A Review and Proof of Concept for Phishing Scam Detection and Response using Apoptosis

A Yahaya Lawal Aliyu
Faculty of Science and Technology (FST),
Universiti Sains Islam Malaysia (USIM),
71800 Nilai, Negeri Sembilan, Malaysia.

Madihah Mohd Saudi
Faculty of Science and Technology (FST),
Universiti Sains Islam Malaysia (USIM),
71800 Nilai, Negeri Sembilan, Malaysia.

Ismail Abdullah
Faculty of Science and Technology (FST),
Universiti Sains Islam Malaysia (USIM),
71800 Nilai, Negeri Sembilan, Malaysia.

***Abstract*—Phishing scam is a well-known fraudulent activity in which victims are tricked to reveal their confidential information especially those related to financial information. There are various phishing schemes such as deceptive phishing, malware based phishing, DNS-based phishing and many more. Therefore in this paper, a systematic review analysis on existing works related with the phishing detection and response techniques together with apoptosis have been further investigated and evaluated. Furthermore, one case study to show the proof of concept how the phishing works is also discussed in this paper. This paper also discusses the challenges and the potential research for future work related with the integration of phishing detection model and response with apoptosis. This research paper also can be used as a reference and guidance for further study on phishing detection and response.**

Keywords*—Phishing; apoptosis; phishing detection; phishing response

I. Introduction

As online technology is growing at a faster level, so have other numerous online activities such as advertising, gaming, and e-commerce. As online financial activities are on the rise, so have online fraudulent activities in which phishing is playing a major role for illegally obtaining private individual details. Phishing activities against financial institutions have become a regular occurrence leading to a rising concern about how to increase security on these sectors which could relate to banks and online shopping such as Ebay and Amazon. Fraudulent schemes conducted via the Internet are generally difficult to trace and prosecute, and they cost individuals and businesses millions of dollars each year. From computer viruses to web site hacking and financial fraud, Internet crime became a larger concern than ever in the 1990s and early 2000s. In response to such issue, different anti phishing tools were developed in order to counter such illegal online activities [1].

As for the phishing activities, it has also been evolving on a rapid level in order to evade other anti-phishing tools that are been developed to counter the phishing tricks. Phishing emails are also known to contain links to the infected website. Phishing email directs the user to the infected website where they are asked to type in their personal information such as username and password of account details, so that the website will hack the information related to whatever the user enters. Phishing email is also sent to a large number of people and the phishers will also try to count the percentage of people who read that email and entered the information. It is very difficult to find that the individuals are actually visiting an actual site or malicious site. Phishing is also understood to be a sort of brand spoofing or carding. As a result researchers are attempting to reduce the risk and vulnerabilities of such fraudulent phishing activities [2]. Some researchers also define phishing as a new type of network attack. The attacker creates a replica of an existing Web page to fool users for example by using specially designed e-mails or instant messages into submitting personal, financial, or password data to what they think is their service provides' Website [3]. According to [4], phishing is a social engineering crime which is carried out by impersonating a trusted third party in order to attain access to private data or information. These are numerous definitions by different researchers depending on their point of view relating to their research. It could also depend on the trends the researchers are facing during their study due to the fact that the phishing techniques are always changing. It is important to understand why phishing has taken a lot of interest on targeting financial sectors. Numerous reports have shown that financial sectors are always under constant attacks through phishing techniques. According to Laidlaw and colleagues the share of phishing messages intended against the financial sector which consist of banks, payment systems and online stores have been rapidly on the rise for several quarters in a row [5]. In their porting period, it has raised to 50.96% of the total number of reported phishing attacks against various organizations, which is 4.73% higher than the value for the second quarter of 2016 as displayed in Fig. 1.

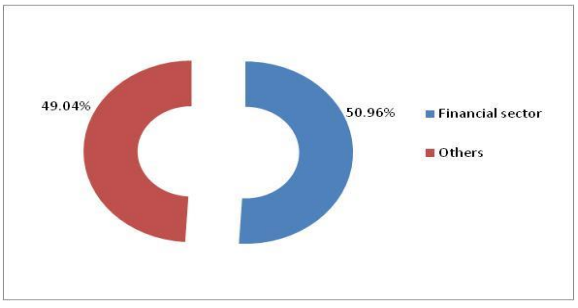

Fig. 1. Phishing target distribution of 3rd quarter (Q3) of 2016.

