A Systematic Review on Phishing Detection: A Perspective Beyond a High Accuracy in Phishing Detection

Daniel Alejandro Barreiro Herrera(✉) and Jorge Eliecer Camargo Mendoza(✉)

Universidad Nacional de Colombia, Bogotá, Colombia
{dabarreiroh,jecamargom}@unal.edu.co
https://unal.edu.co/

Abstract. Phishing is one of the cyberattacks most feared by users who use transactional services over the Internet, although there are a lot of studies focused on detecting phishing attacks showing high accuracy, those have problems acting with the effectiveness required to prevent people to fall into these attacks in the early stages. In this article, a state-of-the-art overview of phishing detection is shown using a systematic literature review methodology for studies addressed between 2016 and 2022, such as other survey papers between 2020 and 2022, focused on the different detection stages, information sources, phishing characterization, and different methods used in the literature. Found that 83% of applications works selected are focused on the mitigation stage, where the methodologies act in reactive ways using statics features that provides high accuracy but turn the models fail through time. Finally, conclusions will be presented to highlight the importance of using brand information and mixing different methods to improve stage detection and assure durability in the detection model. The article's contribution is focused on establishing another perspective that encourages future research and future related works to consider their models beyond a high accuracy and start thinking about how these models can to provide effective solutions that could be integrated into production environments to protect the users.

Keywords: Phishing · State of the art · Detection · Stages · Systematic review

1 Introduction

Our environment and how we interact as a society has lived a great transformation in recent years. Technology and especially advances in communications have been largely responsible for this, providing alternatives to carry out actions that previously were carried out in physical sites consuming a lot of time on a daily basis, and now they can be carried out with just one click. The portfolio of services available on the web every day is more diverse and complete, after

H. Florez and H. Gomez (Eds.): ICAI 2022, CCIS 1643, pp. 173–188, 2022.
https://doi.org/10.1007/978-3-031-19647-8_13

the pandemic episode it is not just an alternative, this issue has accelerated the digital transformation of many organizations, retail and small business, which has been motivated to publish via web site their services, for that reason safety for people who interact with them is a subject of great importance.

In this field there are two main actors, providers that offer services to increase the number of clients who make use of them, seeking a good experience that ensures continued use on posterity, and on the other hand, there are the users of these services who despite the comfort that these services can provide, they are not willing to sacrifice security in their transactions that may affect them monetarily or violate their data privacy. According to studies such as [1] Anti-Phishing Working Group in Q4 of 2021, phishing as an attack is one of the most suffered by users on the web, this study shows an amount of 316,747 attacks in December 2021 and it is the month with the most count of phishing attacks in the history. This report also mentions that the number of attacks at the end of 2021 has tripled the number of early attacks in the early months of 2020. It also makes an analysis most targeted industry sectors and found that the financial sector is one of the most suffered (23.2%) followed closed by SAAS/Web-mail (19.5%) and e-commerce/retail services (17.3%)

The same previous analysis was done by [2] and [3] in his introduction 2 and 4 years ago respectively and surprisingly these stats have continued growing in the following years. That shows that, although the research on phishing detection is varied and not especially a recent topic, the application of mechanisms that allow not only mitigating but preventing users from falling into the early stages of phishing should be strengthened. This review paper intends to conduct a study of the state of the art in detecting phishing and establish key points where research still has shortcomings.

In [4]it is described a review of research related to phishing and what it calls security challenges, Fig. 1 shows the aspects that this article wants to highlight.

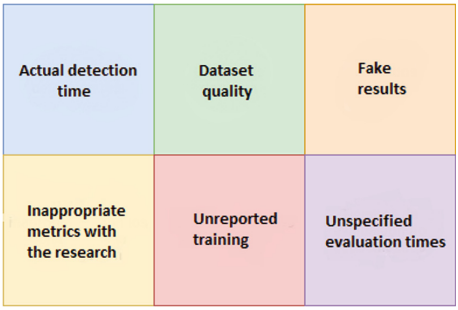

Fig. 1. Challenges in phishing detection models

The reality is that some research, especially engineering research, should be concerned with innovating previously unused methods, looking for different configurations that provide better results, or trying new approaches, it should

also prioritize risks and reevaluate objectives. Phishing detection research is not particularly new and targeting efforts to detect as many malicious URLs as possible might seem like a good goal, seeking to remove the manual effort from security SOCs and ranking potentially dangerous URLs rather than a list of websites that are not. Far from the purpose of this article is to point out that this objective is bad or unnecessary, otherwise, it is considered very important and it is proposed a thought related to changing the approach in which the detection of phishing must act by mixing various methods, various actors in the process and overall considering of vital importance to act in the stage of prevention.

