Hybrid Model of Phishing Email Detection: A Combination of Technical and Non-Technical Anti-Phishing Approaches

Melad Mohamed Al-Daeef, Nurlida Basir, and Madihah Mohd Saudi

Abstract—**Phishing is a cybercrime in which, attackers try to fraudulently retrieve users' credentials by mimicking trusted communication channels. The problem with phishing is that attackers still able to bypass anti-phishing automated systems through the human factor. It is not enough, therefore, to only add new technologies, aware users might play the key role in stopping phishing attacks. Based on that, phishing problem requires defense solutions that to be applied at both of the technical (automated systems) and non-technical (human) aspects. Phishing attacks, in general, are initiated through simulated emails with a false claim of being sent from trusted parties. The work in this paper is dedicated to fighting phishing threats at email's level in order to kill this type of attacks in the cradle. Users, therefore, are protected at a level which is prior of browsing phishing web pages. This paper proposes an anti-phishing model that designed based on the general taxonomy of the technical and non-technical aspects of phishing detection approaches. This paper, in addition, presents the general structure of the proposed anti-phishing system that developed based on the herein proposed model. The novelty of this model is the approach of combining both of the automated procedures with users' anti-phishing training method to detect phishing emails.**

Index Terms— **immunity approach, phishing email, technical solutions, user awareness, URL-based classification feature**

I. INTRODUCTION

Phishing is a complicated problem that requires defense solutions to be applied at both of the technical (automated) and non-technical (human) aspects. Researchers have implemented the automated-based solutions such as client-side toolbars, classifiers, authentication mechanisms, artificial immune systems, etc. It, however, widely claimed that automated solution alone cannot be relied upon to stop phishing attacks since phishers can change their attacking techniques to bypass such automated systems through users' unawareness, inattention, and ignorance factors [1]. Users' mistakes cannot be avoided by only adding new technologies; anti-phishing training approach has, therefore, been widely utilized to mitigate the bad impact of phishing attacks. Security awareness training is a promising choice to alleviate the limitations of the technical aspects of phishing solutions [2], [3].

Phishing attacks are usually launched through simulated emails that falsely claim sent from trusted parties such as organizations or banks that the victims deal with. It is a useful countermeasure, therefore, to fight phishing attacks at the email level and kill phishing attacks in the cradle. It is a common scenario when phishing emails contain fake URLs to deliver the victims to phishing websites [4]. Anti-phishing systems that solely operated using URL-based features can perform better than the systems that operated based on the content-based and text-based classification features. That is due that the content of phishing emails is usually constructed to look like legitimate ones. It is difficult, therefore, for the classification systems to correctly classify emails using only content-based classification features [5], [6]. URL-based classification systems can also eliminate many of operating costs and security risks associated with anti-phishing systems [7].

Besides of the automated anti-phishing systems, Internet users need to be trained on how to protect themselves and how to react against fraudulent activities [8]. Many of challenges, however, might limit the desired benefits from applying anti-phishing training approach. The most obvious challenge might be is how to make training process as an ongoing activity that makes the trainees retain the acquired knowledge for a longer time. Another challenge is how to help them to transfer this acquired knowledge to other related security contexts [9], [10]. The technical and non-technical aspects of anti-phishing solutions should complement each other since phishing threats cannot be eliminated by only adding more technologies. This paper introduces a novel anti-phishing model that combines both of the technical and non-technical approaches in one anti-phishing solution.

The rest of this paper is organized as follows; section II presents the related work to this paper's topic, section III introduces the proposed anti-phishing model that aims to minimize the probabilities of the phishers' success to bypass anti-phishing systems through users' unawareness factor, evaluation process of the proposed system which was developed based on this model is briefly presented in section IV, this paper has lastly concluded in section V.

Manuscript received July 18, 2017; revised July 26, 2017.

