Types of anti-phishing solutions for phishing attack

IOP Conf. Series: Materials Science and Engineering **769** (2020) 012072 doi:10.1088/1757-899X/769/1/012072

Types of anti-phishing solutions for phishing attack

Siti Hawa Apandi, Jamaludin Sallim and Roslina Mohd Sidek

Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang, Gambang, Kuantan, Pahang, Malaysia

E-mail: sitihawa.apandi@gmail.com

Abstract. Nowadays, many people use Internet to do online activity. This scenario exposed them to danger in Internet which is phishing attack. In order to solve phishing attack, the anti-phishing solutions are needed. Based on our review, there are still lacks of articles that review on the types of anti-phishing solutions in detail. In this paper, a general idea of phishing attack and anti-phishing solutions is presented. The phishing attack can be classified into two categories which are social engineering and malware-based phishing attack. The anti-phishing solutions can be differentiating into two types which are phishing prevention and phishing detection. Compared to phishing prevention, the phishing detection is more important to solve the phishing attack. In phishing detection, there are two categories which are user awareness and software detection. The software detection has two methods which are traditional and automatic. There are two types for automatic method of software detection which are public phishing detection toolbars and academic phishing detection / classification schemes. Based on the comparison of all types of phishing detection, the academic phishing detection / classification schemes are more useful for phishing detection. For future work, we can do more research on the academic phishing detection / classification schemes that utilize deep learning to see its potential of accuracy to detect phishing websites.

1. Introduction

The people lifestyle has changed with the arrival of the Internet. Nowadays, the people spent a lot of time on the Internet. The people surf Internet by using smartphone, laptop or computer. The common things that people do with the Internet are online shopping, emails, surfing social media, chatting and many more. The people also use Internet to access information in the World Wide Web, or simply called as Web.

Internet users are exposed to cybercrime. The cybercrime is also referred as computer crime. It is a crime that involves computer and network. One of the cybercrime is phishing. The phishing is an act to trick Internet users to visit the phishing website for the purpose to obtain the personal information of Internet users such as username and passwords to login into the website. The attacker of phishing is known as phisher. The phisher exploit the retrieved information of the Internet users for their own gain such as financial gain, sell the stolen identities of the Internet users, fame and notoriety [1].

The most targeted for the phishing website are financial and online payment companies. The reason for why the Internet user fall the phishing attack is that the phishing website look similar with the legitimate website [2]. There is increasing number of phishing attack even though there are many research have been done to solve the phishing attack [3]. The anti-phishing solutions for the phishing attack can be divided into two types which are phishing detection and prevention.

The remainder of this paper is organized as follows. In Section 2, the phishing attack is explained. This is followed by the discussion of anti-phishing solutions in Section 3. Lastly is the conclusion in Section 4.

2. Phishing attack

One of the popular cybercrime is phishing attack. It was first discovered in 1996 [3]. The term phishing sound similar with fishing. This is because the idea of phishing attack is same as the idea of fishing where someone will throw out a bait for the user to catch the bait [2, 4, 5]. The phisher lures the Internet user to visit the phishing website in order to steal the personal information of the user. Figure 1 shows the life cycle of phishing attack.

There are three components in the phishing attack which are medium, vector and technical approaches. For the medium of phishing, there are three medium which are internet, voice and short messaging service (SMS). The internet is commonly used medium for phishing because it has opened a huge chance for the phisher to deploy the phishing attack. The vector means a place to launch the phishing attack. The vector for internet are email, website and social networks [4]. The technical approaches for the phishing attack can be differentiate into two category which are social engineering and malware-based phishing attack [6]. The social engineering exploit the user emotion of fear to lose something valuable, thus the user reveal the personal information to the phisher [4]. For the malware-based phishing attack, it secretly install malicious programs in order to give the phisher access to the user computer [6].