Another aspect to consider is the change of the characteristics of phishing in the time [4]. Phishing of 2022 is different from the phishing of 2018 in just 4 years the attackers have found ways to steal the information of the users through forms, hacked sites, free hosting services, and tunneling of local sites. For that reason, it is necessary to consider characteristics that could give us not only high accuracy in the detection currently if else also can achieve mechanisms that can be adapted to different techniques and that can identify key elements of the features of phishing to act in the early stages allowing to keep the performance of the solution for an acceptable time after the implementation.

This paper will check the studies related to phishing detection, taking into account the chronology of the different studies, and identifying the stages in which they act. The article is structured with an initial explanation of the methodology used, with the research questions formulation, and next with the literature selection to build a frame in phishing detection that allows analysis and synthesis using 4 pillars found in the review. Finally, the findings will enable it to reflect on the features identified in state of the art and adjust a model proposal for phishing identification that manages to act in the early stages of detection.

2 Methodology

2.1 Research Question Formulation

Let's launch a premise: "It is necessary to avoid that phishing attacks catch the people". So, there are many ways to be approached a possible solution, but particularly we are thinking here in some collection of programs, algorithms, and validations that alert previously a user that could fall on the attack. So the following guiding questions were raised taking into account the phishing detection broad topic:

- What are the characteristics of a phishing attack?
- What are the currently used mechanisms to identify phishing?
- What are the characteristics found in the literature used in phishing detection?
- When phishing attacks are found?
- What characteristics are attributable to the brand affected by phishing and which are typical of a generic phishing attack?
- What are the challenges and/or gaps in phishing detection?

2.2 Sourcing of Relevant Literature

The chosen articles were obtained from a search equation based on the research questions: TITLE-ABS-KEY ("Characteristics of a phishing" OR "mechanisms to identify phishing" OR "extract features phishing" OR "features brand phishing" OR "features generic phishing" OR "Machine learning phishing detection") Bibliographic database such as IEEE explorer where used in this literature search. Finally another articles where added from the citations of the first articles, taking into account some review articles and organizations mentioned in the first articles such as AWPG.

2.3 Literature Selection

Based on the research question formulation, search equations were executed to reach a reference frame that allowed identifying 4 pillars of this research, these are: When? (detection stage), Where? (information sources), What? (Phishing features), How? (Phishing methods detection). 244 research articles were collected from the first decade of 2000 up to date. However, in the first review, it was found that more recent works already covered previous studies taking as global categories detection methods as List, Heuristically, Machine learning and it was also necessary to consider a great variation in the techniques used by cyber-criminals. The same happens concerning the technologies and procedures used in the implementation of web pages. So 57 articles were identified as key in phishing detection literature in recent years, taking into account that the objective given by the thematic is identifying the 4 pillars in the articles, identifying the point of action of each one of the investigations, identifying the information sources, the characterization of the attack and the methodology used for detection.

3 Phishing Detection Stages, When?

To understand when the studies of phishing attacks are acting, it is important to identify different types of approaches in the literature that deal with the problem, and according to that analyze and present their results. Lets to show it through an example:

Lets to analyze different studies acting in completely different detection stage, studies such as [5] focuses its efforts on validating phishing from an already deployed URL, it is a mean which already exists the attack. Probably some people already also has received the same URL and some of them have fallen into phishing, meanwhile, some others have been instructed to avoid falling into it. On the other side studies such as [6] focus on creating domain generation algorithms that allow it to act in zero time; it is mean that this algorithm could identify a potential attack of phishing even before this attack would be deployed.

