A Comprehensive Survey of AI-Enabled Phishing Attacks Detection Techniques

Manjunath A C[1] , Dinesh Kumar K[2], Prajwal Kanthan T[3], Akash S[4], Dr. Kayalvizhi[5]

[1,2,3,4,5] Presidency School of Computer Science And Engineering, Presidency University Bangalore – 560064

---****************---

ABSTRACT

In recent times, a phishing attack has become one of the most prominent attacks faced by internet users, governments, and service-providing organizations. In a phishing attack, the attacker(s) collects the client's sensitive data (i.e., user account login details, credit/debit card numbers, etc.) by using spoofed emails or fake websites. Phishing websites are common entry points of online social engineering attacks, including numerous frauds on the websites. In such types of attacks, the attacker(s) create website pages by copying the behavior of legitimate websites and sends URL(s) to the targeted victims through spam messages, texts, or social networking. To provide a thorough understanding of phishing attack(s), this paper provides a literature review of Artificial Intelligence (AI) techniques: Machine Learning, Deep Learning, Hybrid Learning, and Scenario-based techniques for phishing attack detection. This paper also presents the comparison of different studies detecting the phishing attack for each AI technique and examines the qualities and shortcomings of these methodologies. Furthermore, this paper provides a comprehensive set of current challenges of phishing attacks and future research direction in this domain.

Keywords Phishing Attack · Security Threats · Advanced Phishing Techniques · Cyber Attack · Internet Security · Machine Learning · Deep Learning · Hybrid Learning

INTRODUCTION

The process of protecting cyberspace from attacks has come to be known as Cyber Security . Cyber Security is all about protecting, preventing, and recovering all the resources that use the internet from cyber-attacks . The complexity in the cybersecurity domain increases daily, which makes identifying, analyzing, and controlling the rel- evant risk events significant challenges. Cyberattacks are digital malicious attempts to steal, damage, or intrude into the personal or organizational confidential data Phish- ing attack uses fake websites to take sensitive client data, for example, account login credentials, credit card numbers, etc. In the year of 2018, the Anti-Phishing Working Group (APWG) detailed above 51,401 special phishing websites. Another report by RSA assessed that worldwide associations endured losses adding up to $9 billion just due to phishing attack happenings in the year 2016. These stats have demonstrated that the current anti-phishing techniques and endeavors are not effective. Figure 1 shows how a typical phishing attack activity happens.

Personal computer clients are victims of phishing attack because of the five primary reasons:

(1) Users do not have brief information about Uniform Resource Locator (URLs),
(2) the exact idea about which pages can be trusted,
(3) entire location of the page because of the redirection or hidden URLs, (4) the URL possess many possible options, or some pages accidentally entered, (5) Users cannot differ- entiate a phishing website page from the legitimate ones.
Phishing websites are common entry points of online social engineering attacks, including numerous ongoing web scams

In such type of attacks, the attackers create web- site pages by copying genuine websites and send suspicious URLs to the targeted victims through spam messages, texts, or online social networking. An attacker scatters a fake vari- ant of an original website, through email, phone, or content messages , with the expectation that the targeted victims would accept the cases in the email made. They will likely target the victim to include their personal or highly sensi- tive data (e.g., bank details, government savings number, etc.). A phishing attack brings about an attacker acquiring bank card information and login data. In any case, there are a few methods to battle phishing . The expanded utilization of Artificial Intelligence (AI) has affected essen- tially every industry, including cyber-security. On account of email security, AI has brought speed, accuracy, and the capacity to do a detailed investigation. AI can detect spam, phishing, skewers phishing, and different sorts of attacks uti- lizing previous knowledge in the form of datasets. These type of attacks likely creates a negative impact on clients' trust toward social services such as web services. According to the APWG report, 1,220,523 phishing attacks have been reported in 2016, which is 65% more expansion than 2015 .

Figure 2 shows the Phishing Report for the third quarter of 2019.

As per Parekh et al. a generic phishing attack has four stages. First, the phisher makes and sets up a fake website that looks like an authentic website. Secondly, the person sends a URL connection of the website to a targeted victim pretending like a genuine organization, user, or associ- ation. Thirdly, the person in question will be tempted to visit the injected fake website. Fourth, the unfortunate targeted victim will click on the fake source link and give his/her valuable data as input. By utilizing the individual data of the person in question, impersonation activities will be performed by the phisher. APWG contributes individual reports on phishing URLs and analyzes the regularly evolving nature and procedures of cybercrimes. The Anti-Phishing Working Group (APWG) tracks the number of interesting phishing websites, an essential proportion of phishing over the globe. Phishing locales dictate the interesting base URLs. The absolute number of phishing websites recognized by APWG in the 3rd quarter-2019 was 266,387 . This was 46% from the 182,465 seen in Q2 and in Q4-2018 practically twofold 138,328 was seen.

Figure 3 shows the most targeted industries in 2019. Attacks on distributed storage and record facilitating websites, financial institutions stayed more frequent, and attacks on the gaming, protection, vitality, government, and human services areas were less prominent during the 3rd quarter [3]. MarkMonitor is an online brand insurance association, verifying licensed innovation. In the 3rd quarter of 2019, the greatest focus of phishing remained Software as a service (SaaS) and webmail websites. Phishers keep on collecting credentials to these sorts of websites, using them to execute business email compromises (BEC) and to enter corporate SaaS accounts.

This survey covers the four aspects of a phishing attack: communication media, target devices, attack technique, and counter-measures as shown in Fig. 4. Human collaboration is a communication media with an application targeted by the attack. Seven types of communication media which include Email, Messenger, Blog & Forum, Voice over internet protocol, Website, Online Social Network (OSN), and Mobile platform are identified from the literature. For the selection of attack strategies, our devices play a significant role as victims interact online through physical devices. Phishing attack may target personal computers, smart devices, voices devices, and/or WiFi-smart devices which includes VOIP devices as well as mobile phone device.

Attack techniques are grouped into two categories: attack launching and data collection. For attack launching, several techniques are identified such as email spoofing, attachments, abusing social settings, URLs spoofing, website spoofing, intelligent voice reaction, collaboration in a social network, reserve social engineering, man in the middle attack, spear phishing, spoofed mobile internet browser and installed web content.