This research is funded by Ministry of Higher Education (MOHE), Malaysia and Universiti Sains Islam Malaysia (USIM).

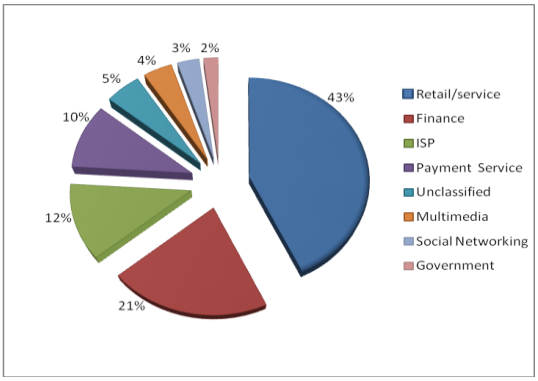

Fig. 2. Most recently targeted industries by phishing.

According to [6] there are many cyber-attacks targeted retail service sector, financial and payment service as displayed in Fig. 2. Financial gain is still known to be one of the major reasons behind most cybercriminal activities and there is no sign of this problem changing in the coming years or near future. Cybercriminals are continuously trying to make money. This is proved by the black market industry which has risen around different payments and card fraud. Cybercriminals have no problem coming up with different scams to make money [7].

Based on the phishing implication to financial sectors, therefore this paper aims to evaluate existing works related with the phishing detection and response techniques together with apoptosis. This paper also comes with a proof of concept (POC) on how the phishing attacks the victim. POC is important to help researcher to have a better understand on the phishing architecture. Hence the researcher will be able to grasp the idea how to detect and protect against phishing in future. This paper is organised as follows: Section 2 presents the related works with existing phishing detection methods and challenges. Section 3 explains the Apoptosis concept and benefits in applying it for protection against phishing. While Section 4 explains the POC of phishing attacks and Section 5 concludes and discusses future work for this paper.

II. Related Works

Different anti-phishing techniques have being on the rise in recent times due to the coming of advanced technological tools leading to an increase in phishing strategies invented by the perpetrators. Table 1 summarised the related works on anti-phishing tools.

In regards to the proposed research, despite the fact that there are existing works been implemented, applying the apoptosis concept against phishing activities will hopefully make improvements on the security aspect which will consist of not relying on black listing or white listing of website for threat identification. For the proposed research, implementing apoptosis should be able to identify any phishing threat through analysing of any slight change of message patterns within a network which could be either in the form of Domain Name Server (DNS) or a malware based phishing software from fraudsters. Another improvement to hoping to be made by using apoptosis concept is to improve protection against phishing by using its optimisation capability in which it will be able to measure its performance and policies to attempt to improve itself by reacting to any system changes by the user.

TABLE. I. Challenges Faced by Different Anti-Phishing Methods

Phishing Detection Tools	**Methods Used for detection**	**Challenges for improvement**
Proof Point [8]	• Offers a comprehensive solution for data protection and governance through an integrated, security-as-a-service platform.	Proofpoint solutions are complex and can include numerous modules that work together
CANTINA [9]	• Examine the content of a web page to find out whether it is legitimate or not. • Makes use of the well-known TF- IDF (term frequency/inverse document frequency) algorithm used in information retrieval.	Phishers are able to design their attacks to avoid CANTINA'S heuristic detection.
Auntie Tuna[10]	• It used with web browser plug-in that provides anti-phishing alerts whenever a user browses. • Indexes the target site's content and watches for this content to appear at incorrect sites which will identify a sign of active phishing.	Need to keep signature up to date against malware based phishing.
PhiGARO [11]	• It checks up on victims of phishing and prevents further harm related to the incident.	Depends on reports of phishing incidents from users.
Anomaly Based Phishing detection tool[12]	• Examine the anomalies in web pages, particularly, the difference between a web site's identity and its structural features and HTTP transactions.	False positives can become difficult using anomaly based setup.