Although both of them are aimed at detecting phishing, they differ greatly on: methodology, stage detection, sources, and techniques used. Considering the

result [6] presents accuracy less than 5% and [5] accuracy above 90%, however in where no person had to have fallen on fraud, while the accuracy can cover a wider range of brands taking into account that cost can not be quantified on how many people might fall on the attack before it is being detected. Taking this into account here a proposal of the stages for phishing detection will be proposed:

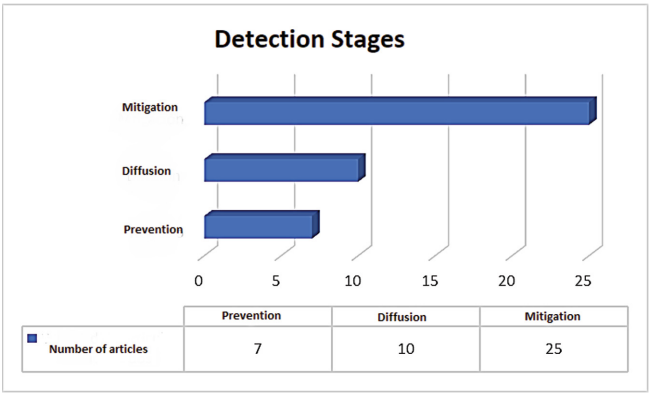

Fig. 2. Description of the detection stage in which the reviewed papers are located

3.1 Prevention Stage

Where there are jobs for generating domains such as nakamura 2019 [6], Buber 2017 [7], Adil 2020 [8], Spaulding 2016 [9], Starov 2019 [10], Ginsberg 2018 [11] and Li 2016 [12] based on features extracted, this stage is perfect pipelining for applications that actually prevent the spread of phishing before reaching a user or a propagation medium.

3.2 Diffusion Stage

In this stage there are works such as Ya 2019 [5], Li 2017 [13], Li 2020 [14], Eshmawi 2019 [15], Balim 2019 [16], Dalgic 2018 [17], Yan 2020 [18], Sahoo 2018 [19], Baykara 2018 [20], Lingam 2018 [21] and Lingam 2019 [22] related to identify mechanisms in which phishing reaches end users; this is how analyzes are presented on social networks or email dissemination, among others.

3.3 Mitigation Stage

Final stage of action on which 83% models studied act based on community databases or reported URLs such as Phishtank or Openphish.

Currently, the 35 articles and investigations here studied have focused on the last two stages, seeking to identify and study the means by which phishing

spreads and how it is dispersed or analyzing the final URL in which users have already fallen.

For practical purposes throughout this paper, we will refer to these 3 previous stages as prevention, propagation, and mitigation, as it is depicted in Fig. 2.

4 Phishing Information Sources ¿Where?

In studies such as Li 2017 [13], Sharma 2017 [23] and Pande 2017 [24] different sources were found, used either for the own study of phishing characteristics, or for the validation of results such as Adil 2020 [8] and Li 2016 [12]. For this, it is essential to count the sources of information that link phishing sites or at least that allow extracting of URLs related to phishing features. Here are some sources considered useful at different stages of detection:

Table 1. Some sources used in phishing search

Source name	Description	Ref
Phishtank	PhishTank is a collaborative clearinghouse for data and information about phishing on the Internet	[12,23,24,33]
APWG/ecx APWG	Different types of anti-phishing working group tools focused on the detection and centralization of information about phishing	[4]
Openphish	OpenPhish provides cyber-threat intelligence services	

Table 1 shows some sources found on the web to obtain websites reported by the community or specialized teams, where usually researchers can put together their data sets and make a preliminary analysis of the characteristics of phishing attacks. However, it is necessary to mention that these types of sources present different types of utilities for different users involved in the detection. However, although it is a good way to centralize information regarding active phishing, acting with these data for the detection of phishing would help only in the mitigation stage. Thus, it is possible to highlight what stage of detection certain investigations are at, based on the choice of their data sources. An investigation that wants to act in stages of dissemination, will look for sources related to social networks, emails, or web advertising, while a mitigation stage would use sources related to Table 1 that would allow automating classification processes where it would help in more systematized processes to classify URLs with more elements; while for early stages it would be ideal to act within the domain registry itself, where clues begin to be given that the domain is focused on impersonating another web page.

Table 2 shows some sources focused on domain detection from its prevention stage. They are sources that, based on a keyword, can allow searches for recently

Table 2. Some sources that can be used in prevention stages

Source name	Description	Ref
Domainwatch	Useful tool to search information of a domain from a keyword. Provides historical and current information	[25]
Urlscan	urlscan.io is a service for scanning and analyzing websites that can search for domains using a keyword	[26]

registered domains, as well as displaying whois, associated security certificates, and other domains.