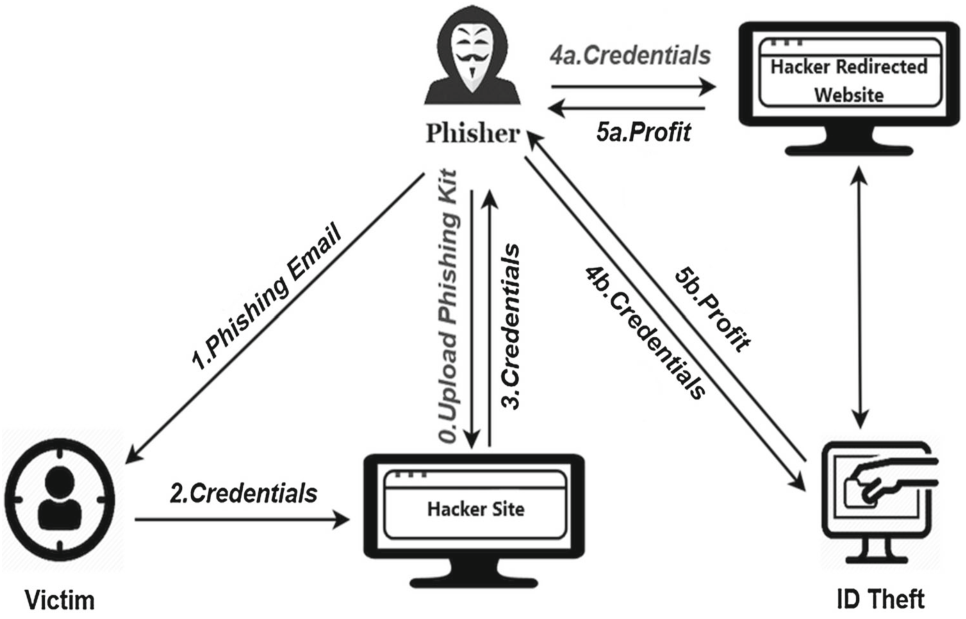

Fig. 1 Phishing Attack Diagram

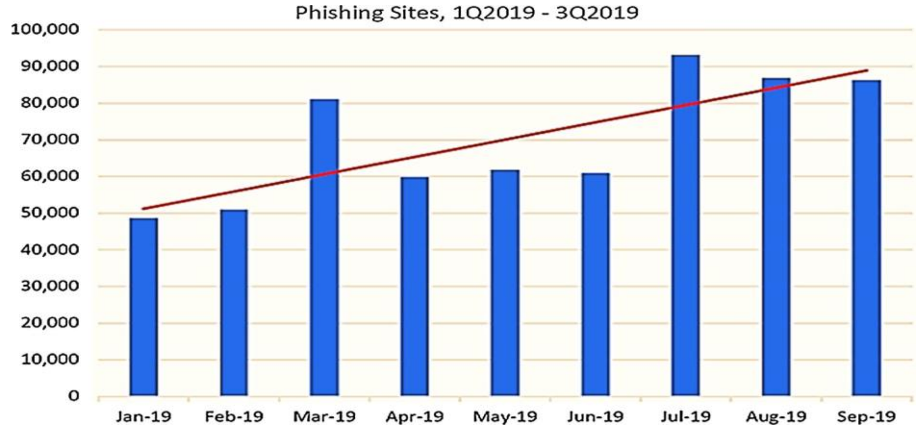

Fig. 2 Phishing report for third quarter of the year 2019

Meanwhile, for data collection during and after the victim's interaction with attacks, various data collection techniques are used . There are two types of data collec- tion techniques, one is automated data collection techniques (such as fake websites forms, key loggers, and recorded mes- sages) and the other is manual data collection techniques (such as human misdirection and social networking). Then, there are counter-measures for victim's data collected or used before and after the attack. These counter-measures are used to detect and prevent attacks. We categorized counter-measurement into four groups (1) Deep learning- based Techniques, (2) Machine learning Techniques, (3) Scenario-based Techniques, and (4) Hybrid Techniques.

To the best of our knowledge, existing literature include a limited number of surveys focusing more
on providing an overview of attack detection techniques. These surveys do not include details about all deep learn- ing, machine learning, hybrid, and scenario based techniques. Besides, these surveys lack in providing an extensive discus- sion about current and future challenges for phishing attack detection.

Keeping in sight the above limitations, this article makes the following contributions:

Provide a comprehensive and easy-to-follow survey focusing on deep learning, machine learning, hybrid learning, and scenario-based techniques for phishing attack detection.
Provide an extensive discussion on various phishing attack techniques and comparison of results reported by various studies.

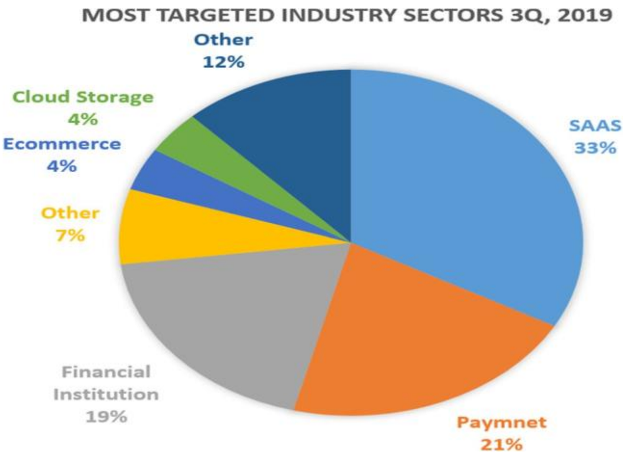

Fig. 3 Most targeted industry sectors—3rd quarter 2019

Provide an overview of current practices, challenges, and future research directions for phishing attack detection.

The study is divided into the following sections:

Sect. 1 present the introduction of phishing attacks.
Section 2 presents the literature survey focusing on deep learning, machine learning, hybrid learning, and scenario-based phish- ing attack detection techniques and presents the comparison of these techniques.
Section 3 presents a discussion on var- ious approaches used in literature.
Section 4 present the current and future challenges.
Section 5 concludes the paper with recommendations for future research.

Machine learning (ML) for phishing attack detection

ML approaches are popular for phishing websites detection and it becomes a simple classification problem. To train a machine learning model for a learning-based detection sys- tem, the data at hand must-have features that are related to phishing and legitimate website classes. Different classifiers are used to detect a phishing attack. Previous studies show that detection accuracy is high as robust ML techniques are used. Several feature selection techniques are used to reduce features. Figure 6 shows the working of the machine learning model. A batch of input data is given as input for training to the machine learning model to predict the phishing attack or legitimate traffic.

