer Science Issues*, 7(3), pp. 1-7.

A Critical Evaluation of Current Research into Improving Botnet Detection Rates

Andrew Thompson

Abstract

Botnets pose a serious cyber-security threat to all internet users due to the malicious damage they can cause, ranging from DDOS attacks to email spam. The detection of these botnets is therefore essential. This paper has focused on critically analysing and evaluating three pieces of current research aimed at detecting Peer-to-Peer botnets focusing specifically on the detection rates achieved. Additional research papers are used to back up this evaluation. The paper then provides a comparison of the evaluated techniques before conclusions on future research and a decision on which approach to botnet detection is the most promising is provided.

1 Introduction

Botnets have evolved continuously in capability and complexity since the advent of the first malicious botnet, Pretty Park, in 1999. (Elliott, 2010).

Initial detection methods had focused on exploiting the Command and Control architecture used by botnets. With more botnets evolving to avoid these detection techniques via HTTP and Peer-to-Peer communication (Wang and Paschalidis, 2016) new detection methods have been devised.

Sheng et al. (2012) conducted research on detecting botnets via Traffic Flow Analysis. Whilst the proposed detection method can detect a variety of botnets, the low detection rate and high false positive rate shown when detecting Peer-to-Peer botnets means that it will not be evaluated in this paper.

The work by Muthumanickam and Ilavarasan, (2012) proposed a Peer-to-Peer detection method based on Host and Network level analysis, the paper however did not conduct any testing on the proposed method and thus will not be evaluated in this paper.

The following survey paper will focus on critically analysing and evaluating botnet detection techniques that detect non-conventional botnets utilising Peer-to-Peer protocols. Papers will be analysed by initially presenting the work done by the researchers before evaluating the experiments conducted, results and conclusions reached.

2 Botnet Detection Methods for Peer-to-Peer Botnets

The following section will analyse research papers aimed at improving Peer-to-Peer botnet detection. Each method will be discussed and evaluated before conclusions are drawn.

2.1 Botnet Detection via Machine Learning

Research by Singh et al. (2014) aimed to improve Peer-to-Peer botnet detection by utilising machine learning. The proposed system utilised a number of existing technologies to achieve its goal.

The proposed system contained three unique elements; a ‘traffic sniffer’ module used to capture and extract data from packets travelling through a network. A feature extraction module used to extract key data from the information collected in the ‘traffic sniffer’ module. And a machine learning module, utilising a Random Forest approach to analyse the data to identify botnet traffic.

Singh et al. (2014) tested their detection method by initially creating a testbed which mirrored traffic on the campus’ network. Standardised data from the Center for Applied Internet Data Analysis (CAIDA) and packet captures from known botnet attacks were injected into the network via TCP Replay.

The proposed detection method was run on the testbed a total of 84,030 times. 90% of the tests formed a training set used to improve the machine learning algorithm and 10% were used as the testing set.

The researchers stated that the classifier had an accuracy of 99.7%, the results of testing can be found in table 1. Figure 1 shows performance of the Random Forest approach in relation to other machine learning algorithms.

Other research into botnet detection via machine learning, such as work by Kirubavathi and Anitha (2016) provided comparisons of their method against various alternatives allowing them to conclude on how well their system performed when compared to existing detection techniques.

Accuracy measure of the classifier.

True positive rate	False positive rate	Precision	Recall	Class
0.998	0.003	0.999	0.998	Malicious
0.997	0.002	0.996	0.997	Non-malicious

Table 1 Results of testing the proposed botnet detection method (Singh et al., 2014)

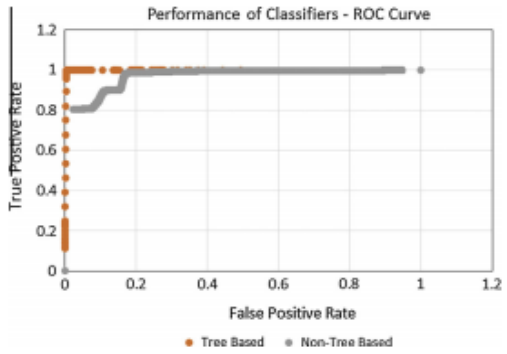

Figure 1 Performance of used algorithm vs alternatives (Singh et al., 2014)

Singh et al. (2014) concluded that the work conducted made three key contributions to botnet detection. A method for capturing packets in high bandwidth scenarios, a framework for analysing captured packets and a detection module for detecting Peer-to-Peer botnets. The paper also concluded that they suffered from packet drops with one of the technologies used but did not address this issue as it would have been outside the papers scope. The researchers also stated that a recent botnet, Stuxnet, communicates stealthily. Following this the researchers proposed future research aimed at exploring stealthy botnet communication.