5 Phishing Characterization What?

5.1 Challenges for Feature Selection

In studies such as Zhu 2018 [27] and Yang 2019 [28] authors seek to diversify the characteristics used in searching for higher detection accuracy. However, the more characterized phishing today is more susceptible to future attack changes. Within the features mentioned in the literature such as Aung 2019 [29], Eshmawi 2019 [15], McGahagan 2019 [30] and Yuan 2018 [31] as a phishing alarm, a wide variety of options are presented. They are considered to act at different stages since some depend on whether the URL is already deployed with a phishing attack, while others, such as those extracted in heuristic methods, know what they are specifically looking for and this could allow them to identify these characteristics in early stages, These features have different extraction mechanisms and different requirements to be able to quantify them. Ideally, a complete system should encompass the extraction of characteristics in all possible stages, although the ideal would be to identify in the prevention stage, so the greatest number of possible features in the first stage would be ideal. However, features of this type should be searched in the sources shown in Table 2. It is difficult to search in these sources if it is not sure what it is looking for, it is at this point where features related to the brand can help us search in the great number of domains registered every second and can help us validate phishing before it is in a diffusion stage.

6 Phishing Detection How?

6.1 Surveyed Papers for Detecting Phishing

It is considered important to start with the current approaches in phishing detection since it will give a general idea of what is being implemented and possible approaches to these methods (Fig. 3).

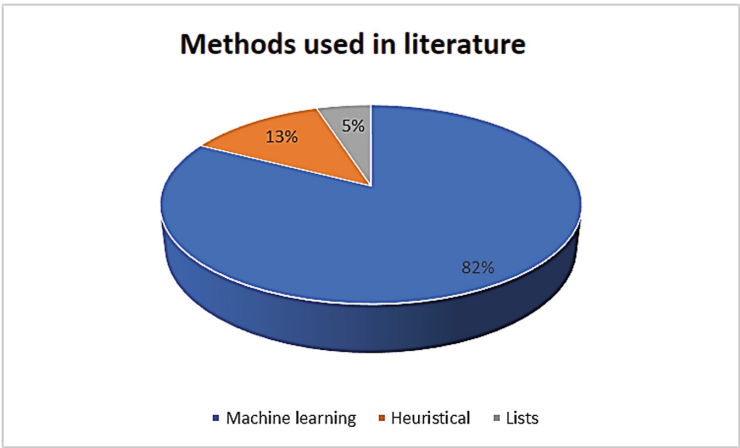

Fig. 3. Distribution of methods used in the literature.

6.2 Blacklist-Whitelist Approach

It is inevitable to mention the topic corresponding to the use of stored data for blocking IP or domains already detected as fraudulent since in articles as Buber 2017 [7], Mondal 2019 [32] and Patil 2018 [3] they are considered important features to take into account in more structured systems, where they can help mitigate an attack, being the most relevant fact of this model has a very effective mitigation potential if there is a system of interconnected information browsers. [33]

These types of strategies are still useful because they allow acting where the other validation systems have failed and although in lesser numbers, they may be able to act in early stages based on notorious precedents, such as past attacks reactivation or malicious IPs blocking.

6.3 Heuristic Approaches

The heuristic approaches depend on the quality of the features that are extracted some articles such as Ali 2019 [34], Huang 2019 [35], nakamura 2019 [6], Nathezhtha 2019 [36], Baral 2019 [37]. In these works results allow visualizing expected behaviors, that is, in the case of these implementations, it is necessary to know what is being looked for and based on this build the algorithms that allow the identification of these expected characteristics. This type of implementation can act in any of the three stages depending on the approach for which it is designed as in [6], where it is used for early detection. But it can also be used in mitigation stages since it can be based on features of a URL of phishing already deployed. In this type of method, it is important to have information about keywords and patterns that can be effective in the prevention stage. The biggest disadvantage is the static detection that results in evaluating additional non-obvious aspects.

6.4 Machine Learning

The other mechanism is the use of tools that monitor URLs and seek to give a risk based on characteristics detected on a certain web page. The tools that use machine learning have proven to have the best results in recent years [38]. Based on this risk, decisions can be made to mitigate the impact of fraud [33].

Although machine learning detection algorithms give the best results, there is still a gap in differentiating between phishing detection and validation. And the difference between these two concepts lies in the stage where the tests of these models are implemented. In 25 models studied, tests are performed on deployed URLs, so rather than detection, they are phishing validation systems acting in the mitigation stages. There is a general absence of evaluating results, being implemented in real environments and over a long period of time, to evaluate the accuracy that machine learning algorithms give against the changes that attackers show in their attacks.