By reducing features, dataset visualization becomes more efficient and understandable. The most significant classi- fiers that were used in various studies and are found to give good phishing attack detection accuracy are C4.5, k-NN, and SVM. These classifiers are based on DTs such as C4.5, so it gives the maximum accuracy and efficiency to detect a phishing attack. To further explore the detection of phishing attacks, researchers have mentioned the limitations of their work. Many highlighted a common limitation that ensem- ble learning techniques are not used, and in some studies, feature reduction was not done. Authors in James et al. used different classifiers such as C4.5, IBK, NB, and SVM. Similarly, authors in Liew et al. used RF to distinguish phishing attacks from original web pages. Authors in Ade- bowale et al. used the Adaptive Neuro-Fuzzy Inference

LITERATURE SURVEY

This paper explores detailed literature available in prominent journals, conferences, and chapters. This paper explores rel- evant articles from Springer, IEEE, Elsevier, Wiley, review is formulated after an exhaustive search on the exist- ing literature published in the last 10 years.

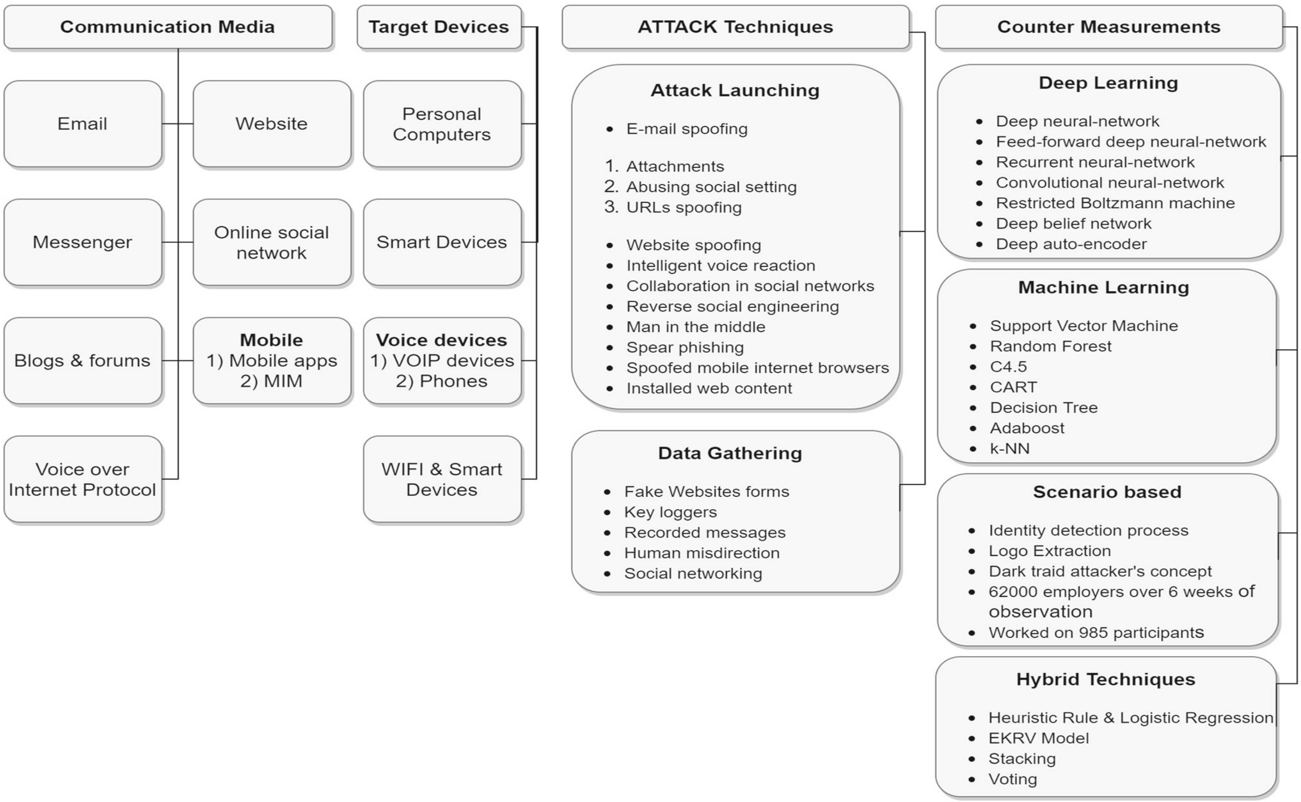

Fig. 4 Taxonomy of this survey focusing on phishing attack detection studies

A phishing attack is one of the most serious threats for any organization and in this section, we present the work done on phishing attacks in more depth along with its different types. Initially, the phishing attacks were performed on telephone networks also known as Phone Phreaking which is the reason the term “fishing” was replaced with the term “Phishing”, *ph* replaced *f* in fishing. From the reports of the anti-phishing working group (APWG) , it can be confirmed that phish- ing was discovered in 1996 when America-on-line (AOL) accounts were attacked by social engineering. Phishing turns into a danger to numerous people, especially individuals who are unaware of the dangers while being in the internet world. In light of a report created by the Federal Bureau of Inves- tigation (FBI) , from October-2013 to February-2016, a phishing attack caused severe damage of 2.3 billion dollars.

In general, users tend to overlook the URL of a website. At times, phishing tricks connected through phishing websites can be effectively prevented by seeing whether a URL is of phishing or an authentic website. For the situation where a website is suspected as a targeted phish, a client can escape from the criminal’s trap.

The conventional approaches for phishing attack detec- tion give low accuracy and can recognize only about 20% of phishing attacks. Machine learning approaches give good outcomes for phishing detection but are time-consuming even on the small-sized datasets and not scale-able. Phishing recognition by heuristics techniques gives high false-positive rates. Client mindfulness is a significant issue, for resistance against phishing attacks. Fake URLs are utilized by phisher, to catch confidential private data of the targeted victim like bank account data, personal data, username, secret password, etc.

Previous work on phishing attack detection has focused on one or more techniques to improve accuracy however, accuracy can be further improved by feature reduction and by using an ensemble model. Existing work done for phishing attack detection can be placed in four categories:

Deep learning for phishing attack detection
Machine learning for phishing attack detection
Scenario-based phishing attack detection
Hybrid learning based Phishing attack detection

Deep learning (DL) for phishing attack detection

This section describes the DL approaches-based intrusion detection systems. Recent advancements in DL approaches suggested that the classification of phishing websites using deep NN should outperform the traditional Machine Learning (ML) algorithms. However, the results of utilizing deep NN heavily depend on the setting of different learning parameters .

