Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

An effective security alert mechanism for real-time phishing tweet detection on Twitter

Seow Wooi Liew, Nor Fazlida Mohd Sani*, Mohd. Taufik Abdullah, Razali Yaakob, Mohd Yunus Sharum

Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, UPM Serdang, Selangor Darul Ehsan, Malaysia

ARTICLE INFO

Article history:
Received 13 August 2018
Accepted 10 February 2019
Available online 19 February 2019

Keywords:
Phishing
Twitter
Classification model
Random Forest (RF)
Security alert mechanism

ABSTRACT

Phishing is a form of social engineering crime uses to deceive victims by directing them to a fraudulent website where their private and confidential information are collected for further illegal actions. Phishing attacks have now targeted users at Online Social Networks (OSN)s such as Twitter, Facebook, Myspace, etc. which traditionally, targeting email users. Twitter has become so prevalent to phishers to spread phishing attacks nowadays due to its vast information dissemination and difficult to be detected unlike email. As such, the effectiveness of security alert to prompt Twitter users for the tweet containing phishing Uniform Resource Locator (URL) in real-time is crucial. Many solutions have been proposed but their effectiveness are inadequate and doubtful. In this paper, we propose an effective security alert mechanism making use of a classification model derived from a supervised machine learning technique of Random Forest (RF) and the identified 11 best classification features yielded 94.75% accuracy higher than 94.56% yielded by other researchers who used more than 11 features trained on the same dataset collected from Twitter. To determine its effectiveness, we used 200 phishing URLs collected from Twitter and PhishTank respectively. From our experiment, we are able to justify that such proposed security alert mechanism managed to prompt 97.50% effectively the security alert to Twitter users in real-time.

1. Introduction

The rapid growth of Information Technology indeed created many conveniences to us, but on the other hand it also resulted and increased security challenges to us to protect our information securely especially from social engineering attacks nowadays.

Social engineering is an art of getting users to compromise information systems (Krombholz et al., 2015), an information gathering form that involves human intervention to breach security without victims realizing they have been manipulated and a real problem we encounter today. In general, it can be regarded as "people hacking" (Hasle et al., 2005) and interpreted as a method of launching attacks against information and information systems (Janczewski and Fu, 2010). Social engineering has been created challenging risks to both individuals for personal data and organizations for customer data. Individuals are more vulnerable to social engineering attacks because they never expect themselves ever be the victims and never know they are victims of such attacks (Bezuidenhout et al., 2010). Normally, a social engineering attack success often depends on a target either being willing or tricked into revealing personal information (Junger et al., 2017). In summary,

* Corresponding author.
E-mail addresses: liewsw28@gmail.com (S.W. Liew), fazlida@upm.edu.my (N.F.M. Sani), taufik@upm.edu.my (Mohd.T. Abdullah), razaliy@upm.edu.my (R. Yaakob), m_yunus@upm.edu.my (M.Y. Sharum).

https://doi.org/10.1016/j.cose.2019.02.004
0167-4048/© 2019 Elsevier Ltd. All rights reserved.

social engineering is an art of influencing people to divulge private and confidential information, and social engineering attacks are processes of doing so (Mouton et al., 2014).

Phishing is a form of social engineering crime called semantic attack and well recognized as online identity theft that deceives victims by directing them to a fraudulent website appears legitimate one (Arachchilage and Love, 2013, 2014; Arachchilage et al., 2016) which will then collect their private and confidential information. It is a huge problem that is getting bigger each month (Toolan and Carthy, 2009), a growing problem on internet today because it is one of the most common approaches uses to capture and gather personal information from victims (Shekokar et al., 2015), an effort for accessing people important information (Dadkhah et al., 2015) and a cyber-threat that was first identified in 1996 (Varshney et al., 2016). Phishers use fraudulent websites accurately mimic genuine ones to create the phishing attacks unnoticed (Moreno-Fernandez et al., 2017).

