Phishing email detection using Natural Language Processing techniques : a literature survey

Salloum, S, Gaber, TMA, Vadera, S and Shaalan, K

http://dx.doi.org/10.1016/j.procs.2021.05.077

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 189 (2021) 19–28

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on AI in Computational Linguistics

Phishing Email Detection Using Natural Language Processing Techniques: A Literature Survey

Said Salloum[a*], Tarek Gaber[a,b], Sunil Vadera[a], and Khaled Shaalan[c]

[a]*School of Science, Engineering, and Environment, University of Salford, UK*
[b]*Faculty of Computers and Informatics, Suez Canal University, Ismailia 41522, Egypt*
[c]*Faculty of Engineering &IT, The British University in Dubai, Dubai, UAE*

Abstract

Phishing is the most prevalent method of cybercrime that convinces people to provide sensitive information; for instance, account IDs, passwords, and bank details. Emails, instant messages, and phone calls are widely used to launch such cyber-attacks. Despite constant updating of the methods of avoiding such cyber-attacks, the ultimate outcome is currently inadequate. On the other hand, phishing emails have increased exponentially in recent years, which suggests a need for more effective and advanced methods to counter them. Numerous methods have been established to filter phishing emails, but the problem still needs a complete solution. To the best of our knowledge, this is the first survey that focuses on using Natural Language Processing (NLP) and Machine Learning (ML) techniques to detect phishing emails. This study provides an analysis of the numerous state-of-the-art NLP strategies currently in use to identify phishing emails at various stages of the attack, with an emphasis on ML strategies. These approaches are subjected to a comparative assessment and analysis. This gives a sense of the problem, its immediate solution space, and the expected future research directions.

Keywords: Phishing email; Natural Language Processing; Machine Learning.

1. Introduction

The immense growth of internet technologies has dramatically altered online user interaction while creating more severe security problems. Newly emerging threats of the present world not only target the user's computer but could potentially steal their identity and money. Phishing threats make use of not only technology but also social engineering to pilfer data related to the victim's identity and accounts; hence, it is imperative to curtail the threat and criminal activity associated with phishing. The Anti-Phishing Working Group (APWG)'s report released in Q3 of 2020 stated that "The total number of phishing email detected by the APWG in Q3 of 2020 was 128,926, that was up notably from the 44,497 seen in Q2 of 2020, and from the 44,008 seen in Q1 of 2020" [1]. Phishing attacks with the subject of coronavirus disease of 2019 (covid-19) have been deployed since mid-September of the following year. Phishing attacks mostly have the textual composition of subjects such as internet and security technologies, as

10.1016/j.procs.2021.05.077

well as information regarding the covid-19 pandemic to attract their targets [1]. Phishing has increased exponentially according to the data, so has the harm caused by it.

Phishing extracts sensitive information from unwary victims and is a social engineering threat. The most used communication channels for such attacks are emails, instant messaging, etc. The attackers appear as legitimate and credible individuals. As email is the most common channel for these attackers, we focus on email communication in this study [2], [3]. Detecting phishing emails and messages automatically is difficult work, as observed [4]. The literature discusses several methods for detecting phishing emails. These predominantly involve three techniques: blacklisting, ML-based classification algorithms, and deep learning (DL) [5]. Current blacklist mechanisms mainly depend upon people identifying and reporting phishing email links, which involves a large workforce and long hours. Moreover, while ML classification algorithm-oriented phishing detection techniques make use of artificial intelligence (AI), feature engineering is essential for the manual identification of representative features, which is not feasible in the case of the application of data migration scenarios. Furthermore, the detection technique developed in DL is restricted by the word embedding in the email's content representation. The output is not optimal, as such methods fall short in identifying the relevance of the phishing emails and specifically interchanged the NLP and DL technology [5]–[7]. Following exponential growth in DL and its highly accurate applications, it has gained popularity in pinpointing phishing [8]. In comparison with traditional ML techniques, DL collects handcrafted aspects inherently so that the ML experts can insert the data without even having to acquire knowledge of cyber-security.