There exist multiple DL approaches used for cybersecurity intrusion detection , namely, (1) deep neural-network, (2) feed-forward deep neural-network, (3) recurrent neural-network, (4) convolutional neural-network, (5) restricted Boltzmann machine, (6) deep belief network, (7) deep auto-encoder. Figure 5 shows the working of deep learning models. A batch of input data is fed to the neurons and assigned some weights to predict the phishing attack or legitimate traffic.

Authors in Benavides et al. work to incorporate a combination of each chosen work and the classification. They characterize the DL calculations chosen in every arrange- ment, which yielded that the most regularly utilized are the Deep Neural Network (DNN) and Convolutional Neural Net- work (CNN) among all. Diverse DL approaches have been presented and analyzed, but there exists a research gap in the use of DL calculations in recognition of cyber-attacks.

Authors in Shie worked on the examination of dif- ferent techniques and talked about different strategies for precisely recognizing phishing attacks. Of the evaluated strategies, DL procedures that used feature extraction shows good performance because of high accuracy, while being robust.

Classifications models also depict good performance. Authors in Maurya and Jain proposed an anti-phishing structure that depends on utilizing a phishing identification model dependent on DL, at the ISP’s level to guarantee secu- rity at a vertical scale as opposed to even execution. This methodology includes a transitional security layer at ISPs and is set between various workers and end-clients. The pro- ficiency of executing this structure lies in the way that a solitary purpose of blocking can guarantee a large number of clients being protected from a specific phishing attack. The calculation overhead for phishing discovery models is restricted distinctly to ISPs and end users are granted secure assistance independent of their framework designs without highly efficient processing machines.

Authors in Subasi et al.proposed a comparison of Adaboost and multi boosting for detecting the phishing website. They used the UCI machine learning repository dataset having 11,055 instances, and 30 features. AdaBoost and multi boost are the proposed ensemble learners in this research to upgrade the presentation of phishing attack cal- culations.

Ensemble models improve the exhibition of the classifiers in terms of precision, F-measure, and ROC region. Experimental results reveal that by utilizing ensemble mod- els, it is possible to recognize phishing pages with a precision of 97.61%. Authors in Abdelhamid et al. proposed a com- parison based on model content and features. They used a dataset from PhishTank, containing around 11,000 exam- ples. They used an approach named enhanced dynamic rule induction (eDRI) and claimed that dynamic rule induction (eDRI) is the first algorithm of machine learning and DL which has been applied to an anti-phishing tool.

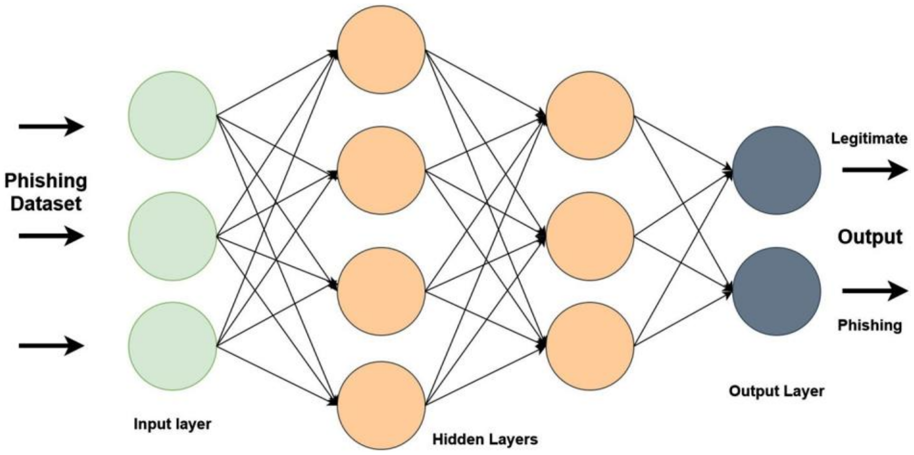

Fig. 5 Deep learning for phishing attack detection

This algorithm passes datasets with two main threshold frequencies and rules strength. The training dataset only stores “strong” features and these features become part of the rule while others are removed.

Authors in Mao et al. proposed a learning-based sys- tem to choose page design comparability used to distinguish phishing attack pages. for effective page layout features, they characterized the guidelines and build up a phishing page classifier with two conventional learning-algorithms, SVM and DT. They tested the methodology on real web- site page tests from phishtank.com and alexa.com. Authors in Jain and Gupta proposed techniques and have per- formed experiments on more than two datasets. First from Phishtank containing 1528 phishing websites, second from Openphish: which contains 613 phishing websites, third from Alexa: which contains 1600 legitimate websites, fourth from payment gateway: which contains 66 legitimate websites, and fifth from top banking website: which contains 252 legitimate websites. By applying machine-learning algorithms, they improved accuracy for phishing detection. They used RF, SVM, Neural-Networks (NN), LR, and NB. They used a feature extraction approach on the client-side.

Authors in Li et al. proposed a novel approach in which the URL is sent as input and the URL, as well as HTML related features, are extracted. After feature extrac- tion, a stacking model is used to combine classifiers. They performed experiments on different datasets: The first one was obtained from Phishtank, with 2000 web pages (1000 legitimate and 1000 phishing). The second dataset is a larger one with 49,947 web pages (30,873 legitimate, and 19,074 phishing) and was taken from Alexa. They used a support vector machine, NN, DT, RF, and combined these through stacking to achieve better accuracy. This research achieves good accuracy using different classifiers.

Some studies are limited to few classifiers and some used many classifiers, but their techniques were not effi- cient or accurate. Two datasets have been commonly used by researchers in past and these are publicly accessible from Phishtank and UCI machine learning repository. ML tech- niques have been used but without feature reduction, and some studies used only a few classifiers to compare their results.

2.1 Machine learning (ML) for phishing attack detection

ML approaches are popular for phishing websites detection and it becomes a simple classification problem. To train a machine learning model for a learning-based detection sys- tem, the data at hand must-have features that are related to phishing and legitimate website classes. Different classifiers are used to detect a phishing attack. Previous studies show that detection accuracy is high as robust ML techniques are used. Several feature selection techniques are used to reduce features. Figure 6 shows the working of the machine learning model. A batch of input data is given as input for training to the machine learning model to predict the phishing attack or legitimate traffic.
By reducing features, dataset visualization becomes more efficient and understandable. The most significant classi- fiers

that were used in various studies and are found to give good phishing attack detection accuracy are C4.5, k-NN, and SVM. These classifiers are based on DTs such as C4.5, so it gives the maximum accuracy and efficiency to detect a phishing attack. To further explore the detection of phishing attacks, researchers have mentioned the limitations of their work. Many highlighted a common limitation that ensem- ble learning techniques are not used, and in some studies, feature reduction was not done. Authors in James et al. used different classifiers such as C4.5, IBK, NB, and SVM. Similarly, authors in Liew et al. used RF to distinguish phishing attacks from original web pages. Authors in Adebowale et al. used the Adaptive Neuro-Fuzzy Inference