Traditionally, phishing attacks target email which serves as the primary vector (Wilcox and Bhattacharya, 2015). However, they have now exposed into popularity of OSNs. According to Anti-Phishing Working Group (APWG) – a non-profit organization's survey, OSNs have become a significant platform where phishers launching phishing attacks. Twitter has become a medium used by phishers today to spread phishing due to its vast information dissemination and difficult to be detected unlike email because of its fast spread in the network, short content size and short URL (Aggarwal et al., 2012; Nair and Prema, 2014). Basically, Twitter is an immensely popular micro-blogging network where people post short messages of 140 characters called tweets (McCord and Chuah, 2011; Aggarwal et al., 2012; Lee and Kim, 2013; Nair and Prema, 2014) and a critical source for real-time information sharing and news dissemination (Liu et al., 2017). Unlike other OSNs such as Facebook or Myspace, Twitter relationship in term of following and being followed do not require reciprocation (Kwak et al., 2010). As such, it becomes so popular used by people. Due to its popularity, it is always being focused by malicious users who often try to find a way to attack (Lee and Kim, 2013). Since Twitter has become so prevalent to phishers to spread phishing attacks, an effective security alert mechanism for real-time phishing tweet detection to prompt effectively the security alert to Twitter users is deem required.

The contributions of this paper contain 2 aspects. Firstly, we provide a literature study on some existing solutions using machine learning classification features for phishing detection and security alert for real-time phishing tweet detection on Twitter. Secondly, we propose an effective security alert mechanism for real-time phishing tweet detection on Twitter using a classification model derived from a supervised machine learning technique of RF and the identified best classification trained on a dataset collected from Twitter (Sharma et al., 2014).

2. Literature study

This section emphasized on the related works pertaining to the use of machine learning classification features for phishing detection and security alert for real-time phishing tweet detection on Twitter.

2.1. Machine learning classification features for phishing detection

In McCord and Chuah (2011), they selected user based and content based features in their machine learning classification for phishing detection. Basically, they referred the user based features to the user's relationships such as follower and followee or user behaviors, and content based features to the average length of a tweet, number of URLs, replies or mentions, keywords or wordweight, retweets or tweetlen and hashtags.

Aggarwal et al. (2012) managed to achieve high phishing detection accuracy with the use of many different features like suspicious URL properties, tweet content (a short message with 140 characters), Twitter user posting the tweet attributes and details regarding the phishing domains to detect phishing tweets. Table 1 shows a list of the features used by them in more details.

As for Sharma et al. (2014), they used 6 sets of features containing URL based, tweet based, WHOIS based, user based and network based features in their machine learning classification phishing detection experiment. According to them, they carried out the classification started with 1 set of features followed by adding on another set of features in the next classification activity until the total 6 sets of features were used completely. From their experiment, they concluded that the performance of the phishing detection significantly improved when more sets of features typically the tweet based features were added in the classification.

Basnet et al. (2014) grouped features into lexical based, keyword based, search engine based and reputation based. Each and every group contains number of features. As emphasized by them, they used 138 features (Table 2) in their phishing detection experiment.

In Sananse and Sarode (2015), they used lexical based, WHOIS based, pagerank based, Alexa rank based and PhishTank based features for phishing detection. Table 3 shows a list of the features used by them in more details.

Another experiment from Akanbi et al. (2015) indicating that they managed to yield high phishing detection accuracy using only 9 selected features (Table 4).

From the study, it was noted that the 9 selected features used by Akanbi et al. (2015) could be effective machine learning classification features because such features allowed them to yield high phishing detection accuracy with less features compared to other researchers discussed earlier; which may be possible to be explored in our machine learning training experiment.

2.2. Security alert for real-time phishing tweet detection on Twitter

Aggarwal et al. (2012) proposed a technique called "PhishAri" works in real-time to detect phishing tweet on Twitter using machine learning technique. Basically, the proposed technique contained 2 critical components – "PhishAri" Application Programming Interface (API) where the machine learning

Table 1 – Features Used by Aggarwal et al. (2012).