Melad Mohamed Al Daeef is a PhD Student at the Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai 71800, Negeri Sembilan Darul Khusus, Malaysia Hand phone: 0060-18250-3435; e-mail: meladmohalda@gmail.com

Dr. Nurlida Basir is a lecturer at the Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai 71800, Negeri Sembilan Darul Khusus, Malaysia e-mail: nurlida@usim.edu.my

Assoc. prof. Dr. Madihah Mohd Saudi is a lecturer at the Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai 71800, Negeri Sembilan Darul Khusus, Malaysia e-mail: madihah@usim.edu.my

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

II. Related work

Many approaches have been utilized to stop phishing attacks. Automated systems are usually developed using the common anti-phishing methods that include blacklists, whitelists, and heuristics [11]. Blacklists are updated databases of previously known phishing URLs, IP addresses, or keywords. These lists are usually maintained by online communities such as Anti-Phishing Working Group (APWG) [12]. Although of their accurate detection results, the blacklists, however, cannot identify the fresh, zero day, phishing instances due to the update time lag of the lists' content [11], [13]. Whitelists on the other side are less common in implementation than blacklists, they usually contain the trusted URLs those the Internet users wish to visit. The list entry error, however, limits the wider implementation of this method, users, in addition, might be annoyed when they are frequently promoted to update their whitelists' contents. The users, therefore, might choose the auto update pattern, or they might simply disable this function [14]. The heuristics method is used on the other side to check emails' or websites' characteristics that include, URLs, HTML code, or page content to determine whether they pose a threat or not [15]. The heuristics based systems are more efficient than blacklist or whitelist based systems in detecting fresh phishing instances [11], phishers, however, are sophisticated enough to bypass the heuristic-based methods of detection [16]. Microsoft IE phishing filter [17], Mozilla Firefox's [18], SpoofGuard [19], and PhishCatch [20] are examples of the client-side toolbars that built up using blacklists, whitelists, and/or heuristic methods. The accuracy of detection results of anti-phishing tools is error prone and can be affected by the technology changes and also to the changes in phishing patterns.

The non-technical aspect of phishing solutions on the other side focuses on the human factor. In order to increase users' awareness about phishing phenomenon, anti-phishing training is a widely applied approach in this aspect of solutions. During the training practice phase, permanent effects may be confounded with the temporary performance effects that quickly disappear after the finish of practice session or the change in test conditions [9]. Awareness training program success relies significantly on the method by which training materials are delivered to the trainees [21]. Training materials can be delivered through many channels such as emails, posters, classroom training, web seminars, games, etc. Each of these methods, however, has its associated limitations. The classroom sittings method, for example, is insufficient when it comes to training large numbers of people due to the high cost and consumed time [22], trainees need to touch, feel, and experience the content. Posted articles can help the trainees only if they have actually read them. The users might wrongly believe that they are aware enough and they know how to protect themselves [23]. The users, therefore, do not read such posted articles and such training information becomes ineffective, particularly against the new sophisticated phishing approaches [24]. It is believed that the game-based training delivery method offers an effective alternative to the traditional training methods [25]. The game method, however, lacks the knowledge transfer characteristic, and impose players to gain required security knowledge before they start the game [26]. An effective training experiment should help the trainees to acquire a new knowledge, retain this knowledge for a long lasting time, and transfer it into other related contexts [9]. Results in [24] show that the participants in the embedded training condition were better in making decisions than participants in the non-embedded training condition.

In this paper, the proposed anti-phishing model is designed to complement the automated phishing email detection process with the daily email browsing activity. That is to make end users directly involved in email's classification process, they, thus, involved in ongoing anti-phishing training experiment.

III. Proposed Anti-Phishing Model

Using the proposed model, a given email is classified as either legitimate or phishing one after many steps of checking. Firstly; all URLs in the checked email's content are extracted, the Full_Domain_Name part of this email's Message-ID field is extracted as well. Secondly; running of the immunity based, blacklist based, and heuristic based checking modules in order to detect phishing indications. Thirdly; determine the suspicion level (SL) of the emails that identified as phishing, the proposed suspicion level module is implemented for that process. Fourthly; the user is informed about the initial automated-base decision. Fifthly; enable the users to request further WHOIS-based information about the extracted URLs before the users make the final decision. Lastly, update the immunological memory cells (IMCs) to easily detect subsequent attacks from the previously known or detected phishing sources. User's knowledge at this phase is also updated and their awareness is definitely improved.

