Identifying phishing attacks in communication networks using URL consistency features

Nureni Ayofe Azeez* and
Balikis Bolanle Salaudeen

Department of Computer Sciences,
University of Lagos,
Lagos, Nigeria
Email: nazeez@unilag.edu.ng
Email: salaudeentitibola8529@gmail.com
*Corresponding author

Sanjay Misra

Department of Electrical and Information Engineering,
Covenant University,
Ota, Nigeria
and
Department of Computer Engineering,
Atilim University,
Ankara, Turkey
Email: sanjay.misra@covenantuniversity.edu.ng

Robertas Damaševičius and
Rytis Maskeliūnas

Faculty of Informatics,
Kaunas University of Technology,
Kaunas, Lithuania
Email: robertas.damasevicius@ktu.lt
Email: rytis.maskeliunas@ktu.lt

Abstract: Phishing is a fraudulent attempt by cybercriminals, where the target audience is addressed by a text message, phone call or e-mail, requesting classified and sensitive information after presenting himself/herself as a legitimate agent. Successful phishing attack may result into financial loss and identity theft. Identifying forensic characteristics of phishing attack can help to detect the attack and its perpetuators and as well as to enable defense against it. To shield internet users from phishing assaults, numerous anti-phishing models have been proposed. Currently employed techniques to handle these challenges are not sufficient and capable enough. We aim at identifying phishing sites in order to guard internet users from being vulnerable to any form of phishing attacks by verifying the conceptual and literal consistency between the uniform resource locator (URL) and the web content. The implementation of the proposed PhishDetect method achieves an accuracy of 99.1%; indicating that it is effective in detecting various forms of phishing attacks.

Keywords: phishing attacks; risk assessment; cybersecurity; digital forensics; digital evidence.

Reference to this paper should be made as follows: Azeez, N.A., Salaudeen, B.B., Misra, S., Damaševičius, R. and Maskeliūnas, R. (xxxx) 'Identifying phishing attacks in communication networks using URL consistency features', *Int. J. Electronic Security and Digital Forensics*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes:

Comment [t1]: Author: Please provide the biographical details of each author (not more than 100 words for each author).

This paper is a revised and expanded version of a paper entitled [title] presented at [name, location and date of conference].

Comment [t2]: Author: If a previous version of your paper has originally been presented at a conference please complete the statement to this effect or delete if not applicable.

1 Introduction

Information and communication facilities are easy to use and efficient despite the fact that a large number of people are using the internet in their daily activities. However, as the internet facilitates convenient access to data and classified and confidential information, such access can also cause internet users to lose their personal data and money through the nefarious activities of cybercriminals like hackers and phishers. According to the RSA's online fraud report (RSA, 2018), phishing attacks are on the rise

and accounted for 50% of all cyber attacks observed by RSA in Q3 2018, a 70% increase from the previous quarter.

Phishing attacks often employ social engineering techniques and leverage directing links represented by fake domain names and malicious URLs. Phishing is a strategy being used by criminals to lure and deceive many internet users into revealing their online financial transaction details. Whenever a user is trapped by their deception, the user tends to lose either a big amount of money or confidential information or even both (Prakash et al., 2010). Since phishing attacks aim at taking advantage of human weaknesses it is hard to abate them even for trained users (Sheng et al., 2010). However, the technical devices and methods used by the phishers such as clickjacking, wiphishing, spear phishing and sound-squatting (Chiew et al., 2018) leave traces and can be investigated by the digital forensics methods. Emerging communication infrastructures such as mobile networks (Goel and Jain, 2018), Internet-of-Things (IoT) (Gupta et al., 2017), wireless sensor networks (WSN) (Grover and Sharma, 2016; Fan et al., 2018), fog network (Pham et al., 2018) or power supply and distribution infrastructure (Aichhorn et al., 2018) also have become the target of phishing attacks, while security threats to augmented reality systems (Połap et al., 2017), e-healthcare systems (Yaseen et al., 2018) and assisted living systems (Lauraitis et al., 2019) are causing increasing concern.

Majority of phishing attacks belong to communication network-based phishing and employ identity spoofing. Any form of phishing attack is usually introduced by sending a link believing to be from a genuine source (Chen et al., 2014). The moment a feedback is given by the user, his personal information would have been revealed to the phishers. The concept of phishing which came into existence in 1996 has remained a common term in cybersecurity (Garera et al., 2007). Apart from e-mail, other platforms being used to perpetrate phishing attacks include but are not limited to social networking sites, voice messaging, multiplayer games, SMS (Odusami et al., 2018) and instant messaging (Xiang and Hong, 2009). A phishing website can be used is to clandestinely and fraudulently obtain confidential information such as personal identification numbers (PINs), credit card details and passwords (Abbasi and Chen, 2009) and use it nefariously on behalf of the rightful owner (Liu et al., 2006). In an effort to safeguard internet users from various attacks and subsequently prevent them from losing money through their online transactions, several anti-phishing solutions have been proposed (Khonji et al., 2013; Odun-Ayo et al., 2017). The approaches can be categorised according to the techniques employed as machine learning (Woźniak et al., 2018), text mining and profile matching (Aleroud and Zhou, 2017). Some of the approaches are efficient but not reliable and dependable to cater for the current trend of global phishing strategies:

- CANTINA which is a unique content oriented technique to identify and detect various phishing sites by analysing and examining what is contained in a given webpage for classification as either legitimate or illegitimate. The classification is determined by the term frequency-inverse document frequency (TF-IDF) algorithm that uses search engine for retrieval of information as it does not rely only on surface level characteristics. It also identifies most weighted words, which generate lexical signature (Xiang et al., 2011).
- Phish tester-based approach (Shreeram et al., 2010) provides a finite sate machine model which is based on known legitimate and phishing websites behaviours in order to differentiate legitimate and phishing websites in terms of forms submission that is based on random inputs. A set of heuristics combinations was developed to

capture the recent up-to-date behaviour of suspicious websites. The technique can assist with information and warning about the likelihood of a phishing sites present in a trusted sites and can detect a cross-site scripting (XSS)-based attack.

- GoldPhish (Dunlop et al., 2010) is aimed at identifying and detecting new faces of phishing attacks based on the observation that phishing sites exist only for a few days or hours. GoldPhish captures the image of the website to convert to text data which will serve as input and finally used to retrieve search result to determine if a website is legitimate.
- Garera et al. (2007) adopted the composition and structure of URLs phishing website identification. They combined numerous heuristics and Google PageRank to decide the status (legitimate or phishing) of a URL. The motive behind this is that any established website will be high-ranked while fake websites will be low-ranked. The results show 1.2% false positive rate (FPR), and a 4.2% false negative rate (FNR) (Sheng et al., 2010).
- Machine learning methods such as neural networks have been used to detect fake websites with an accuracy of 86% (Aksu et al., 2018), and 98.7% using recurrent neural networks (RNN) (Bahnsen et al., 2017), and 90% accuracy in detecting phishing websites using support vector machine (SVM) classifier (Jain and Gupta, 2018), a 95.80% recognition rate using SVM (Zouina and Outtaj, 2017), principal component analysis random forest (PCA-RF) achieved an accuracy of 99.55% (Rao and Pais, 2018), extreme learning machine (ELM)-based classification achieved the accuracy of 95.34% (Sönmez et al., 2018).
- Rule-based approaches use a list of rules derived, e.g., using association rule mining, which are interpreted to emphasise the features that are more prevalent in phishing URLs, such as the existence of special characters in the URL, the URL length is more than 75 and having more than four dots in the host name, and achieving a true recognition rate of 93% (Jeeva and Rajsingh, 2016).
- Natural language processing (NLP) techniques have been applied to URL analysis to extract syntactic features such as word count, average (longest, shortest) word length, and number of special characters. Classification using NLP features random forest (RF) classifier achieved 97.2% accuracy (Buber et al., 2018), RF with only NLP-based features gives the 97.98% accuracy rate for detection of phishing URLs (Sahingoz et al., 2019).
- Blacklist-based technique keeps detailed and up-to-date information about all the phishing websites. Anytime there is information regarding a phishing website, the internet protocol (IP) address where the suspected phishing website is initiated will immediately be included among the blacklisted IP addresses. The system, which has client side proxy as browser, can authenticate sites by confirming its status without any warning or notice to user (Alnajim and Munro, 2009). Anytime, a user enters a URL into the browser, its status will automatically be determined. If the site is found to be a blacklisted one, it will be blocked immediately.
- Heuristic-based approach uses web content and URL signatures to detect phishing behaviours through the extraction of different features like visual similarity. Heuristic can produce high true positive and true negative rates and has the ability to

detect the moment an attack is launched. For example, an approach that uses URL tokens as discriminating features achieved 77% of accuracy (Daeef et al., 2017). The disadvantage of heuristic approach is that it can label legitimate websites as phishing websites by producing false positives.

- Visual similarity-based approaches focus on the resemblance of layout as well as the general style between likely phishing and legitimate sites is usually calculated and determined by visual similarity (Liu et al., 2006).
- Multi-stage models combine several methods are combined for detecting malicious websites such as SSL/TLS features and JavaScript-based visual clues (Mensah et al., 2015), domain name of URL and the text extracted from screenshots by optical character recognition (OCR) (Wu et al., 2016). PhishBox (Li and Wang, 2018) uses an ensemble model to validate the phishing data, and then the suspect sites are submitted for crowdvoting for final decision. PhishLimiter (Chin et al., 2018) performs deep packet inspection (DPI) and combines it with software-defined networking (SDN) to recognise suspect phishing actions in e-mail and web-based communication. Li et al. (2019) constructed a stacking model by combining gradient boosting decision trees (GBDT), distributed gradient boosting library XGBoost and gradient boosting framework LightGBM to recognise phishing website.