Feature Group	Feature	Feature Detail / Description
URL based	Length of URL	Length of expanded URL in number of characters.
	Number of dots	Number of dots (.) used.
	Number of subdomains	Number of subdomains (marked by /) in the expanded URL.
	Number of redirections	Number of hops between the posted URL and the landing page
	Levenshtein distance between redirected hops	Avg levenshtein distance between length of redirected URLs between original & final URL.
	Presence of conditional redirects	Whether the URL is redirected to different landing page for browser or an automated program.
WHOIS based	Registering domain name	Name of the domain provider.
	Ownership period	Age of the domain.
	Time taken to create Twitter account	How much time elapsed between creation of domain and the Twitter account.
Tweet based	Number of #tags	Number of topics mentioned in tweet.
	Number of @tags	Number of Twitter users mentioned in tweet.
	Presence of trending #tags	Number of topics mentioned which were trending at that time.
	Number of RTs	Number of times the tweet was reposted.
	Length of tweet	Length of tweet in number of characters.
	Position of #tags	Number of characters of tweets after which the #tag appears.
Network based	Number of Followers	Number of Twitter users who follow this Twitter user.
	Number of Followees	Number of Twitter users who are being followed by this Twitter user.
	Ratio of Followers-Followees	Number of Followers / Number of Followees.
	Part of Lists	Whether the Twitter user is part of a public list.
	Age of account	How old the Twitter account is.
	Presence of description	Whether the Twitter account has a profile description.
	Number of Tweets	Number of tweets posted by the Twitter user.

Table 2 – Features used by Basnet et al. (2014).

Feature Group	Number of features	Example of features
Lexical based	24	Length of URL, Length of host, Length of path, Digit in host, etc.
Keyword based	101	Login, Signin, Confirm, Verify, etc.
Search engine based	6	Google page rank, Age of domain, etc.
Reputation based	7	PhishTank top 10 domain/target in URL, PhishTank top 10 target in URL, IP in PhishTank top 10 IPs, IP in StopBadware top 50 IPs, URL in phishing blacklist, URL in malware blacklist, etc.

phishing tweet detection mechanism resided, and browser extension served as the front-end interface for Twitter users. To achieve such real-time phishing tweet detection on Twitter, the browser extension is first to capture the inputted tweet id and URL followed by sending them to the "PhishAri" API for phishing tweet detection. Once completed, the result of the detection is returned from "PhishAri" API to the browser extension along with an indicator of red for phishing or green for safe to be displayed at the user's Twitter page accordingly. Nevertheless, they declared that their method may misjudge the phishing tweet as a legitimate tweet if a malicious user makes the phishing tweet looks like a legitimate tweet and the Twitter network features stated it that of a legitimate user. In addition to that, it was noted that Twitter users are always required to sign in to the Twitter in order to obtain such indicator displayed in their Twitter's page and Sharma et al. (2014) made claims that "PhishAri" is no longer in use now because its front-end interface could not work with the "PhishAri" API as Twitter now required Oauth for authentication and various security parameters appliance issue with Twitter encountered.

Similar to Aggarwal et al. (2012), Nair and Prema (2014) proposed a system called "Distributed Phishing Detection System" using machine learning technique works in real-time to detect phishing on Twitter too. A browser is built to deliver a real-time phishing detection result by appending a red indicator next to the phishing tweets for Twitter users. Their system served as an extension to "WarningBird" proposed by Lee and Kim (2013) which is not focusing on phishing URL detection. Anyhow, it was noted that the "Distributed Phishing Detection System" has a dependency on the collected tweets as it used a tweet window instead of individual tweets to determine whether they are sufficient enough before proceeding to the next process of the phishing tweet detection processes.

Another real-time phishing tweet detection on Twitter called "Web Framework" proposed by Sharma et al. (2014). This framework basically is an end user phishing detection system which takes the tweet id and specific keyword as an input, and return the result of detection with a red background for phishing URL or green background for safe URL. This is achieved by first checking with the APIs from PhishTank, Google Safe Browsing and Web of Trust (MyWOT) followed by checking with a trained or already classified model (called Classification Model) derived from RF machine learning technique and the best phishing classification features as the last source. Along with these APIs, they also created a database to store all the new URLs which are not found in the mentioned

Table 3 – Features used by Sananse and Sarode (2015).