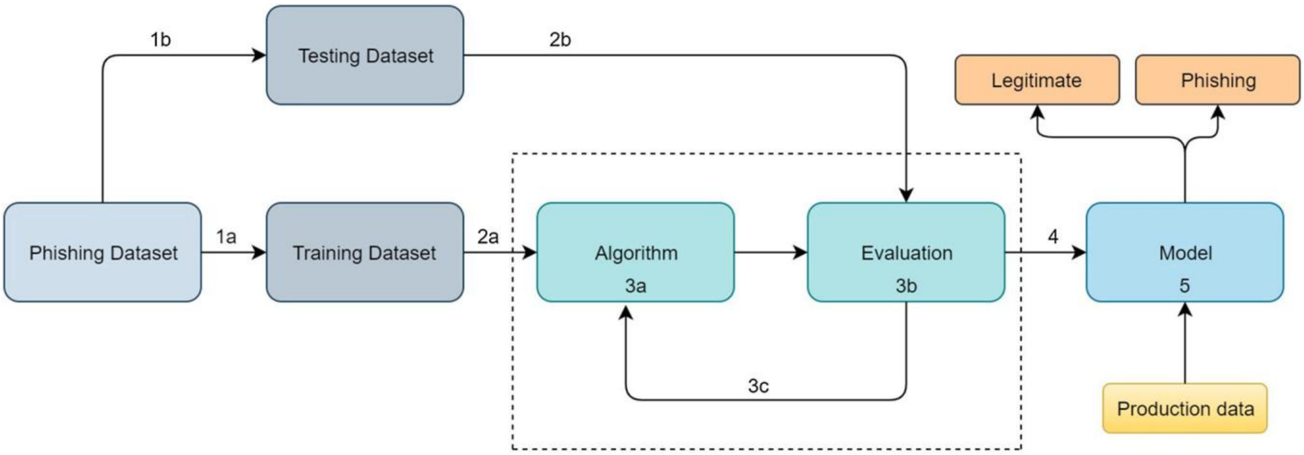

Fig. 6 Machine learning for phishing attack detection

System based robust scheme using the integrated features for phishing attack detection and protection.

Authors in Zamir et al.presented an examination of supervised learning and stacking models to recognize phishing websites. The rationale behind these experiments was to improve the classification precision through proposed features with PCA and the stacking of the most efficient classifiers. Stacking (RF, NN, stowing) outperformed other classifiers with proposed features N1 and N2. The exper- iments were performed on the phishing websites datasets. The data-set contained 32 pre-processed features with 11,055 websites. Authors in Alsariera et al.used four meta- student models: AdaBoost-Extra Tree (ABET), Bagging- Extra tree (BET), Rotation Forest-Extra Tree (RoFBET), and LogitBoost-Extra Tree (LBET), using the extra-tree base classifier. The proposed meta-algorithms were fitted for phishing website datasets, and their performance was tested. Furthermore, the proposed models beat existing ML-based models in phishing attack recognition. Thus, they suggest the appropriation of meta-algorithms when building phish- ing attack identification models.

Authors in El Aassal et al. proposed a benchmark- ing structure called PhishBench, which enables us to assess and analyze the existing features for phishing detection and completely understand indistinguishable test conditions, i.e., unified framework specification, datasets, classifiers, and performance measurements. The examinations indicated that the classification execution dropped when the proportion among phishing and authentic decreases towards 1 to 10. The decrease in execution extended from 5.9 to 42% in F1- score. Furthermore, PhishBench was likewise used to test past techniques on new and diverse datasets.

Authors in Subasi and Kremic proposed an intelli- gent phishing website identification system. They utilized unique ML models to classify websites as genuine or phish- ing. A few classification methods were used to implement an accurate and smart phishing website detecting struc- ture. ROC area, F-measure, and AUC were used to assess the performance of ML techniques. Results demonstrated that Adaboost with SVM performed best among all other classification techniques achieving the highest accuracy of 97.61%. Authors in Ali and Malebary proposed a phish- ing website detection technique utilizing Particle Swarm Optimization (PSO) based component weighting to improve the detection of phishing websites. Their proposed approach recommends using PSO to weigh different websites, effectively accomplishing higher accuracy when distinguishing phishing websites. In particular, the proposed PSO based website features weighting is utilized to separate different features in websites, given how significantly these contribute towards distinguishing the phishing from real websites. Results showed that the ML models improved with the proposed PSO-based component weighting to effectively distinguish, and monitor both phishing and real websites separately.

Authors in James et al.used datasets from Alexa and Phishtank. Their proposed approach read the URL one by one and analyze the host-name URL and path to classify into an attack or legitimate activity using four classifiers: NB, DT, KNN, and Support Vector Machine (SVM). Authors in Sub- asi et al.used Artificial Neural Network (ANN), KNN, SVM, RF, Rotation Forest, and C4.5. They discussed in detail how these classifiers are very accurate in detecting a phishing attack. They claim that the accuracy of the RF is not more than 97.26%. All other classifiers got the same accuracy as given in the study. Authors in Hutchinson et al.proposed a study on phishing website detection focusing on features selection. They used the dataset of the UCI machine learning repository that contains 11,055 URLs and 30 features and divided these features into six groups. They selected three groups and concluded that these groups are suitable options for accurate phishing attack detection.

Authors in Abdelhamid et al.creates a method called Enhanced Dynamic Rule Induction (eDRI) to detect phishing attacks. They used feature extraction, Remove replace feature selection technique (RRFST), and ANOVA to reduce fea- tures. The results show that they have the highest accuracies of 93.5% in comparison with other studies. The research proposed a feature selection technique named as Remove Replace Feature Selection Technique (RRFST). They claim that they got the phishing email dataset from the khoonji's anti-phishing website containing 47 features. The DT was used to predict the performance measures.