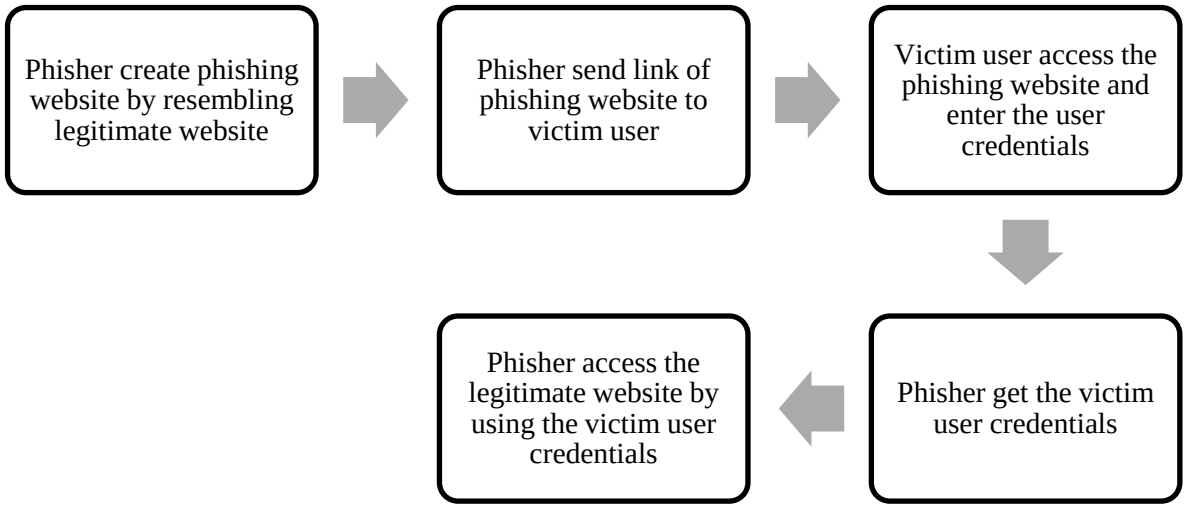

Figure 1. Life cycle of phishing attack [2].

3. Anti-phishing solutions

There are many tools exists to solve phishing attack. The phishing attack solutions also known as anti-phishing solutions. Figure 2 shows the types of anti-phishing solutions. There are two types for anti-phishing solutions which are phishing prevention and phishing detection. For phishing detection, it can be differentiate into two categories which are user awareness and software detection. There is traditional and automatic method of the software detection. For automatic of the software detection, there are two categories which are public phishing detection toolbars and academic phishing detection / classification schemes. Each of the anti-phishing solutions is briefly described.