Feature	Number of features	Feature description
Lexical based	24	URL properties such as Length of URL, Length of host, Length of path, Digit in host, etc.
WHOIS based	48	Properties that explained who, where and how of the websites.
Pagerank based	1	Technique to determine the number and quality of links to a page. The purpose is to decide how essential the particular website can be. If the page is important, it should have more links where other websites link to it. The value of pagerank from Google is robust and updated frequently.
Alexa rank based	1	Ranking set by Alexa with the purpose to audit the visit frequency on numerous websites and available to public for reference. The parameters of the traffic based are reach (the number of Alexa users visiting a particular site in one day) and page views (the number of times a particular URL is viewed by Alexa users).
PhishTank based	>1	Phishing websites historical statistical reports on every month. The URL host belongs to the top IP or domain that hosts phishing websites listed in the reports will be used as reference to determine whether the URL is phishing.

Table 4 – Features used by Akanbi et al. (2015).

Feature	Feature description
Long URL	Length of URL
Dots	Number of dots existed in a URL
IP-address	IP address existed in a URL
SSL connection	Https connection existed in a URL
At “@” symbol	Symbol “@” existed in a URL
Hexadecimal	Symbol “%” existed in a URL e.g. “http://%30%31%30%/paypal/cgi=bin/webscrcmd_login.asp”
Frame	Frame existed in a URL
Redirect	Redirect existed in the URL e.g. “www.facebook.com/2/12432;phish.com”
Submit	Submit button existed in the URL and source code

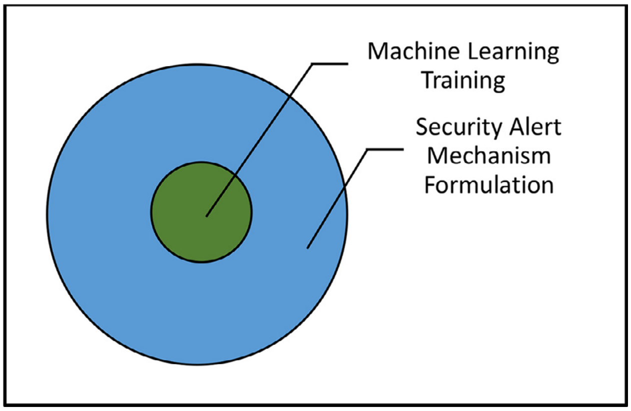

Machine Learning Training
Security Alert Mechanism Formulation

Fig. 1 – Security alert mechanism.

Table 5 – Confusion matrix.

		PREDICTED	
		Phishing	Safe
ACTUAL	Phishing	TP	FN
	Safe	FP	TN

APIs but classified by the trained classification model for future URL phishing detection. “Web Framework” did seem to be an effective real-time system but unfortunately it did not really integrate seamlessly the phishing tweets detection result into the user’s Twitter page as it required user intervention to input the user id or specific keyword manually into the system every time in order to obtain a red background for phishing URL or green background for safe URL.

With the discussed shortfalls, it was evident that there are still rooms to improve the effectiveness of security alert for real-time phishing tweet detection to Twitter users in real-time.

3. Security alert mechanism design

The design of the entire security alert mechanism was divided into 2 component stages. They are machine learning training and security alert mechanism formulation (Fig. 1).

3.1. *Machine learning training*

In this stage, we adopted a supervised machine learning approach; where a machine learning technique of RF was selected to train on a dataset containing 2973 training data of 1573 labelled as phishing URLs and 1400 labelled as safe URLs collected from Twitter (Sharma et al., 2014) using WEKA tool.