The research done by Singh et al. (2014) provides justifications for each of the technologies used within the proposed detection method, showing that the research was carried out in a logical and thorough manner. All tests conducted within the paper have an accompanying table or figure, demonstrating good transparency of the research conducted. The use of standardised data from CAIDA and historical botnet attacks ensures that the tests are as repeatable as possible allowing for independent verification of the methods results.

The paper did compare the performance of the machine learning algorithm against alternative algorithms but no performance comparison against other botnet detection techniques was conducted. This prevented the researchers from concluding whether their method had improved botnet detection.

Overall the research by Singh et al. (2014) was excellent, demonstrated by the high detection rate and low false positive rate achieved. The research achieved both a higher detection rate and lower false positive rate than an earlier piece of research into machine learning based botnet detection conducted by Saad et al. (2011), proving the advancements made by the paper. However, the packet loss documented in the papers conclusion is a cause for concern and shows that the method requires more research.

2.2 Botnet Detection via Fuzzy Inference

Soniya and Wilscy (2014) conducted research into improving botnet detection via Fuzzy Inference. The proposed method would classify destination devices as either benign or malicious after examining the traffic generated by the device.

The detection method is comprised of five unique elements, a diagram of the overall system can be seen in figure 2.

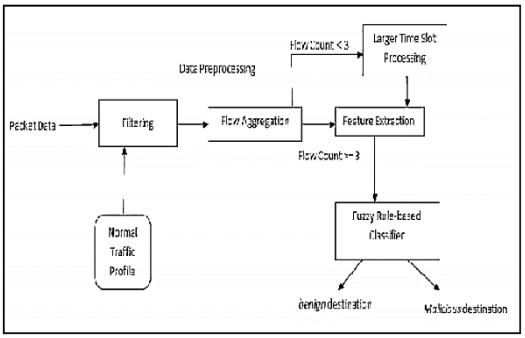

Figure 2 Architecture of the proposed botnet detection system (Soniya and Wilscy, 2014)

Soniya and Wilscy's (2014) proposed detection method collects and processes data in thirty minute intervals. Within processing, a manually created filter is used to prevent benign locations from been flagged as malicious. The filtered packets are then grouped in flow aggregation. The detection method also contains a feature extraction module similar to that used by Singh et al. (2014) that extracts data from the collected packets. Extracted data is

processed by a Fuzzy rule-based classifier which is used to determine whether the device sending the data is malicious or benign.

Soniya and Wilscy's (2014) proposed detection method was tested against ten datasets. Seven of which contained botnet traffic and the remaining three contained benign traffic. The results of testing the proposed method were then compared against two separate botnet detection methods. The results of the tests and the comparison table can be seen in table 2 and table 3 respectively.

The research concluded that the performance of the proposed system proves that a Fuzzy Rule based system for botnet detection is viable but that botnets could theoretically randomize traffic features as a way to evade detection. The researchers go on to recommend that future research should be aimed at detecting botnets that utilize such evasion techniques.

TABLE I. DATASETS

Results Sl. No:	Dataset Name	Detection Rate	False Positive Rate
1	Bot Data	97.64	
2	User-Induced Data		0.8
3	LBNL		1.4
4	DARPA		1.1
5	Banbra	77	
6	Bifrose	84	
7	Dedler	83	
8	Sasfis	45	
9	Koobface	40	
10	Ramnit	72	

Table 2 Proposed detection method test results (Soniya and Wilscy, 2014)

TABLE II. PERFORMANCE COMPARISON

Method	CoCospot[21]	Zhao, David, et al [22]	Proposed FRB bot detection
Detection Rate (DR)	88%	98.30%	97.64%
False Positive Rate (FPR)	0.10%	0.01%	0.80%
Novel Detection of Bots	N/A	DR =100% FPR = 82%	DR = 40-84% FPR = 1.4%

Table 3 Performance comparison of botnet detection methods (Soniya and Wilscy, 2014)

Soniya and Wilscy's (2014) work describes and justifies both the use of a Fuzzy Logic classifier to mark whether traffic is seen as benign or malicious and the Large Time Slots Classifier to detect botnet communication over large time intervals in a scientific manner. However, they appear to neglect a key part of the proposed system; Data Preprocessing, in that they do not define how the data is captured and what technologies are used to perform filtering, flow aggregation and feature extraction. This limits the research's repeatability.