In an effort to detect phishing websites, we propose a method for detection of phishing attacks by check literal and conceptual inconsistencies in website addresses (URLs). Literal checking of a URL means using the contents that the URL itself, without adding or bringing in other properties and features for test running and verification. This is achieved by following the exact words in URL. The conceptual consistency means that each letter, number and symbol of a URL is divided into various categories to allow for a thorough analysis in order to avoid wrong categorisation. To the best of our knowledge, this approach is novel and differs from the methods proposed by other authors. Our contribution to the field of knowledge is the inclusion of the semantic consistency checking step, which verifies if the semantics of URL and website content match. A similar ide was used in Zhu and Dumitras (2018) to detect malware campaigns.

2 Method

2.1 Pre-filtering stage

This stage is the first stage by which possible potential identities and associated domain names will be examined. This stage assists to separate legitimate site from a phishing or suspicious sites and move the latter to the classification stage. Nearly all legitimate sites have their names registered as their second level domain (SLD) names. To effectively carry out what this stage is meant for, two things must be put into play: the identity extraction and checking if the consistency matches with the SLD.

2.2 Identity extraction

Most phishing websites are commonly constructed to confuse their viewer into believing that the URLs are a legitimate one. To identify the possible signs, the first step is to look

only at the URL of the page to determine features, e.g., if the URL is not structured appropriately, or a number is used to represent a domain name instead of the usual alphabet. Phishing websites use a URL that looks like a legitimate one, alternatively by putting a legitimate URL among the illegitimate/phishing with the motive of deceiving the internet user. After this has been checked and verified and subsequently clear of any suspicion, it can also be checked if it matches a whitelist of a high profile or any other safe site hence the identity extraction is passed (Chen and Guo, 2006).

2.3 Check if consistency matches with SLD

Consistency implies that the degree of compatibility of the URL matches with the brand names of the company. The content of the site should be in compatibility with the domain name that appears in the URL of the webpage. To achieve that, some features have to be put into consideration. The page content checks the HTML of the webpage for likely suspicious features and identifies if a site is used for phishing. The company's logo, images and keywords are also factors to be checked for, but observation shows that most phishing sites copy the company's logo, images and keywords word for word, however this method alone is not considered. Phishing websites are short-lived in nature so they rarely get indexed by search engines (Chou et al., 2004). Whenever a click is made by a user, it is usually stored in as search log. The URL that has been frequently and consistently clicked will be classified as legitimate (Basnet et al., 2008).

2.4 Features

The features used to discover if a site is phishing or legitimate include: Randomness of URL (RU), ratio of found domain token (RDT), position of domain token (CPos), and conceptual similarity (CSim).

2.4.1 Randomness of URL (RU)

The URL contents of a malicious website are usually not related to the website and also contains irrelevant long random strings. Take for example, a text string representing a malicious website link:

- http://signin.ebay.com.87ab3540af65fa59167f076ea075f9f7.ustsecurity.info/

The long random string ('87ab3540af65fa59167f076ea075f9f7') in this website URL address is irrelevant because it does not have any role to play in the website being visited. This might be due to the fact that phishers rapidly generate many malicious sites. There are some features to look out for when considering the randomness of URL such as IP address, symbols, and URL length. Presently, research has not convincingly shown us that there is a standard length of a URL to differentiate a legitimate and phishing sites. However, Liu et al. (2006) is of the opinion that any URL with length greater than 54 characters should be considered phishing.

In calculating the randomness of a URL (RU), we adopted the following formula as formulated and used by (Liu et al., 2006):

$$R(t) = \max\left(sl_d(t), sl_s(t) \cdot \log_2 as(t)\right) \tag{1}$$

here $R(t)$ is randomness of token t, $sld(t)$ is the total number of segments formed by splitting token t into digits, $sls(t)$ is the total number of segments formed by splitting token t into symbols, $as(t)$ is the total number of letters in token, and t is the maximal randomness score of the tokens will be the RU of the URL.

2.4.2 Position of domain token (CPos)

Most phishing sites have lots of sub-domain names than legitimate sites. Research has revealed that the highest number of sub-domain name of any site considered legitimate is 5 while 18 is for phishing site. What is more, there is a standard rule for a given hierarchy of the domain name. The higher the domain tokens, the higher the level of categorisation of the webpage. The position of domain token (CPos) is determined as shown in Algorithm 1.