RF was selected because it was claimed by Sharma et al. (2014) as the best machine learning technique allowing them to yield high accuracy of 94.56% and due to RF is an ensemble learning classification method uses to handle problems involving data grouping into classes (Wikipedia - RF; Al-Garadi et al., 2016). Furthermore, RF is one of the most accurate classifier which works efficiently for large databases and the most effective method of machine learning algorithms (Aggarwal et al., 2012; Sharma et al., 2014). Basically, RF uses decision trees to predict an output. In RF training phase, decision trees are created followed by the use of class prediction. This is achievable by considering the voted classes from each of the trees where the highest voted class is considered to be the output (Wikipedia - RF).

Table 6 – 11 Best classification features.

Feature	Proposed By	Feature Description
URL length	WEKA tool	Length of URL
SSL connection	WEKA tool	Https connection existed in a URL
Hexadecimal	WEKA tool	Symbol "%" existed in a URL e.g. "http://%30%31%30%/paypal/cgi=bin/webscrcmd_login.asp"
Alexa rank	WEKA tool	A metric that ranks websites in order of popularity or how well a website is doing over the last 3 months
Age of domain - Year	WEKA tool	Length of time a website has been registered and active
Equal	Feature selection method & Maximal relevance	Symbol "=" existed in a URL e.g. "http://%30%31%30%/paypal/cgi=bin/webscrcmd_login.asp"
Digit in host	WEKA tool	Digits existed in Host e.g. www.3sports.com
Host length	Feature selection method & Maximal relevance	Length of host
Path length	Feature selection method & Maximal relevance	Length of path
Registrar	WEKA tool	Registrar existed in WHOIS
Number of dots in host	Feature selection method & Maximal relevance	Number of dots existed in host

As for the classification features, we selected the machine learning classification features used by McCord and Chuah (2011), Aggarwal et al. (2012), Sharma et al. (2014), Basnet et al. (2014), Sananse and Sarode (2015), and Akanbi et al. (2015) as the basis of features to be explored for best features determination and set 94.56% accuracy achieved by Sharma et al. (2014) as the targeted baseline for classification accuracy improvement in our experiment.

To evaluate the effectiveness of the classification, Standard Information Retrieval Metrics viz. Accuracy, Precision and Recall, and a Confusion Matrix (Table 5) were used.

Where TP - True Positive,
FP - False Positive,
TN - True Negative,
FN - False Negative

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{1}$$

$$\text{Precision}_{(\text{Phishing})} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall}_{(\text{Phishing})} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

In addition, cross validation of 10 folds test mode was selected and used in our experiment because the data available is limited to 2973. The purpose of selecting such test mode is to ensure all available data are used to train the classification model and to compare on the test set in a particular division respectively. In summary, all data in the dataset are used for both training and testing.

3.2. *Security alert mechanism formulation*

With the completion of machine learning training stage, we initiated and conducted the security alert mechanism formulation. In this stage, the classification model derived from the machine learning technique of RF and the identified best classification features was then embedded into our proposed security alert mechanism in order to prompt a security alert for the tweet containing phishing URL to Twitter users in real-time.

Table 7 – RF predicted results in confusion matrix.

		PREDICTED	
		Phishing	Safe
Actual	**Phishing**	1502	71
	Safe	85	1315

Table 8 – RF precision and recall for phishing.

Description	Achieved
$\text{Precision}_{(\text{Phishing})}$	94.64%
$\text{Recall}_{(\text{Phishing})}$	95.49%

To justify the effectiveness of the proposed security alert mechanism, we extracted 100 phishing URLs in the dataset collected from Twitter (Sharma et al., 2014) and 100 phishing URLs from PhishTank respectively to test on it. These 200 phishing URLs were posted one by one on Twitter; where the result for each and every security alert prompted for the phishing URL was observed and recorded accordingly.

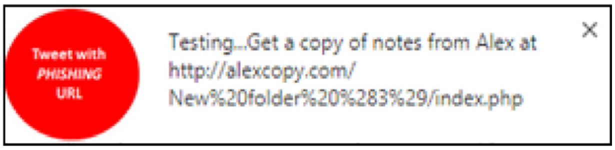
The phishing URL is deemed to be prompted as phishing URL correctly if it matches the actual phishing labelling for the same URL in the dataset collected from Twitter (Sharma et al., 2014) and PhishTank respectively.