Given the vast range of approaches, this survey provides an organized guide to the current state of the literature. The assessment and analysis of various approaches to phishing email identification were given a lot of importance in the literature. This survey compares and analyses the relative merits of these approaches in addition to identifying and categorizing them. It lists capabilities, limitations, and associated implementation situations, for example, to assist readers in developing new anti-phishing detection approaches in the future. This article is not meant to discuss similar subjects such as spam, which has also been the subject of several articles. Phishing by email is a separate issue that requires further focus. Because of the wide scope of the phishing challenge, this phishing email identification survey starts with:

- Defining the phishing problem. It is necessary to remember that the literature's description of phishing is inconsistent, so a comparison of some meanings is presented, which includes phishing history.
- A literature review of anti-phishing detection strategies is presented, which includes features for detecting phishing emails, software detection techniques as well as user-awareness techniques that improve phishing attack detection.
- A comparison of the different proposed phishing email detection that uses NLP techniques.

1.1. Definition of phishing

Phishing is a word that has thousands of references in science journals, a lot of newspaper coverage, and a lot of scrutiny from organizations like banks and law enforcement agencies. This, though, raises the question of what exactly phishing is. The phenomenon of phishing is specifically specified in some publications; in others, it is explained by an illustration, while others presume that the reader already knows what phishing is. Many scholars have proposed their own definition of phishing, resulting in a plethora of meanings in the scientific literature. The literature does not provide a clear description of phishing attacks, which is due to the fact that the phishing issue is broad and encompasses a variety of scenarios. According to PhishTank, for instance: "Phishing is a fraudulent attempt, usually made through email, to steal your personal information" [16]. The definition of PhishTank remains true in a variety of situations that approximately encompass the bulk of phishing attacks (although no accurate studies have been made to reliably quantify this). Nevertheless, the term confines phishing attacks to the theft of personal data, which is not always the case. The majority of situations in which phishers attempt to steal confidential personal details such as login credentials are covered by APWG [17], Xiang et al. [18], and Ramesh et al. [19]. [20] Another description is given: "We define a phishing page as any web page that, without permission, alleges to act on behalf of a third party to confuse viewers into performing an action with which the viewer would only trust a true

agent of the third party". [20] strives to be wider than PhishTank's meaning so that attackers' objectives are no longer limited to stealing sensitive information from victims. The concept, on the other hand, also limits phishing attacks to those who operate on behalf of third parties, which isn't always the case. Social engineering (that is, via fraudulent emails) and technological deception are the most common phishing techniques (e.g., malware infection). Users' personal information is also harvested from the Internet using sophisticated techniques (e.g., pharming [21]). The meanings of [20], [22], but on the other hand, do not restrict the attacker's target (e.g., sensitive personal information). They explain the phishing technique (e.g., phishing website or socially programmed messages) without mentioning the phishing target (e.g., just mentioning the attackers' benefit). To summarize, [25] is the most general of the definitions examined, while APWG [17] describes the most widely used phishing attacks in a particular way.

1.2. Phishing history

As per the APWG, the word phishing was introduced in 1996 as a result of social engineering attacks by web scammers against America On-line (AOL) accounts [23]–[25]. Fishers (i.e. attackers) use traps (i.e. socially-engineered messages) to catch fish (e.g. steal personal information of victims) [22]. The origins of the *ph* substitution of the character f in fishing can be traced back to one of the early ways of hacking, known as Phone Phreaking, which targeted telecommunications networks [22]. As a consequence, *ph* has become a popular hacking character to replace *f*. According to the APWG, by 1997, hackers were using stolen accounts as a form of money to swap hacking codes in exchange for the stolen accounts. Phishing attacks began with the theft of AOL accounts and evolved to include more lucrative targets such as online banking and e-commerce services [22]. Phishing attacks now threaten not just system end-users but also technical staff at service providers, and they use advanced tactics like Man-in-the-Browser (MitB) attacks.

2. Literature on phishing email detection

Several reviews have been published in recent years, each providing critical knowledge for researchers to better understand the phishing epidemic and how it is used with various approaches. The viewpoint of feature selection approaches in hybrid phishing detection was investigated by Zuhair et al. [9]. Also, Varshney, et al. [10] investigated, reviewed, and classified the most important and novel methods proposed in the field of phished website detection, as well as highlighting their benefits and drawbacks. Moreover, Gupta et al. [11] published a literature review on phishing attacks to examine phishing attacks in-depth, the past of phishing attacks, and the motives of the perpetrator behind the attack. They also have a taxonomy of different forms of phishing attacks as well as a taxonomy of different methods suggested in the literature to detect and protect against phishing attacks. Zuraiq & Alkasassbeh [12] conducted a detailed survey of current methods for detecting phishing attacks. Furthermore, Thakur et al. [13] published a new study on phishing identification to look at the backgrounds and motivations of hackers who perpetrated these attacks. Often, categorise the different kinds of phishing attacks. Alabdan [14] also performed a phishing attack survey to provide a critical overview of the key tactics, supplemented by a thorough review of the characteristics of current classic, contemporary, and cutting-edge phishing attack techniques. The objectives of this paper are to raise awareness about phishing tactics, inform individuals about these attacks, and promote the use of phishing prevention techniques, as well as to encourage expert debate on the topic. Korkmaz et al. [15] reviewed the literature for research on phishing web pages classification and the features used in these studies.