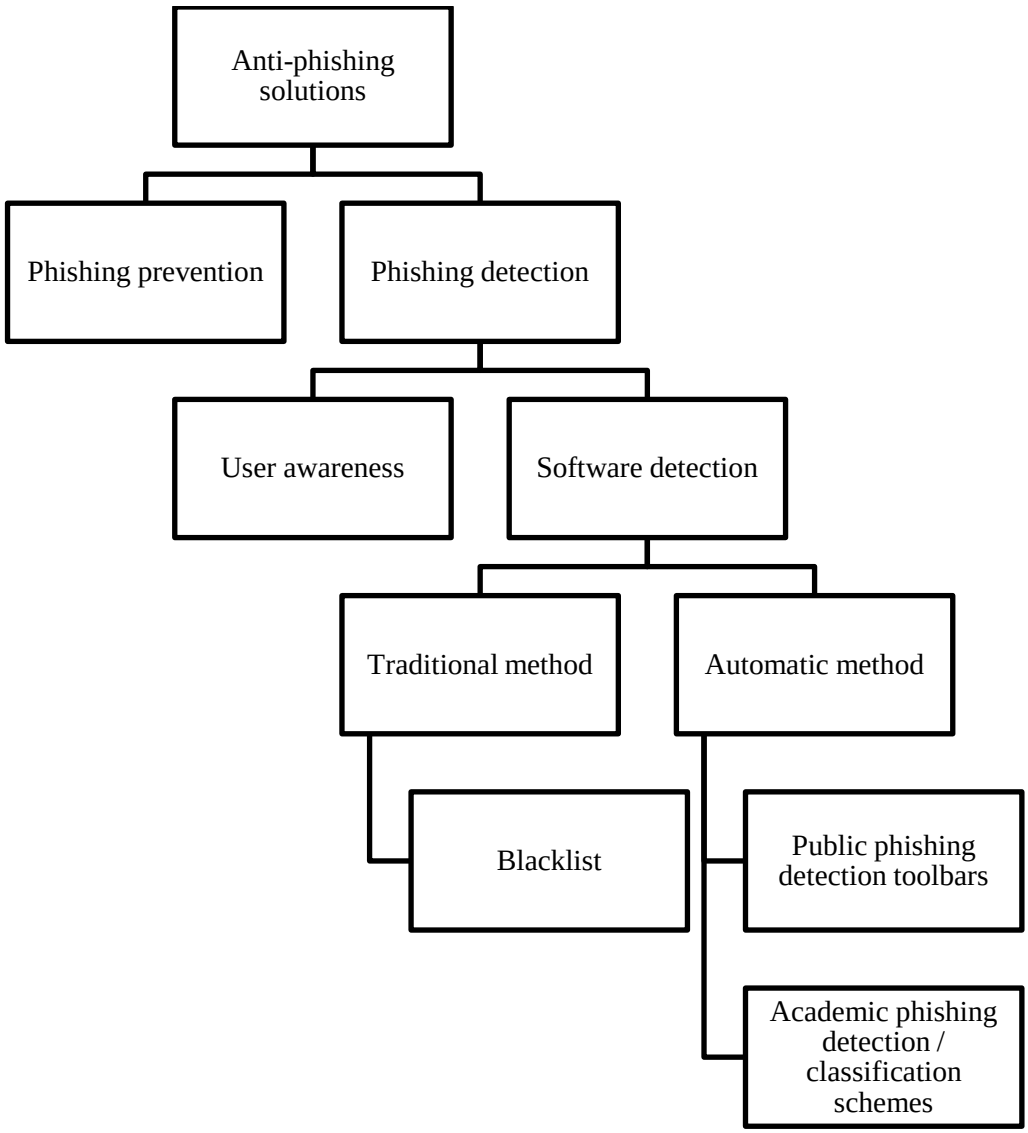

Figure 2. Types of anti-phishing solutions.

3.1. Phishing prevention

In order to prevent phishing attack, the phishing prevention is introduced by providing an extra layer of security when the user login into the website. The extra layer of security is via two-factor authentication, which is a process to confirm the user identity before the user is granted to access its login account in the website.

Figure 3 shows an example of two-factor authentication via SMS. When the user has entered username and password to login into the website, a verification code will be sent to the user's registered mobile phone number via SMS. Then, the user need to enter the verification code before the user can login into the website. The verification code only can be used in a short time before it is expired.

Even though the phishing prevention can provide extra security to prevent the phishing attack, it may hard to be implementing. An extra device is required for the user to get the verification to access into the website. This can cause extra cost to implement the extra security for the phishing attack.