Algorithm 1 CPos calculation algorithm

```
BEGIN
1: token_list = extract_domainName_token(url)
2: if domain is IP then
3:   score = 0
4: else
5:   for each token in token list do do
6:     allIDX_sum + = index of token
7:     if token is included in content then
8:       foundIDX_sum + = index of token
9:     end if
10:  end for
11:  score = foundIDX_sum = (allIDX_sum * number of count of
     token)
12: end if
13: return score.
END
```

2.4.3 Ratio of found domain token

Most phishing sites are accessible for just a couple of hours or days hence the URL of phishing sites cannot have a reasonable number of clicks on the internet because of its short existence. URLs with frequent and uncountable number of clicks can be regarded as legitimate. The concept entities linked to the entities are other domain tokens obtained from similar data source and detailed in the list of domain token (Zhuang et al., 2012). Legitimate sites should have a greater ratio of tokens at the SLD that that of any malicious websites (Liu et al., 2006). The formula for calculating the RDT is given as follows (Chen et al., 2014):

$$RDT = \sum_{i=1}^{n-1} \left[f_{DT}(t_i) + f_{SLD}(t_i) \right] \tag{2}$$

here, $f_{DT}(\bullet)$ and $f_{SLD}(\bullet)$ are the membership functions, $f_{DT}(t_i) = 1$, if t_i is found in the domain token (DT) list, $f_{DT}(t_i) = 0$, if t_i is not found in the DT list, $f_{SLD}(t_i) = 1$, t_i is found in the SLD list, $f_{SLD}(t_i) = 0$, t_i is not found in the SLD list, and n is a number of domain tokens.

2.5 Anti-phishing algorithm

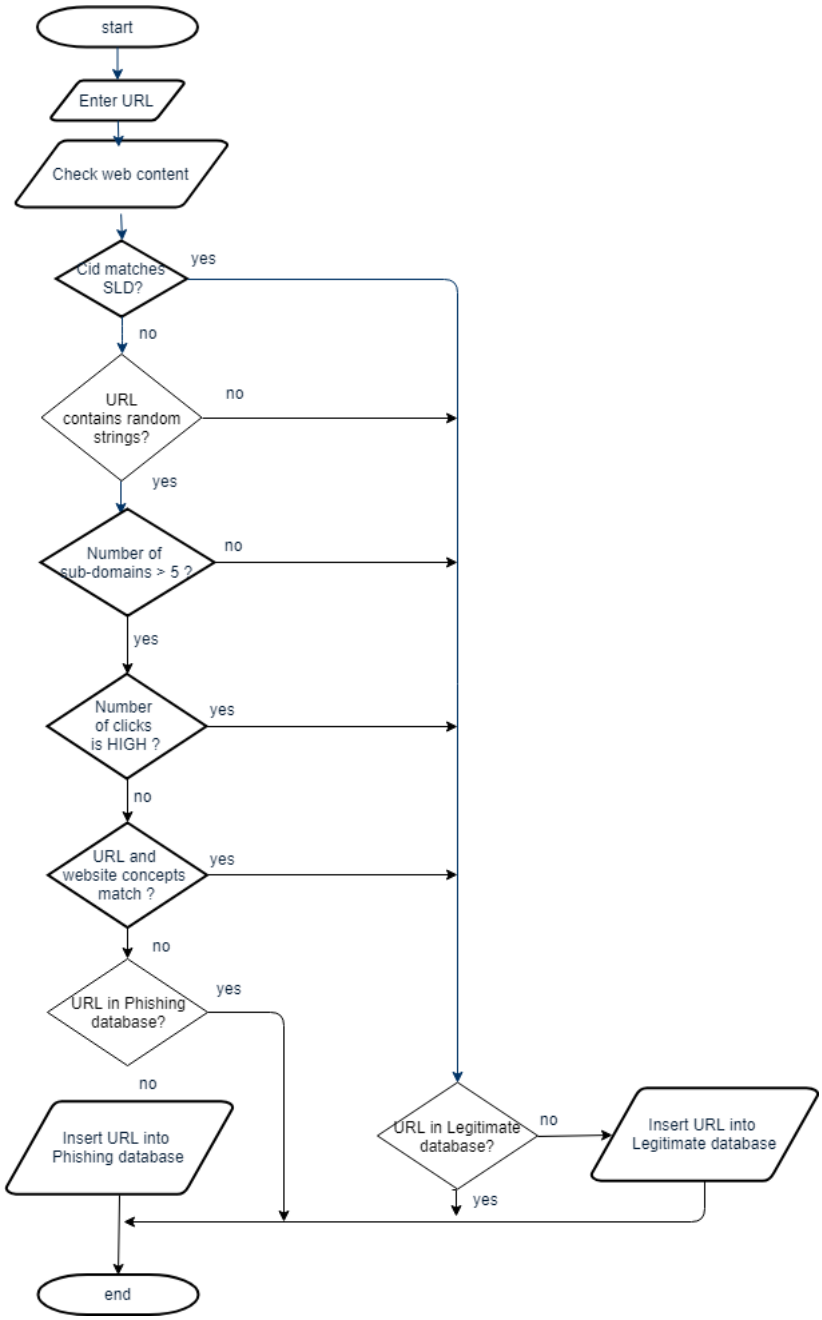
The proposed anti-phishing PhishDetect method is summarised as Algorithm 2.