Based on the existing literature, none of the above review studies have considered the relationship between phishing email detection and NLP techniques. However, in the past years, various studies conducted in this regard provided important information for scholars to understand the effect of phishing email detection techniques to solve the phishing problems. It is important to mention here that the previous researches have overlooked the review of studies pertinent to use NLP techniques. This neglected perspective convinced us to conduct this survey paper. The current review study attempts to add more weightage to the existing body of literature by assessing and including the

latest amalgamation of phishing email detection research studies that were mostly concentrated using NLP techniques.

2.1. Features for detecting phishing emails

Phishing emails are sent with the intent of stealing personal information from the recipients. The majority of users fell victim to phishing attacks as a result of their careless internet surfing. Companies should educate their staff about phishers' traps and strategies. In this section, we'll go into how to defend against phishing attacks as well as how to identify phishing. To weed out phishing emails, some spam filters employ hundreds of features. These features [26] for detecting phishing emails are classified as follows:

1. Email body-based characteristics: These attributes are taken from the email body. They have binary features like shapes, HTML, and specific phrases and links in the email body.
2. Subject-based features: Certain aspects of an email are derived from its subject, such as whether it is a reference to a previous email or the use of terms like verify or debit.
3. URL-based characteristics: These attributes examine when an IP address is used instead of a domain name, the inclusion of @ in links, the number of photos, external and internal links in the email document, the number of cycles in links, and so on.
4. Script-based features: These features look for JavaScript, pop-up window code, on-click activities, and other script-based features in the email.
5. Sender-based characteristics: These characteristics provide information about the sender, such as the difference between the sender's address and the reaction to the address.

2.2. Phishing email detection using NLP

One type of phishing is through spoofing emails, where the phisher emails the user using a fake email address to deceive people so that they end up opening the email [27]–[31]. This allows the phisher to influence the user and gain from their private information [32]. Several anti-phishing technologies have gained traction in countering the problem such as phishing blacklist [33], phishing email detection based on NLP and ML approaches [3], [5], [34]–[38].

The efficiency of the phishing blacklist was investigated by Sheng et al. [33]. The method is based on blacklists of senders and links. Detection involves extracting sender address and link address from the message and cross-checking with the blacklists, for verifying whether the email constitutes a phishing attempt. Blacklists consist of the sender blacklists and link blacklists. The major drawback is the manual revision of the blacklist and users' indication of the website as a phishing website by reporting it. Among all the databases of phishing websites, two major phishing sites are PhishTank [39] and OpenPhish [40]. The efficiency of blacklisting for identifying phishing emails depends greatly on blacklists.

AI has developed a lot, and now phishing email detection has also adopted ML. NLP as well as ML have contributed considerably to combating phishing emails [5]. Earlier, features related to semantics [34], syntax [35], and context [3] had a major role in phishing detection. A study by [36] developed the simplest ML strategies using "Random Forest (RF), Decision Trees (DT), Logistic Regression (LR), and Support Vector Machine (SVM)" containing supervised classification for phishing email identification. One technique using hybrid feature selection simultaneously analyses content as well as behaviour [37]. ML-based anti-phishing techniques train classification algorithms from both phishing and legitimate emails helped by the ML algorithm to attain classifier model email classification. An investigation by [41] divided features into three categories: basic [41], [42], latent topic model [41], [43], [44] and dynamic Markov-chain features [41], [42], [44]. Basic features can be collected through email