6.5 Detection Phishing Methods

The methods used in the detection phishing process describe a set of phases to structure the design, the technical implementation, the results analysis and also the context of the application, and the best conditions for use it is considered. In a General classification, there are three big methods used in current research these are based list detection, heuristics detection, and machine learning detection.

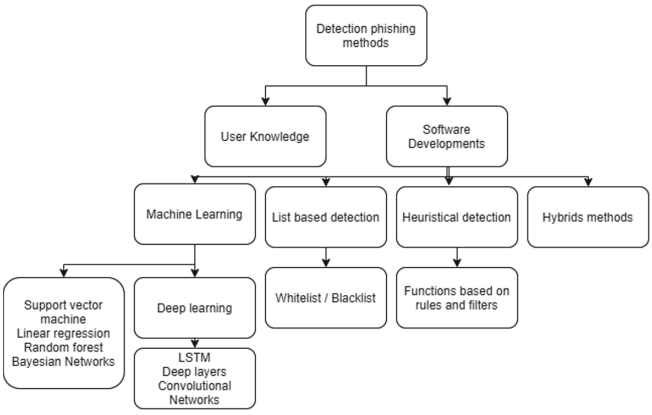

Fig. 4. Distribution of methods used in the literature.

The Fig. 4 can provide a review of methods used in detection phishing researches.

7 Opportunities for Future Researches

7.1 Identifying Challenges in the Review

An important aspect to consider is the pattern found in most of the studied identification systems. It is important to have in mind that there are studies [4] that have identified different challenges in each one of these processes and for the case of this work are considered appropriate to mention.

1. Source extraction stage: This stage presents the obtaining information challenge that is not biased to a certain group of threats, as well as obtaining the information in real-time, in general, shortening the gap that limits the quality of the data and the ideal development of research.
2. Data analysis and relevant data extraction: At this stage, there are still quite a few challenges to explore. [4] raises two main ones that are related to the quality of the data from which these characteristics are extracted and the time scaled since the attack is active.
3. Training and/or adjustment of the system: At this stage, the challenge of configuring the appropriate features are posed so that while the system learns it can be adaptable not only to the data which it was trained-configured but also could help with attacks that come to the future.
4. System evaluation: There are challenges in that proper evaluation parameter must be sought that not only depend on a correct interpretation of the results but also depends on the quality of the past stages to provide information beyond just the effectiveness of "validating phishing".

7.2 Include of Brand Information to Improve Phishing Representation

It is necessary to identify the big majority of features related to phishing at an early stage and consider elements such as changing these features over time, the change in the technology, and security certificates. They should be considered in the analysis of the choice of features to change the global understanding of the problem to protect specific brands that allow user protection before the attack is widespread. For it, the industry should assume the role of protecting the service offered to the users and implement customized systems to address the problem from the characteristics of its fraud threats. In that way, some researchers were found to use a different approach that could be used under the concept of brand features such as Zuraiq 2019 [39], Ginsberg 2018 [11] and Concone 2019 [40]. Related work was found as an example of email-phish with high similarity, demonstrating recurrent neural networks with an accuracy of more than 98% [5]. Figure 5 presents features that can be extracted in each stage.

Additionally [41] shows how from NLP-W2V-based feature extraction it is possible to run a model that can be tracked in real-time. However, a connection is not established with phishing prevention stages, so it is not possible to determine

that it can work in previous stages, but it shows how it could be used for the extraction of possible brand features and their vector representation.

Another approach [42] provides results using the extraction characteristics of logos showing an accuracy of 97%. So consideration should be given to using image validation from early stages.

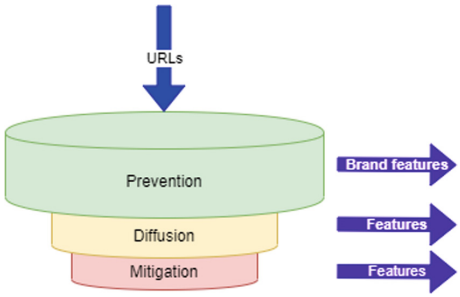

Fig. 5. Features extraction diagram and its importance in the different stages

7.3 How Patterns and Brand Features Can be Used in Prevention Stages

The most valuable improvement that the new research can develop is the capability to increase the detection in prevention stages according to Fig. 2. Additionally, these must increase the accuracy in this stage, the majority of methods used in this stage are associated with heuristics and list methods but Machine Learning methods are low. The machine learning methods need a number of relevant features that can teach the model to learn the characteristics of phishing attacks. But what kind of features can a model take of a recently registered domain? the answer is probably none with the current approach. A recent domain just has a sequence of chars associated and probably information through a WHOIS request.