Authors in Tyagi et al.used a dataset from the UCI machine learning repository that contains unique 2456 URL instances, and 11,055 total number of URLs that have 6157 phishing websites and 4898 legitimate websites. They extracted 30 features of URLs and used these features to pre- dict the phishing attack. There were two possible outcomes whether the user has to be notified that the website is a phish- ing or aware user that the website is safe. They used ML algorithms such as DT, RF, Gradient Boosting (GBM), Gen- eralized Linear Model (GLM), and PCA. The authors in Chen and Chen used the SMOTE method which improves the detection coverage of the model. They trained machine learning models including bagging, RF, and XGboost. Their proposed method achieved the highest accuracy through the XGboost method. They used the dataset of Phishtank which has 24,471 phishing websites and 3850 legitimate websites. Authors in Joshi et al. used a RF algorithm as a binary classifier and reliefF algorithm for feature selection algorithm. They used the dataset from the Mendeley website which is given as input to the feature selection algorithm to select efficient features. Next, they trained a RF algo- rithm over the selected features to predict the phishing attack. Authors in Ubing et al. proposed their work on ensemble Learning. They used ensemble learning through three tech- niques that were bagging, boosting, stacking. Their dataset contains 30 features with a result column of 5126 records. The dataset is taken from UCI, which is publicly accessible. They had combined their classifiers to acquire the maximum accuracy which they got from a DT. Authors in Mao et al. used different machine learning classifiers that include SVM, DT, AdaBoost, and RF to predict the phishing attack. Authors in Sahingoz et al. created their dataset. The dataset con- tains 73,575 URLs, and out of this 36,400 legitimate URLs and 37,175 phishing URLs. As they mentioned that Phishtank doesn't give a free dataset on the web page therefore they cre- ated their dataset. They used seven classification-algorithms and natural-language-processing (NLP) based features for
phishing attack detection.

Table 1 presents the summary of ML approaches for phish- ing websites detection. Table shows that some studies provide highly efficient results for phishing attack detection.

Scenario-based phishing attack detection

In this section, we provide a comparison of scenario-based phishing attack detection used by various researchers. The comparison of scenario-based techniques to detect a phish- ing attack is shown in Table 2. Studies show that different scenarios worked with various methods and provides differ- ent outcomes.

Authors in Begum and Badugu discussed some approaches which are useful to detect a phishing attack. They performed a detailed survey of existing techniques such as Machine Learning (ML) based approaches, Non- machine Learning-based approaches, Neural Network-based approaches, and Behavior-based detection approaches for phishing attack detection. Authors in Yasin et al.con- solidated various studies that researchers have used to clarify different exercises of social specialists.

Moreover, they pro- posed that a higher comprehension of the social engineering attack scenarios would be possible utilizing topical and game-based investigation techniques. The proposed strat- egy for interpreting social engineering attack scenario is one such endeavor to empower people to comprehend gen- eral attack scenarios. Even though the underlying outcomes have demonstrated neutral outcomes, the hypothetically pre- dictable system of this strategy despite everything, merits future augmentation and re-performance.

Authors in Fatima et al.presented PhishI as a pre- cise way to deal with structure genuine games for security training. They characterize a game structure system that incorporates the group of information on social networking, that needs authoritative players. They used stick phishing as a guide to show how the proposed approach functions, and afterward assessed the learning impacts of the produced game dependent on observational information gathered from the student's movement. In the PhishI game, members are needed to trade phishing messages and have the option to remark on the viability of the attack scenario. Results demon- strated that student's attention to spear-phishing chances is improved and that the protection from the first potential attack is upgraded. Moreover, the game demonstrated a beneficial outcome on members' comprehension of extreme online data and information disclosure.

Authors in Chiew et al. concentrated phishing attacks in detail through their features of the medium and vec- tor which they live in and their specialized methodologies. Besides, they accept this information will assist the overall population by taking preparatory and preventive activities against these phishing attacks and the policies to execute approaches to check any further misuse by the phishers. Rely- ing just on client instruction as a preventive measure in a phishing attack is not sufficient. Their survey shows that the improvement of clever frameworks to counter these specialized methodologies is required, as such countermeasures will

Authors	Classification method	Feature selection method	Accuracy (%)
James et al.	J48, JBK, SVM, NB	–	89.75
Abdelhamid et al.	eDRI	–	93.5
Mao et al.	SVM, RF, DT, AB	–	97.31
Jain and Gupta	–	Feature extraction	99.09
Hota et al.	CART, C4.5	RRFST	99.11
Ubing et al.	EL	–	95.4
Chen and Chen	ELM, SVM, LR, C$.5, LC-ELM, KNN, XGB	ANOVA	99.2
Table 1 MLapproaches for phishing websites detection			
Authors	Scenarios	Method	Accuracy
Yao et al.	Identity detection processs	Logo extraction	98.3%
Curtis et al.	Dark traid attacker's concept	Dark traid	–
Williams et al.	62,000 employers over 6 weeks of observation	Theoretical approaches	–
Parsons et al.	Worked on 985 participants	ANOVA	–
Table 2 Comparison of scenario based studies			
Authors	Classification method	Feature selection method	Accuracy (%)
Subasi et al.	ANN, KNN, RF, SVM, C4.5, RF	–	97.36
Tyagi et al.	DT, RF, GBM	PCA	98.4
Mao et al.	SVM, RF, DT, AB	–	97.31

Jagadeesan et al.	RF, SVM	–	95.11
Joshi et al.	RF, RA	RA	97.63
Sahingoz et al.	SVM, DT, RF,	NLP	97.98

Table 3 Comparison of scenario based studies

Authors in Yao et al. used the logo extraction method by using the identity detection process to detect phishing. Two non-overlapping datasets were made from a sum of 726 pages. Phishing pages are from the PhishTank website, and the legitimate website pages are from the Alexa website as they limited their work by not using the DL technique. The authors gave the concept of dark triad attackers. Phishing exertion and execution, and end-users' arrangement of emails are the theoretical approach of the dark triad method. They had limited their work as end-client members may have been hyper-mindful of potential duplicity and in this way progres- sively careful in their ratings of each email than they would be in their normal workplace. Authors in Williams et al. uses a mixed approach to detect a phishing attack. They used ensemble learning to investigate 62, 000 instances over a six-week time frame to detect phishing messages, called spear phishing. As they had a drawback of just taking infor- mation from two organizations, employee observations and encounters are probably going to be affected by a scope of components that might be explicit to the association considered.

Authors in Parsons et al. used the method of ANOVA. In a scenario-based phishing study, they took a total of 985 participants completed to play a role. Two-way repeated- measures analysis of variance (ANOVA) was led to survey the impact of email authenticity and that impact was focused on the study. This investigation included only one phishing and one certifiable email with one of the standards and did not test the impact of numerous standards inside an email. Following are the comparison of specific classifier known as RF which is the most used algorithm by the researchers.