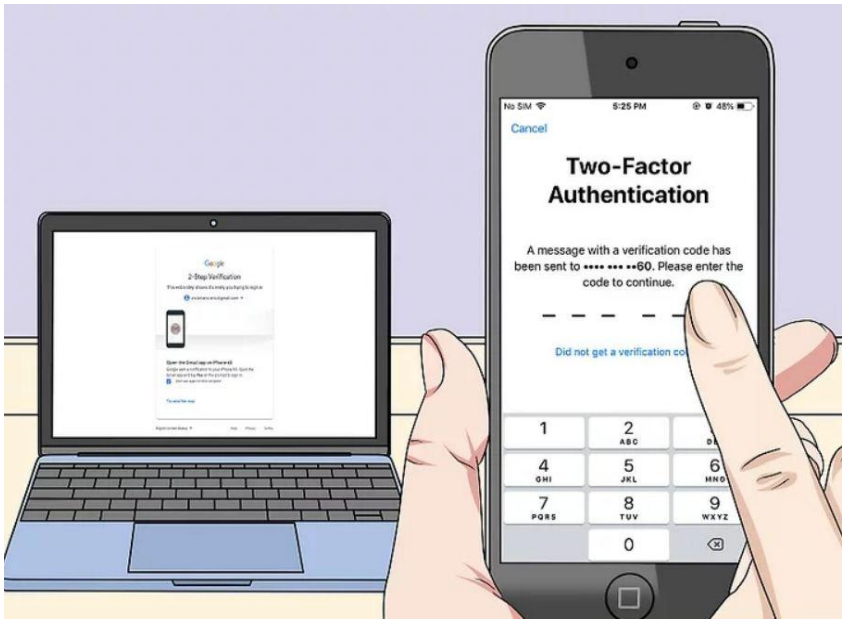

Figure 3. Two-factor authentication via SMS [7].

3.2. Phishing detection

There are two categories for phishing detection which are user awareness and software detection as shown in Figure 2. The user awareness is for to educate users so that they are able to identify phishing attempts targeted at them. The users need to be careful when visiting the web page, for example by checking at the web page URL first. Even though the user has being careful, there is a chance that the user can be deceived by the phisher to visit the phishing web page [3, 8].

Therefore, software detection is introduced to use for distinguish whether the website is legitimate or phishing. The software detection can be differentiating into two methods which are traditional and automatic method as shown in Figure 2 [2].

For traditional method of software detection, the blacklist is used to manage the list of phishing websites, which are manually entered and updated in the system. The advantage of blacklist is it has high accuracy. The drawback of blacklist is it lacking to identify the phishing website that has short lifetime [1, 2]. Furthermore, if no one report about the phishing website, then the blacklist cannot detect the phishing website [9].

For automatic method of software detection, it can be classified into two categories which are public phishing detection toolbars and academic phishing detection / classification schemes as shown in Figure 2. The automatic method of software detection use combinations of heuristic and blacklist based approach. The heuristic based approach examines contents of the website. There are three types of heuristic based approach which are surface level content, textual content and visual content. The heuristic of surface level content means by examine at the URL of website. The heuristic of textual content means by examine the terms or words in the website. Lastly, the heuristic of visual content means by examine the layout of website [1]. Next, the public phishing detection toolbars and academic phishing detection / classification schemes is briefly described.

The purpose of the public phishing detection toolbars is to detect and blocking the phishing website. The user can see these toolbars as a web browser extension. A security warning is displayed to alert the Internet user when the user visits the phishing website. There are two types of security warning which are passive warning and active warning. For passive warning, it does not block the content of the website and just show the warning to notify the user about the phishing attack as shown in Figure 4. While for active warning, it block the content of the website, thus make the user unable to view the website as shown in Figure 5. There are many public phishing detection toolbars that are freely available. Some of them are Google Safe Browsering, McAfee SiteAdvisor, Netcraft Anti-

Phishing Toolbar and many more [1, 10]. This is a useful tool as a protection for the Internet users who has lack knowledge to identify the phishing attack themselves. For public phishing detection toolbars, it is hard to know on the algorithm used for phishing detection.

Figure 4. Passive warning in Internet Explorer (IE) [10].