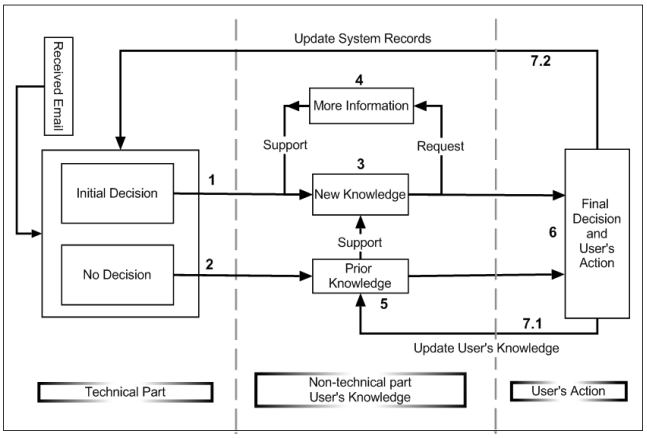

Fig. 1. The Proposed anti-phishing Model

The SL of any identified phishing email is determined by using a novel Reliability Ratio (RR) measure which was introduced in [27]. RR measure was used to evaluate the efficiency of employed email classification features. The process of implementing the RR measure makes users more suspicious and thus more aware than just they blindly classify checked emails as phishing ones.

Fig. 1 shows the structure of the proposed model which was designed based on the general taxonomy of anti-

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

phishing solutions and the necessity to combine both of the technical and non-technical approaches in one solution. The operation steps of the proposed model are explained as follows;

As a new email is received, its content is parsed to extract all URLs in its content. The Full_Domain_Name part of this email's Message-ID field is also extracted to be compared with the IMCs' content to know whether this email has been sent from a previously known phishing source or not. Following steps are then taken to classify this email;

Step 1, an initial decision is to be taken by the system either based on the result of applying the URL-based classification features or based on the result of comparing the Full_Domain_Name information with IMCs' content. The immunity-based module (AIS) is used to detect recurrent attacks from previously known phishers even when they change their attacking patterns to bypass anti-phishing systems, AIS module in such a case can remember the phishing attack source. The initial automated-based decision is then passed to the user at step 3. The user might acquire more supporting knowledge about why this email has identified as either legitimate or phishing one. The user might get more supportive knowledge by either requesting more WHOIS based information, step 4, or by relying on its prior knowledge, step 5. Based on the initial decision and also based on the users' knowledge, the user then can make a true and accurate final decision, step 6. This finally made decision should be utilized to update both of users' prior knowledge, step, 7.1, and also IMCs' content, step, 7.2. If the received email has no URLs in its content, or the Full_Domain_Name part of its Message-ID field has not remembered by the IMCs, the system, in this case, will not initially identify this email as a suspicious one, step 2, the user in such a case should make the final decision based on its prior knowledge.

IV. Evaluation

To evaluate the performance of the proposed model and the system which was developed based on this model. Fig. 2 shows the structure of the developed system in which, the technical and non-technical aspects are combined in one solution to complement each other. Separated evaluation processes have been conducted to individually evaluate each of the modules from which the system is consisting of.

A. Evaluation of the Immune-based module (AIS)

Classification results that obtained from implementing the AIS module have been compared with the classification results that produced by the heuristic-based module on the same dataset of phishing emails. The AIS module was evaluated based on its ability to memorize the phishing sources from the previously detected phishing exposures. Obtained results have shown that many of phishing emails have not detected by the heuristic-based module, same phishing emails, however, were successfully trapped at the AIS checking point. Experimental results show that the AIS was able to trap up to (%93) of phishing emails based on their Full_Domain_Name information. The details of AIS's evaluation experiment can be found in [28].

B. Evaluation of Heuristic and Blacklist Based Modules

The heuristic-based module examines all URLs that extracted from email's content. A given URL is identified as a suspicious if one of the applied URL-based classification features has positively met. The 12 URL-based classification features that involved in the evaluation process can be found in [27]. The heuristic-based module was evaluated based on the number of detected phishing emails out of the number of all emails in the involved phishing email dataset. The blacklist-based module was evaluated by comparing the Domain Name (DMN) of the examined URL against a blacklist of phishing DMNs that maintained by PhishTank [29]. If the DMN of a given URL did not match any of the blacklisted DMNs, this URL, therefore, is considered as a legitimate one. This result, however, is not a final, it should be passed to the user, he/she thus can make the final decision based on its prior knowledge or by initiating a WHOIS query for more information.