4. Experimental results and discussion

From the machine learning training experiment, 11 classification features as listed in Table 6 were identified as the best classification features because they allowed us to achieve

Table 9 – Achieved results for the 200 phishing URLs.

Description	Number of phishing URLs	Number of security alert prompted as phishing URL correctly	Number of security alert prompted as phishing URL wrongly
Number of phishing URLs collected from Twitter (Sharma et al., 2014)	100	98	2
Number of phishing URLs collected from PhishTank	100	97	3
Total number of Phishing URLs	200	195	5



Testing...Get a copy of notes from Alex at http://alexcopy.com/New%20folder%20%283%29/index.php

Fig. 2 – Prompted security alert with red indicator. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

94.75% (Eq. (1)) classification accuracy higher than 94.56% classification accuracy achieved by Sharma et al. (2014) who used more than 11 features trained on the same dataset collected from Twitter. 7 out of the 11 best classification features were proposed by the WEKA tool under Attribute Evaluator as “CfsSubsetEval” and Search Method as “BestFirst”. The remaining 4 were selected from the pool of 22 features gathered from the set used by McCord and Chuah (2011), Aggarwal et al. (2012), Sharma et al. (2014), Basnet et al. (2014), Sananse and Sarode (2015), and Akanbi et al. (2015) manually according to the Feature Selection Method suggested by Dash and Liu (1997), Liu and Yu (2005) and Tang et al. (2014) and base on the Maximal Relevance (Chen et al., 2006; Zhao et al., 2010; Mandal and Mukhopadhyay, 2013) that are most relevant to the target class and highly affecting the classification output.

RF was able to predict 1502 URLs as phishing and 1315 URLs as safe correctly (Table 7), and managed to achieve 94.64% precision and 95.49% recall for phishing using Eqs. (2) and (3) respectively as shown in Table 8.

As for the experiment to determine the effectiveness of the proposed security alert mechanism, we successfully managed to obtain the security alert for real-time phishing tweet detection prompted correctly with 97.50%; in which 98 out of 100 phishing URLs collected from Twitter (Sharma et al., 2014) and 97 out of 100 phishing URLs collected from PhishTank, respectively (Table 9).

In summary, this indicated that our proposed security alert mechanism is effective enough to prompt 97.50% security alerts as phishing URL correctly for phishing URLs collected from Twitter (Sharma et al., 2014) and PhishTank respectively in total. Fig. 2 shows an example of security alert with red indicator prompted by our proposed security alert mechanism.

5. Conclusion

From the experiment, it shows significantly that our proposed security alert mechanism is able to prompt 97.50% effectively the security alert to Twitter users in real-time when making use of a classification model derived from the machine learning technique of RF and the 11 best classification features - URL length, SSL connection, Hexadecimal, Alexa rank, Age of domain - Year, Equal, Digit in host, Host length, Path length, Registrar and Number of dots in host.

Acknowledgment

Our special thanks go to Ms. Smriti Maheshwari who has been contributing of the dataset collected from Twitter used in this study.

REFERENCES

Aggarwal A, Rajadesingan A, Kumaraguru P. PhishAri: automatic realtime phishing detection on Twitter. In: Proceedings of the ECrime researchers summit, ECrime; 2012. p. 1–12. doi:10.1109/eCrime.2012.6489521.

Akanbi OA, Amiri IS, Fazeldehkordi E. A Machine-learning approach to phishing detection and defense. Elsevier Inc.; 2015. doi:10.1016/B978-0-12-802927-5.00002-2.

Al-Garadi MA, Varathan KD, Ravana SD. Cybercrime detection in online communications: the experimental case of cyberbullying detection in the Twitter network. Comput Human Behav 2016;63:433–43. doi:10.1016/j.chb.2016.05.051.

Anti-Phishing Working Group (APWG). (n.d.). Retrieved May 1, 2016, from http://www.antiphishing.org/.

Arachchilage NAG, Love S. A game design framework for avoiding phishing attacks. Comput Human Behav 2013;29(3):706–14. doi:10.1016/j.chb.2012.12.018.