Table 3 provides a comparison of RF classifiers with different datasets and different approaches. Some studies reduced features without creating a lot of impact on accu- racy and the remaining studies focused on accuracy. Authors in Subasi et al.used different classifiers to detect phish- ing attacks and they achieved an accuracy of 97.36% by RF algorithm.

Authors in Tyagi et al.used 30 features to detect the attack by RF. They used other classifiers as well but their result on RF was better than other classifiers. Simi- larly, authors in Mao et al. collected the dataset of 49 phishing websites from *PhishinTank.com*. They used four learning classifiers to detect phishing attacks and concluded that the RF classifiers are much better than others. Authors in Jagadeesan et al. used two datasets one from UCI Machine Learning Repository having 30 features and one target class, containing 2456 instances of phishing and non- phishing URLs. The second dataset comprises of 1353 URLs with 10 features, grouped into 3 classifications: phishing, non-phishing and suspicious. They concluded that RF pro- vides better accuracy than that of support vector machine. Authors in Joshi et al.used the dataset from Mendeley website which is publicly accessible. The dataset contains 5000 legitimate and 5000 phishing records. Authors in Sahin- goz et al.used Ebbu2017 Phishing Dataset containing 73,575 URLs in which 36,400 are legitimate URLs and 37,175 are phishing URLs. They proposed seven different classification algorithms including Natural Language Pro- cessing (NLP) based features. They actually used a dataset which is not used commonly for detecting phishing attack.

Hybrid learning (HL) based phishing attack detection

In this section, we present the comparison of HL models which are used by state-of-the-art studies as shown in Tables 4 and 5 The studies show how the accuracies got improved by ensemble and HL techniques.

Authors in Kumar et al.separated some irrelevant features from the content and pictures and applied SVM as a binary classifier. They group the real and phished mes- sages with strategies like Text parsing, word tokenization, and stop word evacuation. The authors in Jain et al.uti- lized TF-IDF to locate the most significant features of the website to be used in the search question, yet it has been well adjusted to improve execution. The proposed approach has been discovered to be more accurate for their methodology against existing techniques utilizing the traditional TF-IDF approach.

Authors in Adebowale et al. proposed a hybrid approach comprising Search and Heuristic Rule and Logis- tic Regression (SHLR) for efficient phishing attack detection. Authors proposed three steps approach: (1) the most of web-site shown in the result of a search query is legal if the web page domain matches the domain name of the web- sites retrieved in results against the query, (2) the heuristic rules defined by the character features (3) an ML model to predict the web page to be either a legal web page or a phish- ing attack. Authors in Patil et al.used LR, DT, and RF techniques to detect a phishing attack, and they believe the RF is a much-improved way to detect the attack. The draw- back of this system is detecting some minimal false-positive and false-negative results. Authors in Niranjan et al. used the UCI dataset on phishing containing 6157 legiti- mate and 4898 phishing instances out of a total of 11,055 instances. The

EKRV model was used that involves a combi- nation of KNN and random committee techniques. Authors in Chiew et al.used two datasets one from 5000 phish- ing web-pages based on URLs from PhishTank and second OpenPhish. Another 5000 legitimate web-pages were based on URLs from Alexa and the Common Crawl5 archive. They used Hybrid Ensemble Strategy. Authors in Pandey et al. used a dataset from the Website phishing dataset, available online in a repository of the University of California. This dataset has 10 features and 1353 instances. They trained an RF-SVM hybrid model that achieved an accuracy of 94%.

Authors in Niranjan et al.proposed an ensemble tech- nique through the voting and stacking method. They selected the UCI ML phishing dataset and take only 23 features out of 30 features for further attack detection. Out of a total of 11,055 instances, the dataset has 6157 legitimate and 4898 phishing instances. They used the EKRV model to predict the phishing attack. Authors in Patil et al.proposed a hybrid solution that uses three approaches: blacklist and whitelist, heuristics, and visual similarity. The proposed methodology monitors all traffic on the end-user system and compares each URL with the white list of trusted domains. The web- site analyzes various details for features. The three outcomes are suspicious websites, phishing websites, and legitimate websites. The ML classifier is used to collect data and to generate a score. If the score is greater than the threshold, then they marked the URL as a phishing attack and immediately blocked it. They used LR, DT, and RF to predict the accuracy of their test websites.

Authors in Jagadeesan et al.utilized RF and SVM to detect phishing attacks. They used two types of datasets the first one is from the UCI machine learning repository which has 30 features. This dataset consists of 2456 entries of phish- ing and non-phishing URLs. The second dataset consists of 1353 URLs which has 10 features and three categories:

Phishing, non-Phishing, and suspicious. Authors in Pandey et al.used the dataset of a repository of the University of California. The dataset has 10 features and 1353 instances. They trained a hybrid model comprising RF and SVM which they utilize to predict the accuracy.

DISCUSSION

Phishing is a deceitful attempt to obtain sensitive data using social networking approaches, for example, usernames and passwords in an endeavor to deceive website users and get- ting their sensitive credentials [24]. Phishers prey on human emotion and the urge to follow instructions in a flow. Phish- ing is so omnipresent in the internet world that it has become a constant threat. In phishing, the biggest challenge is that the attackers are continuously devising new approaches to deceive clients such that they fall prey to their phishing traps. A comparative study of previous works using different approaches is discussed in the above section with details. Machine learning based approaches, deep learning based approaches, scenario-based approaches, and hybrid tech- niques are deployed in past to tackle this problem. A detailed comparative analysis revealed that machine learning methods are the most frequently used and effective methods to detect a phishing attack. Different classification methods such as SVM, RF, ANN, C4.5, k-NN, DT have been used. Techniques with feature reduction give better performance. Classifica- tion is done through ELM, SVM, LR, C4.5, LC-ELM, kNN, XGB, and feature selection with ANOVA detected phish- ing attack with 99.2% accuracy, which is highest among all methods proposed so far but with trade-offs in terms of computational cost.

The RF method gives the best performance with the high- est accuracy among any other classification methods on different datasets. Several studies proved that more than 95% attack detection accuracy can be achieved using a RF classification method. UCI machine learning dataset is the common dataset that has been used by researchers for phishing attack detection in past.

In various studies, the researchers also created a scenario- based environment to detect phishing attacks but these solutions are only applicable for a particular environment. Individual users in each organization exhibit different behaviors and individuals in the organization are sometimes aware of the scenarios. The hybrid learning approach is another way to detect phishing attacks as it occasionally gave better accuracy than that of a RF. Researchers are of the view that some ensemble models can further improve performance.