Algorithm 2 An implementation of anti-phishing approach

```
BEGIN
   1. Enter URL
   2. Check URL against web content
      Check if content contains below
      Check for title
      Check for link
      Check for styles
      Check if URL has http request: (http, https, :, //, :, [a-
      z, 0-9]
   3. Check if Cid matches with SLD
      If passed, check if URL exists in LEGITIMATE database, else
      INSERT into LEGITIMATE database
      If not passed, proceed to (4)
   4. Check if URL contains long (>54) random irrelevant strings
      If passed, proceed to (5)
   5. Check if numbers of sub-domain is more than 5
      If passed, proceed to (6)
   6. Check if URL has a reasonable amount of click records
      If not passed, proceed to (7)
   7. Check if concept of URL and webpages are similar
      If passed, check if URL exists in LEGITIMATE database, else
      INSERT into LEGITIMATE DATABASE.
      If not passed, check if URL exists in PHISHING DATABASE,
      else INSERT into PHISHING DATABASE.
END
```

The method is summarised as a flow diagram in Figure 1.

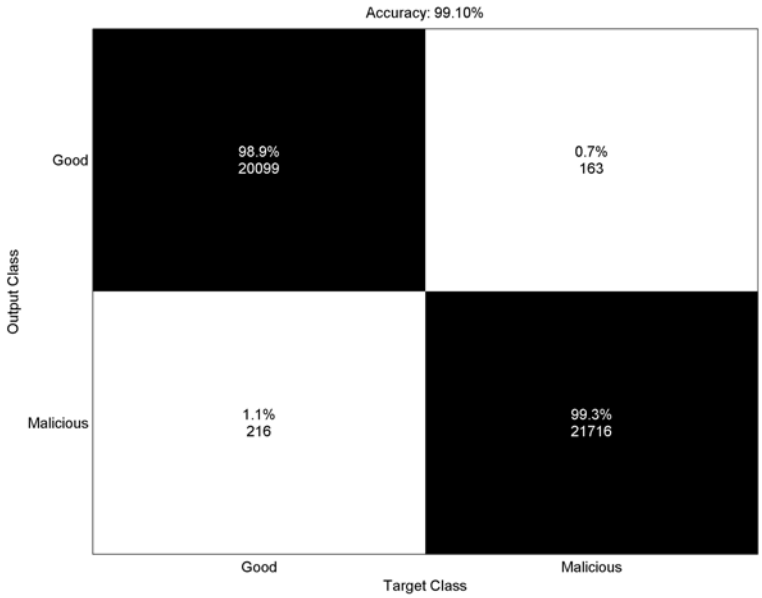
Figure 1 Flow diagram of the proposed approach (see online version for colours)



3 Implementation and evaluation

The URL is entered into the query box of the application, the application checks if the URL exists online by checking the PhishTank database (OpenDNS, 2018). If it exists in the PhishTank, the URL is considered a phishing one else it goes through all the stages (pre-filtering stage and classification stage) where the Randomness of the URL, RDT and the position of the domain token, thereafter the Conceptual similarities before the URL can be confirmed whether it is legitimate or phishing. The presence of the URL on the PhishTank database indicates it is a phishing site. When a URL is entered into the PhishDetect application, the application runs through the submitted URL, if the URL does not have the features of a legitimate site, the application classifies it as a phishing site. Once the URL is submitted, the application uses all the features embedded inside it to check if it's malicious or legitimate. Once it is confirmed as malicious, the application displays all the features checked to tell the user it is a pure phishing website. If the submitted URL does not have any of the attributes of phishing and legitimate or the URL is incomplete, or does not exist online they are classified by the application as invalid and stored in the invalid database.

Figure 2 Confusion matrix of the classification results



To evaluate the system, we used accuracy, which is the measure of overall rate of classified sites in relation to the sum of the actual or correctly classified legitimate sites and phishing sites:

$$Acc = \frac{Ph + O}{Ph + T_{pos} + O + T_{neg}} \tag{3}$$

here T_{pos} is the number of phishing sites accurately categorised as phishing sites, and T_{neg} is the number of legitimate sites accurately categorised as legitimate sites, Ph is the number of phishing sites, and O is the number of legitimate sites.