and need not be processed further. Topic model features are sets of words connected to each other and can occur together; these are not easily detectable in an email. Text features developed on the basis of bag-of-words are also known as "Dynamic Markov Chain features", which involve the modelling of message content to determine the probable association of an email with either legitimate or spam groups. One major flaw with the NLP phishing e-mail detection developed on ML is its high dependence on emails' surface text instead of deep semantics. So, when the structure of a sentence is altered or different synonyms of words are chosen or if different changes are made, it is nearly impossible for NLP built on ML to analyse these changes [45]. This ML method primarily uses feature engineering to create features to accomplish tasks and present emails. Blacklisting and feature engineering both operate manually and a large workforce and expertise are necessary for this task, which restricts the efficiency of detection. Various NLP tasks consisting of text categorisation [46], information extraction [47], and machine translation [48] have been heavily influenced by DL. It can also create features from emails, which will pinpoint phishing attempts automatically, doing away with the need for the manual extraction feature of emails. In phishing email identification, DL helps process emails' text more accurately and efficiently. A study by [49] used DL along with word embedding techniques for reintroducing structure in free text email conversations. This is not used to detect phishing emails but our mode still comprises DL with word embedding to analyse emails. In [6] suggested a phishing detection model prepared on Keras [50], word embedding and also convolutional neural networks (CNN).

NLP technology is currently being employed for detecting phishing emails, rather than using DL techniques, completely disregarding how anti-phishing email and other objectives differ, and partially ignoring contextual information, thus limiting the progress in phishing email detection. The current literature review reveals that the previously mentioned issues are of considerable importance to grasp the research trend of phishing detection utilising NLP and ML. For the most part, we performed a survey to synthesise the research pertaining to the detection of phishing through NLP and ML in order to comprehensively analyse these research works. Some research has been conducted to identify phishing emails utilising diverse ML approaches. Numerous features have been developed for the categorisation of emails into malicious or safe emails [51]–[56]. The researchers in [51] detected phishing emails adopting the best list of features that has high accuracy, but using the least number of features. Their paper suggested a binary search feature selection (BSFS), which assessed with greater accuracy using the least features as well as search time. The findings revealed that the BSFS technique weighed the accuracy of 97.41% in comparison with SFFS (95.63%) and WFS (95.56%). The study still needs more features and sophisticated feature selection methods set to derive the best feature set. An investigation by [52] used document embedding utilising Doc2Vec and performed a parallel arrangement that employed "SVM, LR, RF, and Naive Bayes (NB)". The results of the conducted tests indicated a good classification rate with accuracy and F1 score of 81.6% and 76.6%, respectively using the SVM classifier. Despite considerable study on detecting phishing emails, the research is lacking the use of features that permit an easier interpretation and provide a deeper insight into phishing and legitimate emails. The model put forward in [53] has been developed based on a dynamic approach that adopted an inbuilt dataset gathered from various web sources to equip the work with a dynamic dimension and improve accuracy. A hybrid approach has been proposed for phishing detection integrating feature extraction and classification of the mails using SVM. While compared the proposed hybrid method along with SVM (accuracy-87%, sensitivity-88.5 % & specificity-91%) & Neural Network (accuracy-90.5%, sensitivity-92%, specificity-93.5%) method, we found our proposed hybrid method (accuracy-98%, sensitivity-97%, and specificity-97.5%) performed better [53]. By increasing the dataset, the predicted method would be strengthened. By using a variety of emails, both phished and ham, the scheme will be closer to the real-world scenario, where fraudsters are constantly improving their methods. A study by [54] used a multi-stage method involving both normal NLP and ML to detect phishing emails; the suggested multi-stage strategy comprises of feature engineering inside NLP, "lemmatisation, feature selection, feature extraction", improved learning strategies for resampling and cross validation, and the arrangement of hyper parameters. In Scholars of [54] has introduced two techniques, with the first employing "chi-square statistics and mutual information" to improve dimensionality, whereas the subsequent strategy uses a combination of principal component analysis (PCA) and latent semantic analysis (LSA). "These approaches produce reduced feature sets that, combined with the XGBoost and RF algorithms, lead to an F1-measure of 100% success rate". Validation experiments were conducted using the SpamAssassin Public Corpus and the Nazario Phishing Corpus datasets". An investigation by [55] expects to order spam messages effectively as well as with low latency.

The research applied distinctive ML models like "XGBoost, LightGBM, and Bernoulli Naive Bayes" that are extremely quick and are also marked by lower time unpredictability. Another feature taken for this purpose is the length of messages; Unigram, Bigram, and TF-IDF matrix helped pull such features. Chi-square feature selection enabled diminishing space complexity. The main findings of this study indicate that "Bernoulli Naive Bayes followed by LightGBM with the TF-IDF matrix produced the highest accuracy of 96.5% in 0.157 seconds and 95.4% in 1.708 seconds respectively". In order to boost performance, the models would be more stable when the study use ML models that can be trained using datasets from various resources, as well as datasets with a huge number of documents. An email spam detection strategy that uses NLP and ML techniques is shown in [56]. They depict the outcomes obtained from classification and assessment. Validation experiments have been conducted using the SpamAssassin Public Corpus. SVM algorithm used for detecting spam emails. The accuracy of the applied technique is calculated to 0.90, the precision is equal to 0.90, the recall is equal to and finally, the F-measure is equal as 0.90. The weakness of this study is that it is used only one algorithm to train the proposed model (SVM) for detecting spam emails.