Other limitations could be found in other phishing detection tools such as Anomaly based Phishing detection system which compares the fake website with the legitimate website by using Document Object Model (DOM) objects and Hyper Text Transfer Protocol (HTTP) transactions. The limitations noted are firstly, the network can be in an unprotected state as the system builds its profile, secondly if malicious activity looks like normal traffic to the system it will never send an alarm and false positives can become cumbersome with an anomaly based setup [12]. Although, some anti-phishing tools are useful, they also tend to have some limitations due to other circumstances. Also, as explained by [13], they described these flaws as follows:

- The attack could takes place at the necessary time for new (zero-day) phishing websites to be reported and hence added to the blacklist at that same time.
- The blacklist method may sometimes mislead to inform on a False Negative (FN) results showing that, the email or website is mistakenly identified as phishing.
- The white list method on the other hand is a collection of trustworthy URLs. This method however is a time consuming process. In addition, this method could cause an increased level of False Positive (FP) results, consequently letting phishing emails or websites to pass through; FP is meant to show that, the email or website is inaccurately identified as a legitimate.

Based on the limitations of all these anti-phishing tools that are mentioned, this research intends to come up with a new model on how to identify phishing activities based on using apoptosis algorithm which consist of a new phishing classification and also to optimise the accuracy of phishing detection rate by constructing new parameters. There are numerous phishing tools that are being developed due to the growing complication of phishing activities. One of these anti-phishing tool is the Logo Image Based Approach for Phishing Detection by [14]**.** This tool was built to capture screenshot and then commence with the approach in order to remove the logo. It focuses on to detect replaced images of the logo from downloaded image such as from image income of a query webpage by getting a screenshot to extract the logo. After capturing the screenshot, it will give the actual web content that is usually utilised for optimising the website loading speed.

Another phishing detection method is known as Feature Extraction or Feature Selection for Text Classification which is also a case study for phishing detection which was proposed by [15]. This tool deals with a lot of text classification which could be represented by thousands of token making the classification problem difficult. Therefore, dimensionality reduction is required to make the data representation much shorter and easier. This will make it possible to differentiate between emails. The techniques include feature extraction in which the original feature space is modified into a more compressed new space. Meanwhile, in feature selection technique, the original feature is chosen which will be used for the training and testing of the classifiers. The features that are discarded will not be used for the computations.

A research carried out on a phishing tool that was being performed by [16] is known as fMRI consisting of a study of Phishing and Malware Warning which measures the user's security performance together with the underlying neural activity with the task of distinguishing between a legitimate and phishing website. The phishing control experiment was built to take charge for stimuli offered for the phishing experiment whereby participants are instructed to observe the images shown on the screen without performing any active task. The neural activities were observed in which there were numerous indication of what was called top-down control and attention modulation system. The result of this phishing experiment showed significant activity in phishing activity during the study. It proved that such increased control might be critical for carrying out important judgment relating to the legitimacy of a website.

In regards to curbing out the problems of phishing, some authors are of the opinion that despite the benefits of online security tools, the people tend to be part of the problems which is also an issue to be solved in order to successfully reduce the online fraud effect. The researchers carried out a survey by using established questionnaires in order to evaluate information such as personality characteristics, impulsivity, web and computer based behaviour, previously experienced phishing consequences. In this survey, it was shown that people that are suspicious of others and also showing distrust on people have a low level of being affected by phishing attacks [17].

Dynamic analysis is to be utilised to carry out a test to observe how a phishing dataset will behave once sent by using email client server in the lab architecture. As for the proposed study, apoptosis method detection model aims to improve identification of any suspicious pattern related to phishing through the use of the dynamic analysis process in which any slight suspicious occurrence will be detected even if the fraudsters come up with new phishing procedure. The identity extractor will be further improved using the dynamic process. As for the page classifier, it may sometimes be bypassed due to being in an unprotected state while building its profile but improvement through implementing the apoptosis concept will be able to solve and rectify such limitation through its optimisation capability.