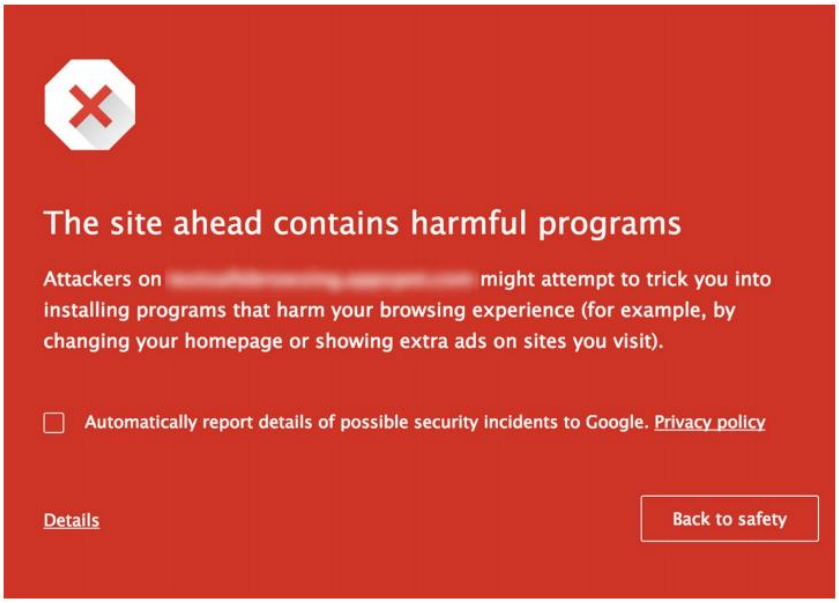

Figure 5. Active warning in Google Safe Browsering [1].

The purpose of the academic phishing detection / classification schemes is to identify and classify whether the website is legitimate or phishing. It utilizes Artificial Intelligence (AI) method which uses supervised learning classification algorithms to do binary classification of website whether it is legitimate or phishing [11]. The algorithm used includes machine learning and deep learning. There are many types of machine learning and deep learning algorithms. The common machine learning algorithms used for the phishing detection techniques are Support Vector Machine, Logistic

Regression and Bayesian-based classifiers [1]. The machine learning algorithm has a limitation which is it become ineffective when dealing with a large scale datasets [12]. In order to improve the performance of the academic phishing detection / classification schemes, the deep learning algorithm is introduced. Compared to machine learning algorithm, the deep learning algorithm is able to handle large scale datasets [1, 8]. However, the deep learning algorithm need more time for training [8]. The comparison performance of phishing website detection by using machine learning and deep learning algorithm can be seen in [13], where it shows deep learning algorithm namely deep belief network achieve high accuracy which is 0.94 compared to the accuracy of the other types of machine learning algorithm which are Random forest achieve 0.92 and J48 achieve 0.89. The challenging task to build algorithm for the academic phishing detection / classification schemes is feature extraction or also known as feature engineering. The website features that are useful for the phishing detection need to be selected in order to improve the algorithm accuracy. The evaluation metrics used to measure the performance of the academic phishing detection / classification schemes are accuracy, precision, recall and F1-score.

Based on the discussion of the phishing detection, we have summarized some important point regarding the category of phishing detection as shown in Table 1.

Table 1. Comparison on category of phishing detection.

Category of phishing detection	User awareness	Software detection		
		Traditional method	Automatic method	
		Blacklist	Public phishing detection toolbars	Academic phishing detection / classification schemes
Aim	Detect phishing attacks based on the user knowledge	Manage list of phishing website	Detect and block phishing website	Identify and classify whether the website is legitimate or phishing
Characteristics	User need to carefully check the URL website before visit the website	Blacklist based approach	Combinations of heuristic and blacklist based approach	
			Little information about algorithm used to detect phishing attack	Algorithm used to detect phishing attack by using machine learning or deep learning
			Freely available to use	Not available for public use
Advantage		High accuracy	Real-time system that determine whether the website is legitimate or phishing	
				Reduce the number of false positive

				rates
Disadvantage	Need to educate the user about phishing attack	Unable to detect zero-hour phishing attacks which mean attacks that were not seen previously	High false positive rates	
	Human error or ignorance to detect phishing attacks	Easy for the phisher to escape from phishing detection	User ignore the security warning displayed	
		Require frequent update on the dataset		
		Too slow to handle large scale datasets		