C. Evaluation of Suspicion Level Module

The suspicion level module was developed to determine the SL of identified phishing emails based on the SL of implemented classification feature(s) that upon these emails were identified as phishing ones. The reliability level of the employed classification features was determined by the RR measure [27] where the results show that the RR measure has outperformed Information Gain measure (IG) which has improperly been used to determine the efficiency of email classification features.

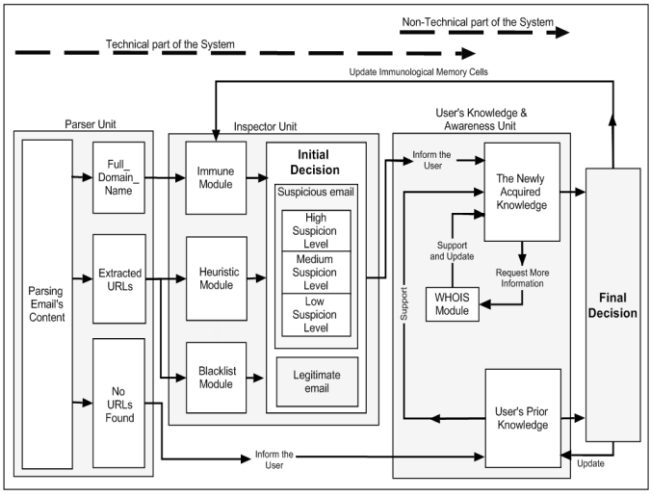

Fig. 2. The Structure of the Proposed Anti-Phishing System

D. Evaluation of WHOIS Module

WHOIS module is implemented in this work to support user's knowledge about URLs in email's content. If there was no useful information returned from the WHOIS query, the questioned URL is therefore considered a suspicious one and the user will be more cautious. This gives the user more confidence to finally make the correct decision. Evaluation criterion of WHOIS module is determined based on user's knowledge and how does he/she utilize it. The experimental evaluation process of WHOIS module is detailed in [30]. It was evaluated based on its ability to retrieve some related information about the DMNs of some instances of phishing and legitimate URLs that extracted from suspicious emails.

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

V. Conclusion and Future Directions

The Internet has extremely impacted the peoples' life patterns. It, however, has opened new avenues for fraudulent activities, phishing is one of such activities in which, the phishers play on users' unawareness and inattention factors to bypass the automated anti-phishing systems. This paper introduces a novel model in which, both of the technical and non-technical aspects of phishing approaches are combined in one solution. Evaluation experimental results have shown the effectiveness of the proposed phishing email detection model and the system which was developed upon it. The AIS module was able to detect new phishing emails although they have passed the heuristic-based checking point. The novel RR measure [27] which was implemented by the proposed system has shown its efficiency in determining the SL of employed phishing classification features. Results show that the RR measure has a privilege over the IG measure which has improperly been used to evaluate the phishing classification features. In our previous work [27], both of the RR and IG measures were applied to evaluate the same feature set over the same legitimate and phishing email datasets. The proposed model has opened a new direction to develop anti-phishing systems that embed the process of phishing email detection in the normal email browsing activity. This approach can continually improve users' awareness about phishing phenomenon.

As a future direction, the usability of the proposed system might be improved by designing a user-friendly interface that eases the systems' usage and performance monitoring process. A well-designed interface will definitely help the users to quickly and conveniently acquire phishing related knowledge. As a consequence, users will be better protected.

Acknowledgment

This work has been supported by Intelligent Spam Detection Model Based on Immunological Memory of Adaptive Immune System (USIM/FRGS/FST/50116-50).

References

[1] Waly, N., R. Tassabehji, and M. Kamala. *Improving organisational information security management: The impact of training and awareness*. in *High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICESS), 2012 IEEE 14th International Conference on*. 2012: IEEE.

[2] Hight, S.D., *The importance of a security, education, training and awareness program, November 2005*.

[3] Luo, X. and Q. Liao, *Awareness education as the key to Ransomware prevention.* Information Systems Security, 2007. **16**(4): p. 195-202.

[4] Yu, W.D., S. Nargundkar, and N. Tiruthani. *A phishing vulnerability analysis of web based systems*. in *Computers and Communications. ISCC 2008. IEEE Symposium on*. 2008: IEEE.