III. APOPTOSIS COMPUTING CONCEPT

For the apoptosis concept to be applied, it is important to note that it is divided into two parts. These parts are known as intrinsic and extrinsic apoptosis. Intrinsic apoptosis is known to be a reaction or response to damages that are internal such as damaged chromosomes or DNA. As for extrinsic apoptosis, it normally occurs as cellular immune response when invaded by external threat [18]. This research will use the extrinsic concept because that is the same process as being attacked by phishing from an external source.

As for the proposed research on phishing scam detection method using Apoptosis, it is meant to rectify these limitations by utilising its autonomous capability which consist of self-configuration, healing, optimisation and protection to provide full time protection against phishing related to financial frauds. Implementing the apoptosis concept is meant to overcome such weakness through its constant monitoring ability. The research proposes for a model is simplified. The process will consist of two phases. The PDA (Phishing Detection Apoptosis) consists of two phases. In regards to phase 1 in the PDA model, after the detection and Analysis, the PDA classification process will consist specifically of phishing site or email. After the analysis, the phishing mail or site will be classified followed by the data matching process which will perform the task of identifying and merging the records which corresponds with the same entities. The sample model can be seen in the Fig. 3.

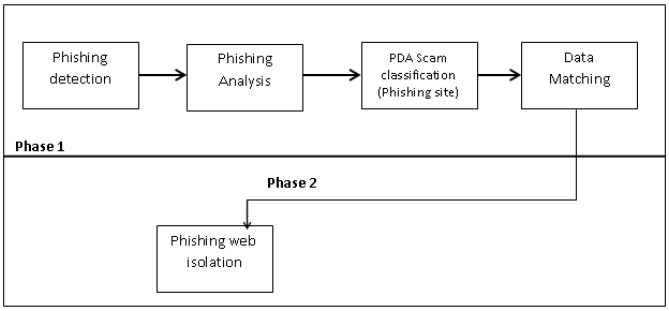

Fig. 3. Phishing Detection Apoptosis (PDA) Model.

Since multi-cellular computing is about virtual interaction rather than physical interaction, a computer only needs to be separated from communication and connection with any other computer. As a result, in computing issues, there should be more focus on quarantining an infected computer by cutting it off from the net. Rapid detection of infection is very important. Any infected computer, on average, could have the capability of infecting less than one other computer. Or else, the epidemic of infection may grow to other systems. Therefore the priority must be to detect infection quickly [19].

In describing Apoptosis implementation by [20], he explained it as a mechanism of a human immune system based on apoptosis that is adopted to build an Intrusion Detection System (IDS) to protect computer networks. Based on his work, features were selected from network traffic using Fisher Score. Also in relation to the selected features, the record/connection is classified as either an attack or normal traffic by the proposed methodology. Simulation results demonstrates that the proposed AIS based on apoptosis performs better than existing AIS for intrusion detection.

In issues that relate to the discipline of Natural Computing, the Apoptosis example can be observed within the context of Artificial Immune Systems [21]. Therefore the conception related to Apoptosis is utilised and embedded into computer system security in which a system is designed and developed from numerous types of small units and in this instance, if one becomes "damaged" through a computer malware, it will either initiate, or will be instructed to initiate Apoptosis concept without affecting other connected system components. This is similar to the analogy of being an animal cell that is invaded by a virus and the immune system knowing it is a foreign body and therefore will attack it. Inventing such type of artificial immune system for a self-managing autonomic computer system, such as a system with the ability to self-configuring, self-healing, self-optimising and self-protecting, is now been stated as a kind of a Holy Grail inspiring a different types of research papers. The autonomic computing paradigm is based on the biological metaphor of the Autonomic Nervous System in that it is self-managing without conscious input from the user, and is gaining ground as a way of designing and building systems capable of dealing with increasing cost and complexity [22].

The benefit of using this detection method is due to its interesting feature of being autonomic (acting involuntarily) without conscious control. It will be used on because Apoptosis detection tends to give full protection from any kind of threat.

The reason for such anticipated efficiency is also because it consists of self-managing cell which it uses for functional measurement with event correlation [23]. Some of its attributes are as follows:

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

Another interesting feature of Apoptosis as an autonomous system is the awareness of its components through its interconnectivity with other systems.

IV. FINDINGS

A case study using client server to perform a sample of phishing scam called phishing testis to shows the proof of concept (POC) on how the phishing works. Reverse engineering process and dynamic analysis were conducted to analyse the code using the architecture as illustrated in Fig. 4.