Here is where a new approach can appear taking into account the patterns and the acknowledgment of the brands, although it could sound like a start of heuristics methods, a machine learning model that can learn the patterns and features associated with a brand could be powerful recognizing phishing for a specific brand. So the importance of studying these features could be of vital importance in developing a detection phishing system in the prevention stage. Here are described global features to be taken into account for this proposal:

Attacker Characterization. A domain recently registered can detect a number of features related to the WHOIS record, such as registrant data, registrar, hosting, country, and actives services as possible MX record, as also can be detected ssl certified and his respective organization.

Character Analysis. Generally, the domains can contain the name of the affected brand or similar chars [6] that can be processed with just the domain considering additionally features as [12] and involving methods such as Ya 2019 [5], McGahagan 2019 [30] and Xiang [43].

Brand Characterization. As in Sect. 5.1 described knowledge about the brand can provide a detection system to provide security to users that want to access the offered services, the knowledge about the common colors used, the official domain or IPs associated, the language used in the official pages, as also the patterns used commonly in phishing attacks as keywords and patterns in paths could provide high-quality features for a robust system detection.

8 Conclusions

After reviewing the research, although there are good results regarding phishing "detection-validation" from a sample of URLs, showing accuracy detection above 90%, in most cases, there are still many challenges that suggest trying other approaches from different perspectives to achieve comprehensive action in the 3 stages described in this paper. Although some approximations may be appropriated for a certain group of threats, it might be good to review the methodologies used to be more effective in seeking benefits applicable to reality. There is a problem in providing models that act in the early stages where the objective is not mitigation, but prevention. Acting from the domain registry itself is not possible if it is not known what is being searched in the registered domains or what content within the domain can help to identify a potential threat. Additionally, the implemented model is required to have adaptability over time, an aspect that was not found in the results of any of the related works. The industry should understand that it should protect the service offered to the users by implementing custom systems to address the problem from its own features. For this reason, a particular approach to avoid phishing from the affected brand should be considered. Identifying particular brand features could help to be focused on early detection.

References

1. apwg: Phishing activity trends report Q4 2021 (2022). http://www.apwg.org
2. Athulya, A.A.: Towards the detection of phishing attacks Praveen K TIFAC-CORE in cyber security Amrita Vishwa Vidyapeetham (2020). ISBN 9781728155180
3. Patil, V., Thakkar, P., Shah, C., Bhat, T., Godse, S.P.: Detection and prevention of phishing websites using machine learning approach. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1–5 (2018). https://doi.org/10.1109/ICCUBEA.2018.8697412
4. Das, A., Baki, S., Aassal, A.E., Verma, R., Dunbar, A.: SoK: a comprehensive reexamination of phishing research from the security perspective. IEEE Commun. Surv. Tutor. **22**(1), 671–708 (2020). https://doi.org/10.1109/COMST.2019.2957750. ISSN 1553-877X VO - 22