During the tests, we used our own dataset collected. To collect legal URLs we employed an approach described by Buber et al. (2018) and performed sending the query words to Yandex Search API, while highly ranked URLs returned by the search engine were regarded as legitimate URLs. The malicious URLs were taken from the blacklist of malicious URLs of cloud-based Human Resource Management system at Covenant University and and checked with Sucuri SiteCheck, a free malicious website scanner.

In the collected data set, there are 42,194 URLs including 21,932 malicious URLs and 20,262 legal URLs. Tests were performed on a Hewlett Packard device with 4 GB of 2,300 MHz DDR3L RAM and 2.2 GHz Intel Core i5 processor. The application was able to detect correctly phishing and legitimate URLs, yielding an accuracy of 99.1% from a dataset of websites submitted for evaluation.

4 Conclusions

To counter phishing attacks in the cyber world we proposed a method for detecting phishing sites with the aim of guiding, guarding and preventing internet users from falling prey of cybercriminals. In this solution, two main stages are involved: the pre-filtering stage and the classification stage. For the pre-filtering stage, a special attention is given to the consistency between the potential identities and the SLD names. The classification stage uses the features of vis-à-vis ratio of the found domain, position of the domain token and the randomness of the URL; all these have their attention focused on what is contained in the URL. Finally, the content of the websites and the content between the URL are examined by the conceptual similarity. If the URL of a webpage as gone through all these features and is being considered phishing, the method further checks the PhishTank database to verify if the website is present then the site is considered as a phishing site. The proposed approach contributes to the development of digital forensics method in the communication networks domain.

References

Abbasi, A. and Chen, H. (2009) 'A comparison of tools for detecting fake websites', *Computer*, Vol. 42, No. 10, pp.78–86.

Aichhorn, A., Etzlinger, B., Unterweger, A., Mayrhofer, R. and Springer, A. (2018) 'Design, implementation, and evaluation of secure communication for line current differential protection systems over packet switched networks', *International Journal of Critical Infrastructure Protection*, Vol. 23, pp.68–78, doi:10.1016/j.ijcip.2018.06.005.

Aksu, D., Turgut, Z., Üstebay, S. and Aydin, M.A. (2018) 'Phishing analysis of websites using classification techniques', *Lecture Notes in Electrical Engineering*, pp.251–258, Springer Singapore, Doi:10.1007/978-981-13-0408-8_21.

Aleroud, A. and Zhou, L. (2017) 'Phishing environments, techniques, and countermeasures: a survey', *Computers & Security*, Vol. 68, pp.160–196, doi:10.1016/j.cose.2017.04.006.

Alnajim, A. and Munro, M. (2009) 'An approach to the implementation of the antiphishing tool for phishing websites detection', *International Conference on Intelligent Networking and Collaborative Systems, INCOS '09*, pp.105–112.

Bahnsen, A.C., Bohorquez, E.C., Villegas, S., Vargas, J. and Gonzalez, F.A. (2017) 'Classifying phishing URLs using recurrent neural networks', *ECrime Researchers Summit, eCrime*, pp.1–8, doi:10.1109/ECRIME.2017.7945048.

Basnet, R., Mukkamala, S. and Sung, A. (2008) 'Detection of phishing attacks: a machine learning approach', *Soft Computing Applications in Industry*, pp.373–383.

Buber, E., Diri, B. and Sahingoz, O.K. (2018) 'NLP based phishing attack detection from URLs', in Abraham, A., Muhuri, P., Muda, A. and Gandhi, N. (Eds.): *Intelligent Systems Design and Applications. ISDA 2017. Advances in Intelligent Systems and Computing*, Vol. 736, pp.608–618, Springer, Cham., doi:10.1007/978-3-319-76348-4_59.

Chen, J. and Guo, C. (2006) 'Online detection and prevention of phishing attacks', *First International Conference on Communications and Networking in China (ChinaCom'06)*, pp.1–7.

Chen, Y., Liu, H., Yu, Y. and Wang, P. (2014) 'Detect phishing by checking content consistency', *IEEE 15th International Conference on Information Reuse and Integration (IRI), IEEE IRI 2014*, 13–15 August, San Francisco, California, USA, pp.109–116.

Chiew, K.L., Yong, K.S.C. and Tan, C.L. (2018) 'A survey of phishing attacks: their types, vectors and technical approaches', *Expert Systems with Applications*, Vol. 106, pp.1–20, doi:10.1016/j.eswa.2018.03.050.

Chin, T., Xiong, K. and Hu, C. (2018) 'Phishlimiter: a phishing detection and mitigation approach using software-defined networking', *IEEE Access*, Vol. 6, pp.42513–42531, doi:10.1109/ACCESS.2018.2837889.