THEMIS is a new DL model for detecting phishing e-mails [5]. The model runs on a better convolutional neural network (RCNN) model including multilevel vectors as also attention mechanism, which allows concurrent modelling of an email at the header, body, character and word levels. THEMIS has an accuracy of 99.848% according to the outcomes of our study. The only flaw of the model is that it cannot detect phishing in emails with an e-mail body but no email header. A model targeting phishers based on the character level convolutional neural network (CNN) was executed by a group of researchers in [38]. In the proposed model, URL is used to extract features of emails that completely outdate the manually created handcrafted features; the model does not depend on network access, which renders it more reliable for clients owing to low response time. It has an accuracy of 95.02%, but this model still has some downsides. The major drawback is that it does not recognise if the URL of the website is active or has any error; it is really important to examine the URL of the website before any major conclusion. The model sometimes misclassifies phishing sites in case of shorter URLs or URLs containing sensitive words such as "login" or "registered", which may result in the misclassification of URLs as phishing websites. Additionally, some URLs of misleading websites which aren't actually replicas of other websites can go un-scanned by the model depending upon the URL string.

3. Discussion

The entire internet community across the globe is highly concerned about the threat posed by phishing attacks. These attacks affect not only internet users but also organizations and ISPs. Despite the discovery of different techniques for phishing detection, the methods being used currently are not adequate to fully overcome this issue; there is a need for discovering comprehensive solutions for countering it [57]. The task of the development of techniques for detecting and countering phishing is becoming more complicated day by day due to the diverse range of techniques continuously emerging in this domain. The same was observed when the phishing rate hit the 3-year high mark in March 2020 with a constant elevation in the average volume of phishing e-mails in all the following months of the same year [1]. The main difference between the current and previous researches [9], [10], [60], [11]–[15], [22], [58], [59] is that the current one conducts a complete analysis of the recommended methods of phishing email detection using NLP techniques. Furthermore, this work provides an analysis of the numerous (NLP) strategies currently in use to identify phishing emails at various stages of the attack, with an emphasis on ML strategies. These approaches are subjected to a comparative assessment and analysis. This gives a sense of the problem, its immediate solution space, and the expected future research directions. According to our existing knowledge, there exists no systematic mapping for the analysis of existing proposed solutions within the current literature for facilitating phishing email detection through NLP techniques. This work performs a quantitative survey of various features of existing research to get an idea about the work already done in this field and the sites where those contents are published, to introduce this new theme to the new researchers.

The current paper presents a critical analysis of the research about phishing email detection techniques particularly the feature selection techniques. The review also sheds light on the ML classifiers mentioned in the included researches along with their corresponding feature selection technique. The shortcomings associated with each included research have also been presented. The outcomes of the included research works have also been discussed concerning various factors including acceptance of redundant or unrelated features, feature hybridity, value diversity, imbalanced data, high dimensionality in data along with the performance depicted by the features in detection of novel kinds of phishing attacks and spam detection in web data. Moreover, it was found in this review that it is possible to facilitate the selection of the most appropriate feature/features by applying an additional feature. Consequently, the issues mentioned above can be easily prevented. Besides this, only a small number of appropriate redundant features are applied by the additional feature to detect phishing leading to better selection results, lower probability of imbalanced dataset selection besides fair reduction in dataset dimensionality. Phishing detection is deemed as efficient if the classification is precise, the runtime is shorter, the expenses due to errors or false detections are lower, and calculations are simple with a minimum storage. This implies that future research will focus on the recommendation of using an additional feature and the benefits offered by these features during the detection of a hybrid phishing email.