5. Ya, J., Liu, T., Zhang, P., Shi, J., Guo, L., Gu, Z.: NeuralAS: DeepWord-based spoofed URLs detection against strong similar samples. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–7 (2019). ISBN 2161-4407 VO. https://doi.org/10.1109/IJCNN.2019.8852416
6. Nakamura, A., Dobashi, F.: Proactive phishing sites detection. In: IEEE/WIC/ACM International Conference on Web Intelligence, Series WI 2019, pp. 443–448. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3350546.3352565. ISBN 9781450369343
7. Buber, E., Demir, Ö., Sahingoz, O.K.: Feature selections for the machine learning based detection of phishing websites. In: 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1–5 (2017). https://doi.org/10.1109/IDAP.2017.8090317. ISBN: VO
8. Adil, M., Khan, R., Ghani, M.A.N.U.: Preventive techniques of phishing attacks in networks. In: 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), pp. 1–8 (2020). https://doi.org/10.1109/ICACS47775.2020.9055943. ISBN: VO
9. Spaulding, J., Upadhyaya, S., Mohaisen, A.: The landscape of domain name typosquatting: techniques and countermeasures. In: 2016 11th International Conference on Availability, Reliability and Security (ARES), pp. 284–289 (2016). https://doi.org/10.1109/ARES.2016.84. ISBN: VO
10. Starov, O., Zhou, Y., Wang, J.: Detecting malicious campaigns in obfuscated JavaScript with scalable behavioral analysis. In: 2019 IEEE Security and Privacy Workshops (SPW), pp. 218–223 (2019). https://doi.org/10.1109/SPW.2019.00048. ISBN: VO
11. Ginsberg, A., Yu, C.: Rapid homoglyph prediction and detection. In: 2018 1st International Conference on Data Intelligence and Security (ICDIS), pp. 17–23 (2018). https://doi.org/10.1109/ICDIS.2018.00010. ISBN: VO
12. Li, X., Geng, G., Yan, Z., Chen, Y., Lee, X.: Phishing detection based on newly registered domains. In: 2016 IEEE International Conference on Big Data (Big Data), pp. 3685–3692 (2016). https://doi.org/10.1109/BigData.2016.7841036. ISBN: VO
13. Li, J., Wang, S.: PhishBox: an approach for phishing validation and detection. In: 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), pp. 557–564 (2017). https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.101. ISBN: VO
14. Li, Q., Cheng, M., Wang, J., Sun, B.: LSTM based phishing detection for big email data. IEEE Trans. Big Data 1 (2020). https://doi.org/10.1109/TBDATA.2020.2978915. ISSN 2332–7790 VO
15. Eshmawi, A., Nair, S.: The roving proxy framewrok for SMS spam and phishing detection. In: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1–6 (2019). https://doi.org/10.1109/CAIS.2019.8769562. ISBN: VO
16. Balim, C., Gunal, E.S.: Automatic detection of smishing attacks by machine learning methods. In: 2019 1st International Informatics and Software Engineering Conference (UBMYK), pp. 1–3 (2019). https://doi.org/10.1109/UBMYK48245.2019.8965429. ISBN: VO

17. Dalgic, F.C., Bozkir, A.S., Aydos, M.: Phish-IRIS: a new approach for vision based brand prediction of phishing web pages via compact visual descriptors. In: 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1–8 (2018). https://doi.org/10.1109/ISMSIT.2018.8567299. ISBN: VO
18. Yan, X., Xu, Y., Xing, X., Cui, B., Guo, Z., Guo, T.: Trustworthy network anomaly detection based on an adaptive learning rate and momentum in IIoT. IEEE Trans. Ind. Inform. 1 (2020). https://doi.org/10.1109/TII.2020.2975227. ISSN 1941-0050 VO
19. Sahoo, P.K.: Data mining a way to solve phishing attacks. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1–5 (2018). https://doi.org/10.1109/ICCTCT.2018.8550910. ISBN: VO
20. Baykara, M., Gürel, Z.Z.: Detection of phishing attacks. In: 2018 6th International Symposium on Digital Forensic and Security (ISDFS), pp. 1–5 (2018). https://doi.org/10.1109/ISDFS.2018.8355389. ISBN: VO
21. Lingam, G., Rout, R.R., Somayajulu, D.V.L.N.: Detection of social botnet using a trust model based on spam content in Twitter network. In: 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), pp. 280–285 (2018). https://doi.org/10.1109/ICIINFS.2018.8721318. ISBN 2164-7011 VO
22. Lingam, G., Rout, R.R., Somayajulu, D.V.L.N.: Deep Q-learning and particle swarm optimization for bot detection in online social networks. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6 (2019). https://doi.org/10.1109/ICCCNT45670.2019.8944493. ISBN: VO
23. Sharma, H., Meenakshi, E., Bhatia, S.K.: A comparative analysis and awareness survey of phishing detection tools. In: 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 1437–1442 (2017). https://doi.org/10.1109/RTEICT.2017.8256835. ISBN: VO
24. Pande, D.N., Voditel, P.S.: Spear phishing: diagnosing attack paradigm. In: 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 2720–2724 (2017). https://doi.org/10.1109/WiSPNET.2017.8300257. ISBN: VO
25. DomainWatch, DomainWatch - Domain WHOIS Search, Website Information. https://domainwat.ch/
26. urlscan, URL and website scanner. https://urlscan.io/
27. Zhu, E., Ye, C., Liu, D., Liu, F., Wang, F., Li, X.: An effective neural network phishing detection model based on optimal feature selection. In: 2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), pp. 781–787 (2018). https://doi.org/10.1109/BDCloud.2018.00117. ISBN: VO
28. Yang, P., Zhao, G., Zeng, P.: Phishing website detection based on multidimensional features driven by deep learning. IEEE Access **7**, 15 196–15 209 (2019). https://doi.org/10.1109/ACCESS.2019.2892066. ISBN: 2169-3536 VO - 7
29. Aung, E.S., Yamana, H.: URL-based phishing detection using the entropy of non-alphanumeric characters. In: Proceedings of the 21st International Conference on Information Integration and Web-Based Applications & Services, iiWAS2019, v. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3366030.3366064. ISBN 9781450371797