[5] Fette, I., N. Sadeh, and A. Tomasic. *Learning to detect phishing emails*. in *Proceedings of the 16th international conference on World Wide Web*. 2007: ACM.

[6] Irani, D., et al. *Evolutionary study of phishing*. in *eCrime Researchers Summit*. 2008: IEEE.

[7] Blum, A., et al. *Lexical feature based phishing URL detection using online learning*. in *Proceedings of the 3rd ACM workshop on Artificial intelligence and security*. 2010: ACM.

[8] Butler, R., *Investigation of phishing to develop guidelines to protect the Internet consumer's identity against attacks by phishers.* South African Journal of Information Management, 2005. **7**(3).

[9] Schmidt, R.A. and R.A. Bjork, *New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training.* Psychological science, 1992. **3**(4): p. 207-217.

[10] Eminağaoğlu, M., E. Uçar, and Ş. Eren, *The positive outcomes of information security awareness training in companies–A case study.* information security technical report, 2009. **14**(4): p. 223-229.

[11] Abbasi, A. and H. Chen, *A Comparison of Tools for Detecting Fake Websites.* IEEE Computer, 2009. **42**(10): p. 78-86.

[12] APWG, *http://www.apwg.org/*. p. http://www.apwg.org/

[13] Zeydan, H.Z. and M.S. Selamat, *Current State Of Anti-Phishing Approaches And Revealing Competencies.* Journal of Theoretical and Applied Information Technology, 2014. **70**(3).

[14] Dunlop, M., S. Groat, and D. Shelly. *GoldPhish: using images for content-based phishing analysis*. in *Internet Monitoring and Protection (ICIMP), 2010 Fifth International Conference on*. 2010: IEEE.

[15] Thrikkakara, C., *Distributed Software agents for antiphishing.* 2013.

[16] Hamid, I.R.A. and J. Abawajy. *Phishing email feature selection approach*. in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*. 2011: IEEE.

[17] IE-Phishing-Filter. *IE. http://support.microsoft.com/kb/930168*.

[18] FirePhish, *https://addons.mozilla.org/en-US/firefox/addon/firephish-anti-phishing-extens/*.

[19] Chou, N., et al. *Client-Side Defense Against Web-Based Identity Theft*. in *NDSS*. 2004.

[20] Yu, W.D., S. Nargundkar, and N. Tiruthani. *Phishcatch, a phishing detection tool*. in *Computer Software and Applications Conference. COMPSAC'09. 33rd Annual IEEE International*. 2009.

[21] Abawajy, J. and T.-h. Kim, *Performance analysis of cyber security awareness delivery methods*, in *Security technology, disaster recovery and business continuity*. 2010, Springer. p. 142-148.

[22] Kumaraguru, P., et al. *Lessons from a real world evaluation of anti-phishing training*. in *eCrime Researchers Summit, 2008*. 2008: IEEE.

[23] Kumaraguru, P., et al., *Teaching Johnny not to fall for phish.* ACM Transactions on Internet Technology (TOIT), 2010. **10**(2): p. 7.

[24] Kumaraguru, P., et al. *Getting users to pay attention to anti-phishing education: evaluation of retention and transfer*. in *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*. 2007: ACM.

[25] Cone, B.D., et al., *Cyber Security Training and Awareness Through Game Play*. 2006: Springer.

[26] Khan, B., et al., *Effectiveness of information security awareness methods based on psychological theories.* African Journal of Business Management, 2011. **5**(26): p. 10862-10868.

[27] Al-Daeef, M.M., N. Basir, and M.M. Saudi. *Evaluation of Phishing Email Classification Features: Reliability Ratio Measure*. in *Proceedings of the World Congress on Engineering*. 2017.

[28] Al-Daeef, M.M., N. Basir, and M.M. Saudi, *Anti-Phishing Immune System.* Advanced Science Letters, 2017. **23**(5): p. 4745-4749.

[29] PhishTank, *" PhishTank home " ; http://www.phishtank.com/*.

[30] Melad Mohamed Al-Daeef, N.B., Madihah Mohd Saudi, *An Anti-Phishing Tool to Verify URLs in Email's Content.* ARPN Journal of Engineering and Applied Sciences, 2015. **10**(3): p. 1378-1382.