It is imperative to consider the outlook of attackers or phishers regarding the phishing problem when carrying out research in the future. The research is needed to investigate the phishers' emotions as well as their motives. Consequently, it will become possible to control phishers' activities. The same was found in the research paper by Hausken et al. [61] presented in 2018 whereby he stated that various tangible factors including financial and intellectual aspects may encourage the phishers to involve in phishing attacks. It is important to note that message contained in phishing e-mails is usually the subject of most research works performed in the domain of phishing prevention and detection techniques, overlooking the phishing attacks made through techniques like malvertizing, squatting and tab-napping. Nowadays, the research mainly investigates the performance of NLP technique which is found to detect and sort out phishing emails more accurately. The phishing is detected by NLP through the application of semantic changes. According to research, NLP is characterized with higher precision in comparison to other techniques when it comes to identification of phishing emails; however, NLP is not so far tried on larger datasets [53]. Another method likely to combat phishing problem is Neural Networks; but, the features of extensive training requirements and special expertise essential for parameter tuning make it a bad choice for phishing detection. Besides this, experts are also considering the use of ML techniques for identifying phishing attacks. Due to extensive focus on novel techniques and tools associated with phishing detection, the impact of human aspect and education methods on phishing has been overlooked in existing literature; hence, these areas must be considered while performing research in future. Researchers, till today, have focused upon non-semantic feature analysis that is not relevant to reading the sender's intention. The task is further complicated by targeted phishing emails being forged by the phishers, like spear-phishing or whaling with the use of personal information obtained from the Social Networking Sites (SNS). Such an activity would not usually include attachments or links, and this creates issues for the detection systems that depend on listings or malware analysis for detecting these emails. When initiating intention analysis, it is necessary to extract the meaning of the body content of the email. The goal of detection is achieved through the introduction of semantic within the research as it helps the email content to be semantically processed. To manage the issues related to phishing, a rigorous study needs to be carried out to extract what the email body text conveys and means. A semantic analysis must be made of the email body text and the legitimacy of the email should be determined using features such as meanings of words and sentences.

ML techniques particularly clustering and classification methods are usually implemented in most techniques associated with the detection of phishing e-mails. Accordingly, such techniques are based on the use of ML-based approaches as well as ML-based evaluation metrics. But, the continuously evolving character of phishers can lead to obsolescence of the evaluation outcomes in the long run. This has amplified the complexity of the cybersecurity domain. However, this problem can be overcome by paying special attention while developing evaluation outcomes. To avoid the drawbacks like inadequate time validity and scope, the experts associated with phishing detection must also make efforts and come up with more flexible evaluation strategies that can be upgraded when required to ensure cybersecurity. In this regard, various categories of the dataset may be formed by experts and researchers based on

the time, place, or nature of the concerned dataset. This may be followed by the analysis of each dataset category. The absence of standard benchmarks makes it difficult to present a precise comparison of various techniques of phishing detection. Additionally, insignificant sharing of sensitive data and evolution of phishers leads to the non-availability of reference datasets. Hence, it is the need of the hour for the researchers to make consistent efforts for coming up with strategies for minimizing the adverse influence of the discussed issues on phishing detection techniques.

4. Conclusion

The modern world faces several threats including the significant one of phishing emails, which cause huge financial losses. The preventive methods commonly used today have not proven effective against this threat despite their constant revision. On the other hand, phishing emails have been increasing at unprecedented levels in recent years. To counter this threat of phishing emails, more advanced phishing detection technology is necessary. Anti-phishing technology developed on the source code features is quite slow in terms of the classification of phishing emails given its dependence on third-party services and scraping of the email content. Many ML methods have been adopted to identify phishing emails, but these cannot effectively detect new phishing scams, which needs significant manual feature engineering. We present a survey analysis of actual phishing email identification works from various perspectives. This survey is unique in the sense that it relates works to their openly available tools and resources. The analysis of the presented works revealed that not much work had been performed on phishing email detection using NLP techniques. Therefore, many open issues are associated with this phishing email detection. An evolving research area is illustrated by the phishing email detection The outcomes shown that further work is required to employ modernized DL techniques in phishing email detection studies, for instance, Recurrent Neural Networks (RNNs), Convolutional neural networks (CNN), and Deep Reinforcement Learning models. The tools and resources are not sufficient in this research area. Hence, the researchers are in dire need to perform more research efforts to assess DL techniques in the phishing email detection domain

References

[1] "Anti-Phishing Working Group.," *Phishing Activity Trends Report 1st Quarter 2020.*, 2020. [Online]. Available: https://docs.apwg.org/reports/apwg_trends_report_q3_2020.pdf.

[2] B. Parno, C. Kuo, and A. Perrig, "Phoolproof phishing prevention," in *International conference on financial cryptography and data security*, 2006, pp. 1–19.

[3] R. Verma, N. Shashidhar, and N. Hossain, "Detecting phishing emails the natural language way," in *European Symposium on Research in Computer Security*, 2012, pp. 824–841.