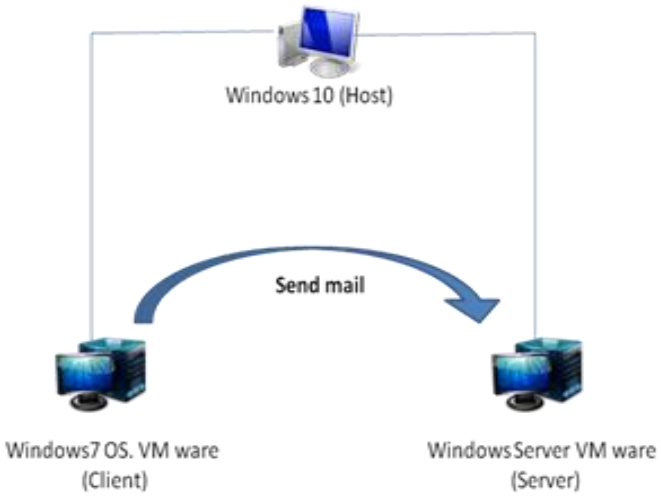

Fig. 4. Lab architecture.

TABLE. II. SOFTWARES FOR TESTING

Software	Function
MS Windows 10	To serve as a Host
MS Mail Server 2012	To receive the email
VM ware	To build virtual system in Host (Windows 10)
Windows 7	To function as Client
Process Monitor	To perform the analysis (Dynamic)
Process explorer	To perform the analysis (Dynamic)
Wireshark	To monitor the network traffic

The reverse engineering was performed in a controlled lab environment by using Windows 10 as the host for installing the testing tools. Windows 7 was used as the client and windows Server 2012 (Mail Server) was utilised as the server. Most of the software tools that were used for this experiment are freeware that can be downloaded freely from the internet. The tools are listed in Table II. The test was conducted within the virtual machine without the network connection. The phishing email dataset used was obtained from the malware traffic analyst website [24]. There are more than 2000 samples that can be freely downloaded for project uses. The dataset used for the test was a .pcap file.

After attaching the phishing mail sample to the virtual mail and sent to the server, numerous activities were observed using dynamic analysis tools. Both the client and server were used within a virtual network of the VM ware. For this experiment, outgoing network was not required. Further analysis showed that the phishing dataset attachment was able to successfully create a thread in which stolen information could be sent through. It also showed a buffer overflow in the result section which might cause loading of files to be slow with the following path at the registry.

HKCU\Software\Windows\CurrentVersion\Internet Settings\Connections\DefaultConnection.

For issue related to computer security, a buffer overflow is a strange activity where a program, during writing of data to a buffe's boiundry and overwrites adjacent memory locations.

Many fraudsters exploit buffer overfloe to gain access to a victims computer in order to gain access to information. The result is shown in Fig. 5. The file activity also made the outlook to do a "Thread Create" in the operation section which opens a registry key. This thread is known to be a holding place of information known to be related to single use program which may be handled by multiple users. Phishers may exploit this operation to retrieve information when using malware based phishing.

Based on the test conducted, it shows that phishing techniques can cause minimal distortions to make a system vulnerable to information theft. Therefore, implementing an efficient phishing detection model is to be performed in order to increase protection against phishing. As for future work, based on 4 concepts of the apoptosis, the concepts are listed and explained as follows:

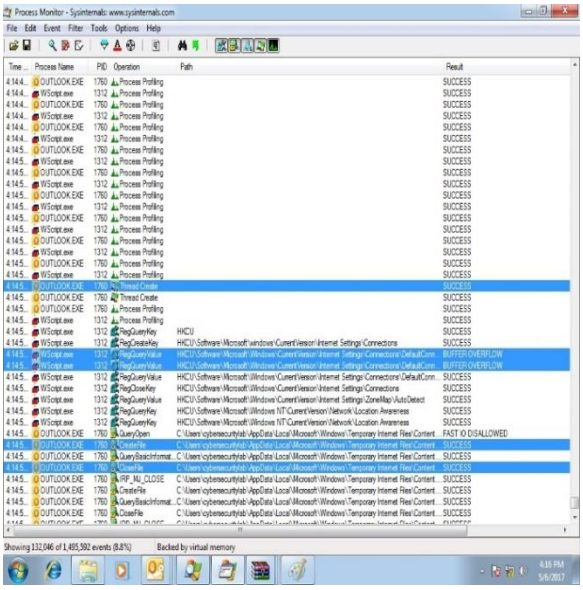

Fig. 5. Screen shot result after sending phishing email.