Arachchilage NAG, Love S. Security awareness of computer users: a phishing threat avoidance perspective. Comput Human Behav 2014;38:304–12 Retrieved from. doi:10.1016/j.chb.2014.05.046.

Arachchilage NAG, Love S, Beznosov K. Phishing threat avoidance behaviour: an empirical investigation. Comput Human Behav 2016;60:185–97. doi:10.1016/j.chb.2016.02.065.

Basnet RB, Sung AH, Liu Q. Learning to detect phishing URLs. Int J Res Eng Technol 2014;3(6):11–24. doi:10.1109/ICMLA.2012.104.

Bezuidenhout M, Mouton F, Venter HS. Social engineering attack detection model: SEADM. Inform Secur South Africa 2010:1–8.

Chen Y, Li Y, Cheng X-Q, Guo L. Survey and taxonomy of feature selection algorithms in intrusion detection system. Inform Secur Cryptol 2006;4318:153–67. doi:10.1007/11937807_13.

Dadkhah M, Sutikno T, Jazi MD, Stiawan D. An introduction to journal phishings and their detection approach. Telkomnika (Telecommun Comput Electron Contr) 2015;13(2):373–80. doi:10.12928/TELKOMNIKA.v13i2.1436.

Dash M, Liu H. Feature selection for classification. Intel Data Anal 1997;1(3):131–56. doi:10.3233/IDA-1997-1302.

Google Safe Browsing. (n.d.). Retrieved May 15, 2016, from https://safebrowsing.google.com/.

Hasle H, Kristiansen Y, Kintel K, Snekkenes E. Measuring resistance to social engineering. Inform Secur Pract Exper 2005:132–43. doi:10.1007/978-3-540-31979-5_12.

Janczewski LJ, Fu L. Social Engineering-based attacks: model and New Zealand perspective. In: Proceedings of the international multiconference on computer science and information technology; 2010. p. 847–53.

Junger M, Montoya L, Overink F-J. Priming and warnings are not effective to prevent social engineering attacks. Comput Human Behav 2017;66:75–87. doi:10.1016/j.chb.2016.09.012.

Krombholz K, Hobel H, Huber M, Weippl E. Advanced social engineering attacks. J Inform Secur Appl 2015;22:113–22. Retrieved from https://pdfs.semanticscholar.org/3266/f05e2e5e785cbab72d2e378059ecc62ef706.pdf.

Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media. In: Proceedings of the international World wide web conference committee (IW3C2); 2010. p. 1–10. doi:10.1145/1772690.1772751.

Lee S, Kim J. Warningbird: a near real-time detection system for suspicious URLs In Twitter stream. IEEE Trans Depend Secure Comput 2013;10(3):183–95. doi:10.1109/TDSC.2013.3.

Liu H, Yu L. Toward integrating feature selection algorithms for Huan liuclassification and clustering. IEEE Trans Knowl Data Eng 2005;17(4):491–502. doi:10.1109/TKDE.2005.66.

Liu S, Wang Y, Zhang J, Chen C, Xiang Y. Addressing the class imbalance problem in twitter spam detection using ensemble learning. Computers & Security 2017;69:35–49. doi:10.1016/j.cose.2016.12.004.

Mandal M, Mukhopadhyay A. An improved minimum redundancy maximum relevance approach for feature selection in Gene expression data. Procedia Technol 2013;10:20–7. doi:10.1016/j.protcy.2013.12.332.

McCord M, Chuah M. Spam detection on Twitter using traditional classifiers. In: Proceedings of the ATC; 2011. p. 175–86. https://doi.org/0.1007/978-3-642-23496-5_13.

Moreno-Fernandez MM, Blanco F, Garaizar P, Matute H. Fishing For Phishers. Improving internet users' sensitivity to visual deception cues to prevent electronic fraud. Comput Human Behav 2017;69:421–36. doi:10.1016/j.chb.2016.12.044.

Mouton F, Malan MM, Leenen L, Venter HS. In: Proceedings of the 2014 information security for South Africa. Social engineering attack framework; 2014. doi:10.1109/ISSA.2014.6950510.