In terms of testing the proposed system, the utilization of multiple datasets that contained either traffic from unique botnets or standardised network traffic was appropriate. Whilst the research states it performs a 3-fold cross validation on datasets 1 and 2, the research fails to state how many times the test was repeated on the remaining datasets, due to this the reliability of their claims relating the remaining datasets is questionable.

The comparison of detection methods allowed the researchers to infer whether the performance of their detection method was good or not; this justifies the use of a comparison. However, the comparison states that the CoCospot method can detect 88% of botnet traffic with a false positive rate of 0.1% this is contradicted by the CoCospot research paper which states it can detect a higher amount; "CoCoSpot can recognize more than 88% of the C&C flows at a false positive rate below 0.1%." (Dietrich et al., 2013) this may be indicative of bias in that the researchers have purposely reduced the results of a separate piece of research in order to make their proposed method appear superior. This does not however invalidate the conclusions reached by the researchers.

In conclusion, the work by Soniya and Wilscy (2014) is justified, both through their own testing and the fact that both the detection rate and false positive rate have been improved when compared to an earlier paper on botnet detection that utilizes Fuzzy Inference (Wang et al., 2011). despite the improvements made the method still needs further research conducted as the paper states it was outperformed in some areas by the work conducted by Zhao et al. (2013)

2.3 Botnet Detection via Negative Reputation Systems

Yahyazadeh and Abadi (2015) proposed improving botnet detection by utilising a reputation system to assign a negative reputation score to host devices located within a singular monitored network.

A score is assigned to each host within the network after an analysis of the host's historical participation in coordinated group activities has been conducted to determine whether any host in question is part of a botnet or not.

The proposed system consisted of five modules; base filtering, flow clustering, negative reputation, malicious activity detection and botnet detection.

The overall structure of the proposed detection method can be seen in figure 3.

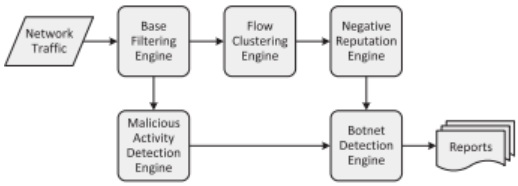

Figure 3 Architecture of the proposed botnet detection system (Yahyazadeh and Abadi, 2015)

protocols. The two datasets collected were combined to create the dataset used for testing the proposed botnet detection method.

The combined dataset of benign and malicious traffic was injected into a test network to test the proposed method's ability to detect malicious botnets. The results of the tests can be found in table 5.

Referring back to table 4 it can be seen that E01, E02 and E03 represent the type of botnet tested. The researchers created a table to compare the proposed

Characteristics of network traces collected from the experiments.

Experiment ID	E01	E02	E03
Number of packets	41,258	36,197	39,741
Duration	160 min	110 min	120 min
Number of hosts	83	94	77
Botnet type	HTTP-based	IRC-based	P2P-based

Table 4 Packets captured for each botnet type (Yahyazadeh and Abadi, 2015)

Experiment ID (%)	E01		E02		E03	
	GA[a]	GA+MA[b]	GA	GA+MA	GA	GA+MA
Detection rate	95.26	100	95.98	100	92.11	97.26
False alarm rate	1.92	2.11	1.41	1.83	2.02	2.31

[a] Coordinated group activities.
[b] Coordinated group activities as well as malicious activities.

Table 5 Test results of the proposed botnet detection method (Yahyazadeh and Abadi, 2015)

Comparison of BotGrab with other well-known botnet detection techniques.

Detection technique	Unknown botnet detection	Early stage detection	Protocol & structure independent	Robust to command encryption	Online detection	Low false alarm rate
Rishi [12]	No	Yes	No	No	Yes	No
BotHunter [14]	No	Yes	Yes	No	Yes	Yes
BotSniffer [15]	No	Yes	No	No	Yes	Yes
BotMiner [16]	Yes	No	Yes	Yes	No	Yes
Castle [17]	No	No	Yes	No	No	Yes
Wurzinger [18]	No	No	Yes	No	Yes	Yes
Yu [19]	Yes	Yes	No	Yes	Yes	Yes
BotGrab	Yes	Yes	Yes	Yes	Yes	Yes

Table 6 Comparison of proposed detection method vs alternative techniques (Yahyazadeh and Abadi, 2015)

To test the proposed method Yahyazadeh and Abadi (2015) created a virtualized environment containing several Virtual PC's running two distinct operating systems; Windows XP and Linux Ubuntu. Each virtual host was infected with three botnets TRiAD a HTTP botnet, rBot an IRC botnet and Immonia a Peer-to-Peer botnet to ensure that each command and control protocol was tested against the proposed detection method.