- Self- Protection: - In regards to this Apoptotic concept which is also known as Autonomic Defence System (ADS), it consist of very useful attributes comprising of protection of information system, actuators for implementing appropriate response, sensors for detecting attacks and lastly a controller for coordinating the sensors with the actuators.
- Self- Healing: - This concept enables the system to be able to automatically recover from any fault. Since fault detection is very important in order to develop an effective self- Healing system, the aim of this part of the system is to be able to do self-repair of components that might have failed without the need to bring down the whole system in order to ensure resource availability is maintained.
- Self- Configuration: - This concept enables the system to be able to adapt to any changes within its software environment or physical topology. This concept will also improve system reliability through rectification of human configuration errors which will in turn, reduce time wasting, therefore, making system resource available [25].
- Self- Optimisation: - For this concept, the system will be able to manage the systems complexity by responding to dynamic changes in order to improve the systems performance. The key aspects in optimisation are its resource utilisation and workload management. The main aims of these activities are to enable maximisation of system's operation.

These have been mapped to phishing detection and response technique as summarised in Table III.

TABLE. III. APOPTOSIS MAPPING TO PHISHING TECHNIQUE

Apoptosis Concept	Map to Phishing Detection and Response Technique
Self-protection	Defend against fake Domains sent by fraudsters
Self healing	Being able to recover from any malware based phishing file attack to the system
Self configuration	Automatically able to readjust itself to support the system (Improvement against any new phishing pattern)
Self-optimisation	Automatically adapt to any change, mostly against system complexities.

V. CONCLUSION

Phishing activities are evolving on a very fast level due to numerous innovations in online technology in recent times. Most malware programmers and online fraudsters are now inventing ways of bypassing many online security tools. For these reasons, it is highly required that online security is designed especially in the dynamic field sector. By utilising the use of Artificial Immune Systems such as Apoptotic computing, there is a high expectation of acquiring efficient result because of its method in detecting any malicious activity on a computer device or network. By thorough research, Apoptotic security concept will be implemented in order to hopefully increase the efficiency of protecting different fraudulent (Phishing) activities by identifying irregularities in both networks and system behavior within the online financial network.

This study will also aid other researchers to assist with coming up with new innovated ideas on how Artificial Immune Systems can be further developed in order to attain a more effective protection against online fraudulent activities. The model to be developed in this study will hopefully aid in coming up with the idea on reducing the effect of phishing on financial sectors. By utilising dynamic analysis for this study, a higher advantage will be attained against other anti-phishing tools especially the ones that rely on static analysis. Based on the findings on the proof of concept and other tests to be carried out for future analysis, the apoptosis model will be applied in order to observe the level of efficiency on how it will be able to detect any phishing pattern on a computer network with the aim of attaining a proper validation outcome.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Ministry of Higher Education (MOHE), Malaysia and Universiti Sains Islam Malaysia (USIM) for their support and facilities provided. This research paper is supported by grants: [USIM/FRGS/FST/32/50114] and [PPP/USG-0216/FST/30/16916].

REFERENCES

[1] Candid Wueest, "The state of financial Trojans 2014," Symantec Corporation, pp. 1-24, 2015.

[2] V. Suganya, "A Review on Phishing Attacks and Various Anti Phishing Techniques," Int. J. Comput. Appl., vol. 139, no. 1, pp. 975–8887, 2016.

[3] U.Naresh, "Intelligent Phishing Website Detection and Prevention System by Using Link Guard Algorithm," IOSR J. Comput. Eng., vol. 14, no. 3, pp. 28–36, 2013.

[4] Whittaker, C., Ryner, B., and Nazif, M. "Large-Scale Automatic Classification of Phishing Pages". In NDSS , Vol. 10, pp. 1, 2010.