Nair MC, Prema S. A distributed system for detecting phishing in Twitter stream. Int J Eng Sci Innov Technol 2014;3(2):151–8.

PhishTank. (n.d.). Retrieved May 15, 2016, from https://www.phishtank.com/.

Sananse BE, Sarode TK. Phishing URL detection: a machine learning and web mining-based approach. Int J Comput Appl 2015;123(13):46–50.

Sharma N, Sharma N, Tiwari V, Chahar S, Maheshwari S. Real-time detection of phishing Tweets. In: Proceedings of the fourth international conference on computer science, engineering and applications; 2014. p. 215–27. doi:10.5121/csit.2014.4727.

Shekokar NM, Shah C, Mahajan M, Rachh S. An ideal approach for detection and prevention of Phishing attacks. Procedia Comput, Sci, 2015;49(1):82–91. doi:10.1016/j.procs.2015.04.230.

Tang J, Alelyani S, Liu H. Feature selection for classification: a review. Data Classif Algorith Appl 2014:37–64 doi:10.1.1.409.5195.

Toolan F, Carthy J. In: Proceedings of the 2009 ECrime researchers summit, ECRIME '09. Phishing detection using classifier ensembles; 2009. doi:10.1109/ECRIME.2009.5342607.

Varshney G, Misra M, Atrey PK. A phish detector using lightweight search features. Comput Secur 2016;62:213–28. doi:10.1016/j.cose.2016.08.003.

Web of Trust (MyWOT). (n.d.). Retrieved May 15, 2016, from https://www.mywot.com/.

Wikipedia - RF. (n.d.). Random Forest (RF). Retrieved May 15, 2016, from https://en.wikipedia.org/wiki/Random_forest.

Wilcox H, Bhattacharya M. Countering social engineering through social media: an enterprise security perspective. In: Proceedings of the ICCCI; 2015. p. 54–64. doi:10.1007/978-3-319-24306-1_6.

Zhao Z, Morstatter F, Sharma S, Alelyani S, Anand A, Liu H. (2010). Advancing feature selection research – ASU feature selection repository. ASU Feature Selection Repository Arizona State University, 1–28. Retrieve from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.642.5862&rep=rep1&type=pdf.

Liew Seow Wooi is a PhD student in the Department of Computer Science at Universiti Putra Malaysia. His research interests include anti-phishing, information security and information technology risk. Liew obtained a Master of Information Technology in Computer Science from The National University of Malaysia and professional certification of CISM and CRISC from Information Systems Audit and Control Association (ISACA). Contact him at liewsw28@gmail.com.

Nor Fazlida Mohd Sani is an associate professor in the Department of Computer Science at Universiti Putra Malaysia. Her research interests include information security, secure coding, authentication system, intrusion detection system, malware analysis, program understanding and debugging. She obtained a PhD from Universiti Kebangsaan Malaysia. Contact her at fazlida@upm.edu.my.

Mohd. Taufik Abdullah is a senior lecturer at Universiti Putra Malaysia (UPM), leader of Information Security Research Group, UPM and collaborators of Digital Forensics Investigation Research Laboratory, UCD, Dublin, Ireland. His research interests include software engineering, security in computing and digital forensics. he holds many professional certificates such as Digital Etiquette Certification, ECSS, ENSA, CEH V8, and CHFI V8. He obtained a PhD from Universiti Putra Malaysia. Contact him at taufik@upm.edu.my.

Razali Yaakob is an associate professor in the Department of Computer Science at Universiti Putra Malaysia. His research interests include neural network, artificial intelligence and game, and evolutionary computation. He obtained a PhD from University of Nottingham. Contact him at razaliy@upm.edu.my.

Mohd Yunus Sharum is a senior lecturer in the Department of Computer Science at Universiti Putra Malaysia. His research interests include artificial intelligence, natural language processing and computational linguistics. He obtained a PhD from Universiti Putra Malaysia. Contact him at m_yunus@upm.edu.my.