30. McGahagan, J.. Bhansali, ,D, Gratian, M., Cukier, M.: A comprehensive evaluation of HTTP header features for detecting malicious websites. In: 2019 15th European Dependable Computing Conference (EDCC), pp. 75–82 (2019). https://doi.org/10.1109/EDCC.2019.00025. ISBN 2641-810X VO
31. Yuan, H., Chen, X., Li, Y., Yang, Z., Liu, W.: Detecting phishing websites and targets based on URLs and webpage links. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 3669–3674 (2018). https://doi.org/10.1109/ICPR.2018.8546262. ISBN 1051-4651 VO
32. Mondal, S., Maheshwari, D., Pai, N., Biwalkar, A.: A review on detecting phishing URLs using clustering algorithms. In: 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), pp. 1–6 (2019). https://doi.org/10.1109/ICAC347590.2019.9036837. ISBN: VO
33. Megha, N., Babu, K.R.R., Sherly, E.: An intelligent system for phishing attack detection and prevention. In: 2019 International Conference on Communication and Electronics Systems (ICCES), pp. 1577–1582 (2019). https://doi.org/10.1109/ICCES45898.2019.9002204. ISBN: VO
34. Ali, W., Ahmed, A.A.: Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting. IET Inf. Secur. **13**(6), 659–669 (2019). https://doi.org/10.1049/iet-ifs.2019.0006. ISSN 1751-8717 VO - 13
35. Huang, Y., Qin, J., Wen, W.: Phishing URL detection via capsule-based neural network. In: 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), pp. 22–26 (2019). https://doi.org/10.1109/ICASID.2019.8925000. ISBN 2163-5056 VO
36. Nathezhtha, T., Sangeetha, D., Vaidehi, V.: WC-PAD: web crawling based phishing attack detection. In: 2019 International Carnahan Conference on Security Technology (ICCST), pp. 1–6 (2019). https://doi.org/10.1109/CCST.2019.8888416. ISBN 2153-0742 VO
37. Baral, G., Arachchilage, N.A.G.: Building condence not to be phished through a gamified approach: conceptualising user's self-efficacy in phishing threat avoidance behaviour. In: 2019 Cybersecurity and Cyberforensics Conference (CCC), pp. 102–110 (2019). https://doi.org/10.1109/CCC.2019.000-1. ISBN: VO
38. Anand, A., Gorde, K., Moniz, J.R.A., Park, N., Chakraborty, T., Chu, B.: Phishing URL detection with oversampling based on text generative adversarial networks. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 1168–1177 (2018). https://doi.org/10.1109/BigData.2018.8622547. ISBN: VO
39. Zuraiq, A.A., Alkasassbeh, M.: Review: phishing detection approaches. In: 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), pp. 1–6 (2019). https://doi.org/10.1109/ICTCS.2019.8923069. ISBN: VO
40. Concone, F., Re, G.L., Morana, M., Ruocco, C.: Assisted labeling for spam account detection on Twitter. In: 2019 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 359–366 (2019). https://doi.org/10.1109/SMARTCOMP.2019.00073. ISBN: VO
41. Yazhmozhi, V.M., Janet, B.: Natural language processing and machine learning based phishing website detection system. In: 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 336–340 (2019). https://doi.org/10.1109/I-SMAC47947.2019.9032492. ISBN: VO

42. Yao, W., Ding, Y., Li, X.: LogoPhish: a new two-dimensional code phishing attack detection method. In: 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), pp. 231–236 (2018). https://doi.org/10.1109/BDCloud.2018.00045. ISBN: VO
43. Xiang, G., Hong, J., Rose, C.P., Cranor, L.: CANTINA+: a featurerich machine learning framework for detecting phishing web sites (2011)