[4] D. Irani, S. Webb, J. Giffin, and C. Pu, "Evolutionary study of phishing," in *2008 eCrime Researchers Summit*, 2008, pp. 1–10.

[5] Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang, "Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism," *IEEE Access*, vol. 7, pp. 56329–56340, 2019.

[6] M. Hiransha, N. A. Unnithan, R. Vinayakumar, K. Soman, and A. D. R. Verma, "Deep learning based phishing e-mail detection," in *Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal.(IWSPA)*, 2018.

[7] C. Coyotes, V. S. Mohan, J. Naveen, R. Vinayakumar, K. P. Soman, and A. D. R. Verma, "ARES: Automatic rogue email spotter," in *Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal.(IWSPA)*, 2018.

[8] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1, no. 2. MIT press Cambridge, 2016.

[9] H. Zuhair, A. Selamat, and M. Salleh, "Feature selection for phishing detection: a review of research," *Int. J. Intell. Syst. Technol. Appl.*, vol. 15, no. 2, pp. 147–162, 2016.

[10] G. Varshney, M. Misra, and P. K. Atrey, "A survey and classification of web phishing detection schemes," *Secur. Commun. Networks*, vol. 9, no. 18, pp. 6266–6284, 2016.

[11] B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting against phishing attacks: state of the art and future challenges," *Neural Comput. Appl.*, vol. 28, no. 12, pp. 3629–3654, 2017.

[12] A. A. Zuraiq and M. Alkasassbeh, "Phishing detection approaches," in *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, 2019, pp. 1–6.

[13] H. Thakur and S. Kaur, "A survey paper on phishing detection," *Int. J. Adv. Res. Comput. Sci.*, vol. 7, no. 4, 2016.

[14] R. Alabdan, "Phishing Attacks Survey: Types, Vectors, and Technical Approaches," *Futur. Internet*, vol. 12, no. 10, p. 168, 2020.

[15] M. Korkmaz, O. K. Sahingoz, and B. Diri, "Feature Selections for the Classification of Webpages to Detect Phishing Attacks: A Survey," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2020, pp. 1–9.

[16] "Available." [Online]. Available: http://www.phishtank.com.

[17] "APWG Phishing Trends Reports, Anti Phishing Working Group," 2016.

[18] G. Xiang, J. Hong, C. P. Rose, and L. Cranor, "Cantina+ a feature-rich machine learning framework for detecting phishing web sites," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 2, pp. 1–28, 2011.

[19] G. Ramesh, I. Krishnamurthi, and K. S. S. Kumar, "An efficacious method for detecting phishing webpages through target domain identification," *Decis. Support Syst.*, vol. 61, pp. 12–22, 2014.

[20] C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," 2010.

[21] C. Karlof, U. Shankar, J. D. Tygar, and D. Wagner, "Dynamic pharming attacks and locked same-origin policies for web browsers," in *Proceedings of the 14th ACM conference on Computer and communications security*, 2007, pp. 58–71.

[22] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013.

[23] M. Langberg, "AOL acts to thwart hackers," *San Jose Mercur. News*, 1995.

[24] K. Rekouche, "Early phishing," *arXiv Prepr. arXiv1106.4692*, 2011.

[25] G. Ollmann, "The phishing guide–understanding & preventing phishing attacks," *NGS Softw. Insight Secur. Res.*, 2004.

[26] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 581–590.

[27] S. Gunawardena, D. Kulkarni, and B. Gnanasekaraiyer, "A steganography-based framework to prevent active attacks during user authentication," in *2013 8th International Conference on Computer Science & Education*, 2013, pp. 383–388.

[28] S. Gupta, A. Singhal, and A. Kapoor, "A literature survey on social engineering attacks: Phishing attack," in *2016 international conference on computing, communication and automation (ICCCA)*, 2016, pp. 537–540.

[29] G. Mujtaba, L. Shuib, R. G. Raj, N. Majeed, and M. A. Al-Garadi, "Email classification research trends: review and open issues," *IEEE Access*, vol. 5, pp. 9044–9064, 2017.

[30] E. S. Gualberto, R. T. De Sousa, P. D. B. Thiago, J. P. C. L. Da Costa, and C. G. Duque, "From Feature Engineering and Topics Models to Enhanced Prediction Rates in Phishing Detection," *IEEE Access*, vol. 8, pp. 76368–76385, 2020.