Once all virtualised hosts had been set up, each botnet was commanded to perform three functions that demonstrated botnet behavior; send basic information about their hosts, sleep and lastly perform some malicious activity. The results of the packet capture can be seen in table 4.

Additionally, packet capture was conducted on the campus network to provide a dataset of normal network traffic that included commonly used

method with seven alternative botnet detection methods, shown in table 6.

The research by Yahyazadeh and Abadi (2015) concluded that botnet detection is a challenging area for research and that most of the detection techniques proposed previously suffer from certain limitations. The researchers state that the proposed detection method addresses the limitations found in the alternative techniques. Yahyazadeh and Abadi (2015) go on to state that the experimental results show that the proposed detection method can detect a high amount of botnet traffic whilst maintaining a low false alarm rate. Finally, the researchers stated that to improve upon their system they would add a host-based negative reputation engine to potentially detect bot infected hosts before any botnet activity is conducted.

The researchers started the description of their proposed botnet detection method by providing a detailed description of each engine (see figure 3) used to form the proposed botnet detection method. These descriptions improve the repeatability of the research as they allow independent researchers to recreate the detection method more precisely.

In terms of testing the proposed method, the use of a virtualized environment to create botnet traffic was the optimum choice as it increased the repeatability of the experiments conducted. Additionally, the use of real world botnets to create the packet traces was appropriate. However, the reliability of the experiment could have been improved had the proposed method been tested against a larger range of botnets. This approach has been seen in a number of botnet research papers such as Singh et al. (2014) who tested their proposed method against five botnets and Soniya and Wilscy (2014) who tested their method against eight botnet data sets.

The creation of a comparison table by Yahyazadeh and Abadi (2015) allowed the researchers to demonstrate the range of advantages their detection method had brought over the alternative methods. However, the use of a polar question for the low false alarm rate comparison is questionable as qualitative data could have been used which would have increased the accuracy of the comparison, this may be indicative of bias by the researchers not wanting to display that the proposed method had a higher false alarm rate than the alternatives compared. Additionally, the researchers do not state what constitutes a low false alarm rate which reduces the transparency of the comparison. This potential bias does not affect the results of the detection method and thus does invalidate the research.

To conclude, the test results of the detection method by Yahyazadeh and Abadi (2015) show that it can detect Peer-to-Peer botnets effectively whilst maintaining a relatively low false detection rate. Future research should be directed at ensuring that this method can maintain its detection rate against a variety of Peer-to-Peer botnets.

3 Comparison of Detection Methods

As this papers focus is on the detection rates achieved by the evaluated papers, the primary method for comparing the papers is a comparison table of both the detection rate and the false positive rate of each detection technique. This comparison is documented in table 7.

Detection technique	Detection rate	False positive detection rate
Singh et al. (2014)	99.70%	0.30%
Soniya and Wilscy (2014)	97.64%	0.80%
Yahyazadeh and Abadi's (2015)	97.26%	2.31%

Table 7 Comparison of detection methods

The table makes it clear that the detection method by Singh et al. (2014) shows the best test results for both detection rates and false positive rates, suggesting that this machine learning approach to botnet detection offers the highest detection rates. However, the packet loss issue discussed by Singh et al. (2014) in their conclusion does question the overall performance of the method as a proportion of the dropped packets may have been malicious and thus avoided detection.

Neither Soniya and Wilscy (2014) or Yahyazadeh and Abadi (2015) stated they suffered from any packet loss issues. It is possible that this is because both techniques utilised a filtering system which removed packets from trusted sources such as Google or Yahoo. The proposed method by Singh et al. (2014) has no filter in place.

The work by Singh et al. (2014) and Yahyazadeh and Abadi (2015) noted and discussed the time taken for their detection technique to complete, however, Soniya and Wilscy's (2014) work did not. Likely because their detection method inherently took a longer amount of time due to the Large Time Slots Classifier which processed traffic for intervals varying from 4 hours to 32 hours. Furthermore, it would be unfair to compare the duration figures as the method by Singh et al. (2014) has a predetermined time constraint in that they attempted to detect botnets in quasi real time, whilst the methods by Yahyazadeh and Abadi (2015) and Soniya and Wilscy (2014) did not.