Chou, N., Ledesma, R., Teraguchi, Y., Boneh, D. and Mitchell, J.C. (2004) 'Clientside defense against web-based identity theft', *11th Annual Network and Distributed System Security Symposium (NDSS'04)*, pp.1–16.

Daeef, A.Y., Ahmad, R.B. and Yacob, Y. (2017) 'Websites phishing detection using URLs tokens as a discriminating features', *Journal of Engineering and Applied Sciences*, Vol. 12, No. 3, pp.513–519, doi:10.3923/jeasci.2017.513.519.

Dunlop, M., Groat, S. and Shelly, D. (2010) 'GoldPhish: Using images for content based phishing analysis', *5th International Conference on Internet Monitoring and Protection (ICIMP)*, pp.123–128.

Fan, X., Wei, W., Wozniak, M. and Li, Y. (2018) 'Low energy consumption and data redundancy approach of wireless sensor networks with bigdata', *Information Technology and Control*, Vol. 47, No. 3, pp.406–418, doi:10.5755/j01.itc.47.3.20565.

Garera, S., Provos, N., Chew, M. and Rubin, A.D. (2007) 'A framework for detection and measurement of phishing attacks', *2007 ACM Workshop on Recurring Malcode*, pp.1–8.

Goel, D. and Jain, A.K. (2018) 'Mobile phishing attacks and defence mechanisms: state of art and open research challenges', *Computers & Security*, Vol. 73, pp.519–544, doi:10.1016/j.cose.2017.12.006.

Grover, J. and Sharma, S. (2016) 'Security issues in wireless sensor network — a review', *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp.397–404, IEEE, doi:10.1109/icrito.2016.7784988.

Gupta, B.B., Arachchilage, N.A.G. and Psannis, K.E. (2017) 'Defending against phishing attacks: taxonomy of methods, current issues and future directions', *Telecommunication Systems*, Vol. 67, No. 2, pp.247–267, doi:10.1007/s11235-017-0334-z.

Jain, A.K. and Gupta, B.B. (2018) 'A machine learning based approach for phishing detection using hyperlinks information', *Journal of Ambient Intelligence and Humanized Computing*, pp.1–14, doi:10.1007/s12652-018-0798-z.

Jeeva, S.C. and Rajsingh, E.B. (2016) 'Intelligent phishing url detection using association rule mining', *Human-Centric Computing and Information Sciences*, Vol. 6, No. 1, doi:10.1186/s13673-016-0064-3.

Comment [t3]: Author: Please provide the volume number and issue number.

Khonji, M., Iraqi, Y. and Jones, A. (2013) 'Phishing detection: a literature survey', *IEEE Communications Surveys & Tutorials*, Vol. 15, No. 4, pp.2091–2121, doi:10.1109/surv.2013.032213.00009.

Lauraitis, A., Maskeliunas, R., Damasevicius, R., Polap, D. and Wozniak, M. (2019) 'A smartphone application for automated decision support in cognitive task based evaluation of central nervous system motor disorders', *IEEE Journal of Biomedical and Health Informatics*, doi:10.1109/jbhi.2019.2891729.

Li, J. and Wang, S. (2018) 'PhishBox: an approach for phishing validation and detection', *IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, IEEE 15th International Conference on Pervasive Intelligence and Computing, IEEE 3rd International Conference on Big Data Intelligence and Computing and IEEE Cyber Science and Technology Congress, DASC-PICom-DataCom-CyberSciTec 2017*, pp.557–564, doi:10.1109/DASC-PICom-DataCom-CyberSciTec.2017.101.

Li, Y., Yang, Z., Chen, X., Yuan, H. and Liu, W. (2019) 'A stacking model using URL and HTML features for phishing webpage detection', *Future Generation Computer Systems*, Vol. 94, pp.27–39, doi:10.1016/j.future.2018.11.004.

Liu, W., Deng, X., Huang, G. and Fu, A.Y. (2006) 'An antiphishing strategy based on visual similarity assessment', *Internet Computing*, Vol. 10, No. 2, pp.58–65, IEEE.

Mensah, P., Blanc, G., Okada, K., Miyamoto, D. and Kadobayashi, Y. (2015) 'AJNA: anti-phishing JS-based visual analysis to mitigate users' excessive trust in SSL/TLS', *4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, pp.74–84.

Odun-Ayo, I., Misra, S., Omoregbe, N.A., Onibere, E., Bulama, Y. and Damasevicius, R. (2017) 'Cloud-based security driven human resource management system', *Frontiers in Artificial Intelligence and Applications*, p.295, *Advances in Digital Technologies*, pp.96–106, doi:10.3233/978-1-61499-773-3-96.