4. Conclusion

This paper has discussed about phishing attack and anti-phishing solutions. This paper can be used as a guideline for the researchers who are interested to know about the anti-phishing solutions exists in order to solve the phishing attack. There are two types of anti-phishing solutions which are phishing prevention and phishing detection. We need to do phishing detection first before we can do phishing prevention. If the phishing attack is not detected first, then it is no use to do the phishing prevention. After the phishing attack has been detected, only then the phishing prevention can be applied. This shows the importance of the phishing detection compared to the phishing prevention. There are many categories of the phishing detection which are user awareness, traditional method of software detection, public phishing detection toolbars and academic phishing detection / classification schemes. Based on the comparison of the categories of the phishing detection, the academic phishing detection / classification schemes are more useful in order to detect the phishing attack. For the future work, we would like to explore more on the academic phishing detection / classification schemes that utilizes on deep learning. We choose to focus on deep learning as the phishing detection technique because it is the recent technique being used and it has shown the improved performance in terms of accuracy of phishing detection compared to the machine learning.

5. Acknowledgment

The work presented in this paper is supported by Universiti Malaysia Pahang Grant: RDU182207-2. This work also partially supported by Adnuri SMA Research Center (M) Sdn. Bhd.

References

[1] Z. Dou, I. Khalil, A. Khreishah, A. Al-Fuqaha, and M. Guizani, "Systematization of Knowledge (SoK): A systematic review of software-based web phishing detection," *IEEE Communications Surveys & Tutorials,* vol. 19, pp. 2797-2819, 2017.

[2] R. Das, M. Hossain, S. Islam, and A. Siddiki, "Learning a deep neural network for predicting phishing website," Degree of B.Sc. in Computer Science, Brac University, 2019.

[3] G. Varshney, M. Misra, and P. K. Atrey, "A survey and classification of web phishing detection schemes," *Security and Communication Networks,* vol. 9, pp. 6266-6284, 2016.

[4] K. L. Chiew, K. S. C. Yong, and C. L. Tan, "A survey of phishing attacks: their types, vectors and technical approaches," *Expert Systems with Applications,* vol. 106, pp. 1-20, 2018.

[5] I. Qabajeh, F. Thabtah, and F. Chiclana, "A recent review of conventional vs. automated cybersecurity anti-phishing techniques," *Computer Science Review,* vol. 29, pp. 44-55, 2018.
[6] H. Thakur and S. Kaur, "A survey paper on phishing detection," *International Journal of Advanced Research in Computer Science,* vol. 7, 2016.
[7] w. Staff. (2019). *How to Prevent Phishing*. Available: https://www.wikihow.com/Prevent-Phishing
[8] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," *Expert Systems with Applications,* vol. 117, pp. 345-357, 2019.
[9] A. I. Abunadi, "Anti-phishing Model for Phishing Websites Detection: Using Pruning Decision Tree," Universiti Teknologi Malaysia, 2013.
[10] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey," *IEEE Communications Surveys & Tutorials,* vol. 15, pp. 2091-2121, 2013.
[11] V. Ra, B. G. HBa, A. K. Ma, S. KPa, P. Poornachandran, and A. Verma, "DeepAnti-PhishNet: Applying deep neural networks for phishing email detection," in *Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal.(IWSPA)*, 2018, pp. 1-11.
[12] L. Safae, B. El Habib, and T. Abderrahim, "A Review of Machine Learning Algorithms for Web Page Classification," in *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, 2018, pp. 220-226.
[13] M. K. Verma, S. Yadav, B. K. Goyal, B. R. Prasad, and S. Agarawal, "Phishing Website Detection Using Neural Network and Deep Belief Network," in *Recent Findings in Intelligent Computing Techniques*, ed: Springer, 2019, pp. 293-300.