[5] S .Laidlaw and M .Hillick, "Profiling cyber threats detected in a target environment and automatically generating one or more rule bases for an expert system usable to profile cyber threats detected in a target environment". U.S. Patent 9,503,472. Cyberlytic Limited, 2016.

[6] Anti-Phishing Working Group, "Phishing Activity Trends Report 1 Quarter," Most, no. March, pp. 1–12, 2010.

[7] R. Mohamad, A. Building, and N. A. Ismail, "Journal of Internet Banking and Commerce," J. Internet Bank. Commer., vol. 15, no. 1, pp. 1–11, 2010.

[8] H. Proofpoint, C. Help, and A. Proofpoint, "Why Today ' s Phishing Attacks are Harder to Detect and How Proofpoint Can Help Why Today ' s Phishing Attacks," no. 2, pp. 1–16.

[9] Y. Zhang, J. Hong, and L. Cranor, "Cantina: a content-based approach to detecting phishing web sites," Conf. World Wide Web, pp. 639–648, 2007.

[10] C. Ardi and J. Heidemann, "AuntieTuna : Personalized Content-based Phishing Detection," no. February, 2016.

[11] M. Husak and J. Cegan, "PhiGARo: Automatic phishing detection and incident response framework," Proc. - 9th Int. Conf. Availability, Reliab. Secur. ARES 2014, pp. 295–302, 2014.

[12] P. Ying and D. Xuhua, "Anomaly based web phishing page detection," Proc. - Annu. Comput. Secur. Appl. Conf. ACSAC, pp. 381–390, 2006.

[13] M. M. Al-daeef, N. Basir, and M. M. Saudi, "An anti-phishing tool to verify urls in email content," vol. 10, no. 3, pp. 1378–1382, 2015.

[14] H. Thakur and S. Kaur, "Logo Image Based Approach for Phishing Detection," vol. 6913, no. December 2016, pp. 129–139.

[15] M. Zareapoor and S. K. R, "Feature Extraction or Feature Selection for Text Classification: A Case Study on Phishing Email Detection," Int. J. Inf. Eng. Electron. Bus., vol. 7, no. 2, pp. 60–65, 2015.

[16] A. Neupane, N. Saxena, K. Kuruvilla, M. Georgescu, and R. Kana, "Neural Signatures of User-Centered Security: An fMRI Study of Phishing, and Malware Warnings," Proc. 2014 Netw. Distrib. Syst. Secur. Symp., no. February, pp. 1–16, 2014.

[17] M. Jakobsson, "The Human Factor in Phishing," Priv. Secur. Consum. Inf. 07, http//www.informatics.indiana.edu/markus/papers/aci.pdf, vol. 7, pp. 1–19, 2007.

[18] Z. Hongmei, "Extrinsic and Intrinsic Apoptosis Signal PathwayReview," Apoptosis Med., pp. 3–22, 2012.

[19] R. Sridevi and G. Jagajothi, "Apoptosis Inspired Intrusion Detection System," vol. 8, no. 10, pp. 1890–1896, 2014.

[20] D. Jones, "Implementing biologically-inspired Apoptotic behaviour in digital objects : An Aspect-Oriented Approach," no. March, 2010.

[21] M .Saudi, M.Woodward, J.Cullen and M .Noor, "An overview of apoptosis for computer security ". In Proc. International Symposium on Information Technology ITSim2008., 2008.

[22] R. Sterritt, "Apoptotic computing: Programmed death by default for computer-based systems," Computer (Long. Beach. Calif)., vol. 44, no. 1, pp. 59–65, 2011.

[23] T. Spotlight, C. Technologies, H. Education, S. Technologies, and S. Technologies, "Autonomic Computing When You Should Expect It," pp. 2001–2002, 2015.

[24] A source for Pcap abd Malware Samples. "Malware Traffic Analysis ". N.p., 2017. Web. 6 May 2017

[25] Dai. Y, Marshall. T & Guan. X, "Autonomic and Dependable Computing: Moving Towards a Model-Driven Approach," Journal of Computer Science., vol. 2, pp. 497-500, 2005.