[31] G. Sonowal and K. S. Kuppusamy, "PhiDMA–A phishing detection model with multi-filter approach," *J. King Saud Univ. Inf. Sci.*, vol. 32, no. 1, pp. 99–112, 2020.

[32] A. Zamir *et al.*, "Phishing web site detection using diverse machine learning algorithms," *Electron. Libr.*, 2020.

[33] and C. Z. S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, "An empirical analysis of phishing blacklists," *Proc. 6th Conf. Email Anti-Spam (CEAS), Sacramento, CA, USA*, pp. 1–10, 2009.

[34] R. V. and N. Hossain, "Semantic feature selection for text with application to phishing email detection," *Proc. Int. Conf. Inf. Secur. Cryptol. Cham, Switz. Springer*, pp. 455–468, 2013.

[35] G. Park and J. M. Taylor, "Using syntactic features for phishing detection," *arXiv Prepr. arXiv1506.00037*, 2015.

[36] A. Vazhayil, N. B. Harikrishnan, R. Vinayakumar, K. P. Soman, and A. D. R. Verma, "PED-ML: Phishing email detection using classical machine learning techniques," in *Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal.(IWSPA)*, 2018, pp. 1–8.

[37] I. R. A. Hamid and J. Abawajy, "Hybrid feature selection for phishing email detection," in *International Conference on Algorithms and Architectures for Parallel Processing*, 2011, pp. 266–275.

[38] A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J.-P. Niyigena, "An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL," *Electronics*, vol. 9, no. 9, p. 1514, 2020.

[39] "No Title." [Online]. Available: www.phishtank.com.

[40] "No Title." [Online]. Available: https://joewein.net/spam/index.htm.

[41] A. Bergholz, J. De Beer, S. Glahn, M.-F. Moens, G. Paaß, and S. Strobel, "New filtering approaches for phishing email," *J. Comput. Secur.*, vol. 18, no. 1, pp. 7–35, 2010.

[42] J. Singh, "Detection of phishing e-mail," *in Proc. IJCST*, vol. 2, no. 1, pp. 547-_549, 2011.

[43] A. Bergholz, J. H. Chang, G. Paass, F. Reichartz, and S. Strobel, "Improved Phishing Detection using Model-Based Features.," in *CEAS*, 2008.

[44] X. Gu and H. Wang, "Online anomaly prediction for robust cluster systems," in *2009 IEEE 25th International Conference on Data Engineering*, 2009, pp. 1000–1011.

[45] C. N. Gutierrez *et al.*, "Learning from the ones that got away: Detecting new forms of phishing attacks," *IEEE Trans. Dependable Secur. Comput.*, vol. 15, no. 6, pp. 988–1001, 2018.

[46] X. Glorot, A. Bordes, and Y. Bengio, “Domain adaptation for large-scale sentiment classification: A deep learning approach,” in *ICML*, 2011.

[47] T. H. Nguyen and R. Grishman, “Relation extraction: Perspective from convolutional neural networks,” in *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 2015, pp. 39–48.

[48] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *arXiv Prepr. arXiv1409.0473*, 2014.

[49] T. Repke and R. Krestel, “Bringing back structure to free text email conversations with recurrent neural networks,” in *European Conference on Information Retrieval*, 2018, pp. 114–126.

[50] F. Chollet, *Deep learning with Python*, vol. 361. Manning New York, 2018.

[51] G. Sonowal, “Phishing Email Detection Based on Binary Search Feature Selection,” *SN Comput. Sci.*, vol. 1, no. 4, 2020.

[52] L. F. Gutiérrez, F. Abri, M. Armstrong, A. S. Namin, and K. S. Jones, “Phishing Detection through Email Embeddings,” *arXiv Prepr. arXiv2012.14488*, 2020.

[53] A. Kumar, J. M. Chatterjee, and V. G. Díaz, “A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing,” *Int. J. Electr. Comput. Eng.*, vol. 10, no. 1, p. 486, 2020.

[54] E. S. Gualberto, R. T. De Sousa, P. D. B. Thiago, J. P. C. L. Da Costa, and C. G. Duque, “The Answer is in the Text: Multi-Stage Methods for Phishing Detection based on Feature Engineering,” *IEEE Access*, 2020.

[55] A. Ora, “Spam Detection in Short Message Service Using Natural Language Processing and Machine Learning Techniques.” Dublin, National College of Ireland, 2020.

[56] S. R. Mirhoseini, F. Vahedi, and J. A. Nasiri, “E-Mail phishing detection using natural language processing and machine learning techniques.”