To conclude, the comparison has found that the method by Singh et al. (2014) shows the best detection rates of the methods evaluated, offering both the highest detection rate and lowest false positive rate. However, the method does suffer from a packet loss which is not documented by the other detection methods in the comparison. Due to the different time constraints of each detection method the paper concludes that it would have been unfair to attempt a comparison on this and thus has not.

4 Conclusions

This paper has analysed and evaluated three research papers aimed at improving botnet detection. The high detection rates seen in each research paper show that each approach to botnet detection has scientific merit.

To improve the detection method proposed by Singh et al. (2014) the paper recommends that future research focus' on eliminating the packet loss issue documented in the researchers' conclusions. Combining the filter module seen in the work by Soniya and Wilscy (2014) with the 'traffic sniffer' module documented by Singh et al. (2014) may prove an effective way to achieve this as Soniya and Wilscy (2014) did not suffer from any packet loss. If this combination was successful in eliminating the packet loss, real world testing would be the next logical step.

Overall the work by Soniya and Wilscy (2014) was the most thoroughly tested of the evaluated papers due to the range of botnets used during testing. This allowed the researchers to provide robust conclusions, the fact that the paper attained higher detection rates when compared to an earlier paper on botnet detection that utilised fuzzy logic (Wang et al., 2011) demonstrates its significance to the field of botnet detection. Future research in this area should focus on improving the detection rate and reducing the false positive rate of this method.

Although the conclusions reached by Yahyazadeh and Abadi (2015) were sound, future research should be conducted on testing the methods ability to detect a larger variety of Peer-to-Peer botnets. This is not to discredit the overall results provided by Yahyazadeh and Abadi (2015) as their focus was on detecting IRC, HTTP and P2P botnets within the same detection method.

All work discussed in this paper requires more research before real world deployment. Of the detection methods analysed, the work by Singh et al. (2014) is the most promising for future development. Not only because it offers both the highest detection rate and lowest false positive detection rate, as documented in table 7, but also due to the high level of detail found throughout the research paper, the use of standardised data during testing and the fact that can detect botnets in quasi real time.

References

Dietrich, C., Rossow, C. and Pohlmann, N., 2013., 'CoCoSpot: Clustering and recognizing botnet command and control channels using traffic analysis'. *Computer Networks*, 57(2), pp.475-486.

Elliott, C., 2010., 'Botnets: To what extent are they a threat to information security?'. *Information Security Technical Report*, 15(3), pp.79-103.

Kirubavathi, G. and Anitha, R. 2016., 'Botnet detection via mining of traffic flow characteristics'. *Computers & Electrical Engineering*, 50, pp.91-101.

Muthumanickam, K. and Ilavarasan, E. (2012). 'P2P Botnet detection: Combined host- and network-level analysis'. *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*.

Saad, S., Traore, I., Ghorbani, A., Sayed, B., Zhao, D., Wei Lu, Felix, J. and Hakimian, P., 2011. 'Detecting P2P botnets through network behavior analysis and machine learning'. *2011 Ninth Annual International Conference on Privacy, Security and Trust*. Pp 174-180

Sheng, L., Zhiming, L., Jin, H., Gaoming, D. and Wen, H. (2012). 'A Distributed Botnet Detecting Approach Based on Traffic Flow Analysis'. *2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control*.

Singh, K., Guntuku, S., Thakur, A. and Hota, C. 2014., 'Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests'. *Information Sciences*, 278, pp.488-497.

Soniya, B. and Wilscy, M. 2014., 'Fuzzy inference system based on entropy of traffic for bot detection on an endpoint host'. *2014 International Conference on Data Science & Engineering (ICDSE)*. pp 112-117

Wang, J. and Paschalidis, I. 2016., 'Botnet Detection based on Anomaly and Community Detection'. *IEEE Transactions on Control of Network Systems*, Issue 99.

Wang, K., Huang, C., Lin, S. and Lin, Y. 2011., 'A fuzzy pattern-based filtering algorithm for botnet detection'. *Computer Networks*, 55(15), pp.3275-3286.

Yahyazadeh, M. and Abadi, M. 2015., 'BotGrab: A negative reputation system for botnet detection'. *Computers & Electrical Engineering*, 41, pp.68-85.

Zhao, D., Traore, I., Sayed, B., Lu, W., Saad, S., Ghorbani, A. and Garant, D. 2013., 'Botnet detection based on traffic behavior analysis and flow intervals'. *Computers & Security*, 39, pp 2-16.