Odusami, M., Abayomi-Alli, O., Misra, S., Shobayo, O., Damasevicius, R. and Maskeliunas, R. (2018) 'Android malware detection: a survey', *Communications in Computer and Information Science*, pp.255–266, doi:10.1007/978-3-030-01535-0_19.

OpenDNS (2018) *PhishTank* [online] https://www.phishtank.com.

Pham, C., Nguyen, L.A.T., Tran, N.H., Huh, E. and Hong, C.S. (2018) 'Phishing-aware: a neuro-fuzzy approach for anti-phishing on fog networks', *IEEE Transactions on Network and Service Management*, Vol. 15, No. 3, pp.1076–1089, doi:10.1109/TNSM.2018.2831197.

Połap, D., Kęsik, K., Książek, K. and Woźniak, M. (2017) 'Obstacle detection as a safety alert in augmented reality models by the use of deep learning techniques', *Sensors*, Vol. 17, No. 12, p.2803, doi:10.3390/s17122803.

Prakash, P., Kumar, M., Kompella, R. and Gupta, M. (2010) 'Phishnet: predictive blacklisting to detect phishing attacks', *IEEE, in INFOCOM, 2010 Proceedings IEEE*, pp.346–350.

Rao, R.S. and Pais, A.R. (2018) 'Detection of phishing websites using an efficient feature-based machine learning framework', *Neural Computing and Applications*, pp.1–23, doi:10.1007/s00521-017-3305-0.

RSA (2018) *RSA Quarterly Fraud Report*, Q3 2018 [online] https://www.rsa.com/en-us/offers/rsa-fraud-report-q3-2018.

Sahingoz, O.K., Buber, E., Demir, O. and Diri, B. (2019) 'Machine learning based phishing detection from URLs', *Expert Systems with Applications*, Vol. 117, pp.345–357, doi:10.1016/j.eswa.2018.09.029.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F. and Downs, J. (2010) 'Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions', *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI '10*, ACM, New York, NY, USA, pp.373–382.

Comment [t4]: Author: Please provide the access details (date when the site was accessed/visited).

Comment [t5]: Author: Please provide the access details (date when the site was accessed/visited).

Shreeram, V., Suban, M., Shanthi, P. and Manjula, K. (2010) 'Anti-phishing detection of phishing attacks using genetic algorithm', *IEEE International Conference on Communication Control and Computing Technologies (ICCCCT)*, pp.447–450.

Sönmez, Y., Tuncer, T., Gökal, H. and Avci, E. (2018) 'Phishing web sites features classification based on extreme learning machine', *6th International Symposium on Digital Forensic and Security, ISDFS 2018 – Proceeding*, pp.1–5, doi:10.1109/ISDFS.2018.8355342.

Woźniak, M., Połap, D., Damaševičius, R. and Wei, W. (2018) 'Design of computational intelligence-based language interface for human-machine secure interaction', *Journal of Universal Computer Science*, Vol. 24, No. 4, pp.537–553.

Wu, L., Du, X. and Wu, J. (2016) 'Effective defense schemes for phishing attacks on mobile computing platforms', *IEEE Transactions on Vehicular Technology*, Vol. 65, No. 8, pp.6678–6691.

Xiang, G. and Hong, J.I. (2009) 'A hybrid phish detection approach by identity discovery and keywords retrieval', *International World Wide Web Conference Committee (IW3C2)*, pp.1–10.

Xiang, G., Hong, J., Rose, C.P. and Cranor, L. (2011) 'CANTINA+: a feature-rich machine learning framework for detecting phishing web sites', *ACM Transactions on Information and System Security*, Vol. 14, No. 2, Article No. 21.

Yaseen, M., Saleem, K., Orgun, M.A., Derhab, A., Abbas, H., Al-Muhtadi, J. and Rashid, I. (2018) 'Secure sensors data acquisition and communication protection in eHealthcare: review on the state of the art', *Telematics and Informatics*, Vol. 35, No. 4, pp.702–726, doi:10.1016/j.tele.2017.08.005.

Zhu, Z. and Dumitras, T. (2018) 'ChainSmith: automatically learning the semantics of malicious campaigns by mining threat intelligence reports', *IEEE European Symposium on Security and Privacy (EuroS&P)*, London, pp.458–472.

Zhuang, W., Jiang, Q. and Xiong, T. (2012) 'An intelligent anti-phishing strategy model for phishing website detection', *32nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*, Macau, China, pp.51–56.

Zouina, M. and Outtaj, B. (2017) 'A novel lightweight URL phishing detection system using SVM and similarity index', *Human-Centric Computing and Information Sciences*, Vol. 7, No. 1, doi:10.1186/s13673-017-0098-1.

Comment [t6]: Author: Please provide the page numbers.