Nowadays phishing attacks defense is probably consid- ered a hard job by system security experts. With low false positives, a feasible detection system should be there to iden- tify phishing attacks. The defense approaches talked about so far are based on machine learning and deep learning algorithms. Besides having high computational costs, these methods have high false-positive rates; however, better at distinguishing phishing attacks. The machine learning techniques provide the best results when compared with other different approaches. The most effective defense for phishing attacks is an educated and well aware employee. But still, people are people with their built features of curiosity. They have a thirst to explore and know more. To mitigate the risks of falling victim to phishing tricks, organizations should try to keep employees away from their inherent core processes and make them develop a mindset that will abstain from clicking suspicious links and webpages.

Current practices and future challenges

A phishing attack is still considered a fascinating form of attack to lure a novice internet user to pass his/her private confidential data to the attackers. There are different mea- sures available, yet at whatever point a solution is proposed toovercome these attacks, attackers consider the vulnerabilities of that solution to continue with their attacks. Several solu- tions to control phishing attacks have been proposed in past. A recent increase in the number of phishing attacks linked to COVID-19 performed between March 1 and March 23, 2020, and attacks performed on online collaboration tools (ZOOM, Microsoft Teams, etc.) has led researchers to pay more attention in this research domain. Most of the working be it at government or the corporate level, educational activ- ities, businesses, as well as non-commercial activities, have switched online from the traditional on-premises approach. More users are relying on the web to perform their rou- tine work. This has increased the importance of having a comprehensive phishing attack detection solution with bet- ter accuracy and better response time .

The conventional approaches for phishing attack detec- tion are not accurate and can recognize only about 20% of phishing attacks. ML approaches give better results but with scalability trade-off and time-consuming even on the small-sized datasets. Phishing detection by heuristics techniques gives high false-positive rates. User cautiousness is a key requirement to prevent phishing attacks. Besides educating the client regarding safe browsing, some changes can be done in the user interfaces such as giving dynamic warnings and consequently identifying malicious emails. As the classified resources are accessible to the IoT gadgets, but their security architectures and features are not mature so far which makes them an exceptionally obvious target for the attackers.

Phishing is a door for all kinds of malware and ran- somware. Malware attacks on organizations use ransomware and ransomware operators demand heavy amount as ran- som in exchange for not disclosing stolen data which is a recent trend in 2020. Phishing scams in 2020 are deliberately impersonating COVID-19 and healthcare-related organiza- tions and individuals by exploiting the unprepared users. It is better to safeguard doors at our ends and be proactive in defense rather than thinking about reactive strategies to com- bat once a phishing attack has happened.

Fake websites with phishing appear to be original but it is hard to identify as attackers imitate the appearance and functionality of real websites. Prevention is better than cure so there is a need for anti-phishing frameworks or plug-ins with web browsers. These plug-ins or frameworks may per- form content filtering and identify as well as block suspected phishing websites to proceed further. An automated reporting feature can be added that can report phishing attacks to the organization from the user's end such as a bank, government organization, etc. The time lost on remediation after a phishing attack can have a damaging impact on the productivity and profitability of businesses. In the current scenario, organizations need to provide their employees with awareness and feasible solutions to detect and report phishing attacks proactively and promptly before it causes any harm.

In the future, an all-inclusive phishing attack detection solution can be designed to identify, report, and block malicious web websites without the user's involvement. If a website is asking for login credentials or sensitive information, a framework or smart web plug-in solution should be responsible to ensure the website is legitimate and inform the owner (organization, business, etc.) beforehand. Web pages health checking during user browsing has become a need of the time and a scalable, as well as a robust solution, is needed.

CONCLUSION

This survey enables researchers to comprehend the various methods, challenges, and trends for phishing attack detection. Nowadays, prevention from phishing attacks is considered a tough job in the system security domain. An efficient detection system ought to have the option to identify phishing attacks with low false positives. The protection strategies talked about in this paper are data mining and heuristics, ML, and deep learning algorithms. With high computational expenses, heuristic and data mining methods have high FP rates, however better at distinguishing phishing attacks. The ML procedures give the best outcomes when contrasted with different strategies. A portion of the ML procedures can identify TP up to 99%. As malicious URLs are created every other day and the attackers are using techniques to fool users and modify the URLs to attack. Nowadays deep learning and machine learning methods are used to detect a phish- ing attack. classification methods such as RF, SVM, C4.5, DT, PCA, k-NN are also common. These methods are most useful and effective for detecting the phishing attack. Future research can be done for a more scalable and robust method including the smart plugin solutions to tag/label if the website is legitimate or leading towards a phishing attack.

REFERENCES

[1]. (2016). Apwg trend report. http://docs.apwg.org/reports/apwg_ trends_report_q4_2016.pdf. Accessed from 20 July 2020

[2]. (2018) Phishing activity trends report. http://docs.apwg.org/ reports/apwg_trends_report_q2_2018.pdf. Accessed from 20 July 2020

[3]. (2019) Apwg trend report. https://docs.apwg.org/reports/apwg_ trends_report_q3_2019.pdf. Accessed from 20 July 2020

[4]. (2019) Fbi warns of dramatic increase in business e-mail compromise (bec) schemes—fbi. https://www.fbi.gov/contact- us/field-offices/memphis/news/press-releases/fbi-warns-of- dramatic-increase-in-business-e-mail-compromise-bec-schemes. Accessed from 20 July

[5]. (2019) What is phishing? https://www.phishing.org/what-is- phishing. Accessed from 20 July 2020

[6]. (2020) Coronavirus-related spear phishing attacks see 667% increase. https://www.securitymagazine.com/articles/92157- coronavirus-related-spear-phishing-attacks-see-667-increase-in- march-2020. Accessed from 20 July 2020

[7]. (2020) Cost of black market phishing kits soars 149% in 2019. https://www.infosecurity-magazine.com/news/black-phishing-

[8]. kits/. Accessed from 20 July 2020

[9]. (2020) Recent phishing attacks. https://www.infosec.gov.hk/

[10]. english/anti/recent.html. Accessed from 20 July 2020

[11]. Abdelhamid, N., Thabtah, F., Abdel-jaber, H. (2017). Phishing detection: A recent intelligent machine learning comparison based on models content and features. In *2017 IEEE international con- ference on intelligence and security informatics (ISI)* (pp. 72–77). IEEE.

[12]. Adebowale, M. A., Lwin, K. T., Sanchez, E., & Hossain, M. A. (2019). Intelligent web-phishing detection and protection scheme using integrated features of images, frames and text. *Expert Systems with Applications*, *115*, 300–313.

[13]. Aleroud, A., & Zhou, L. (2017). Phishing environments, tech- niques, and countermeasures: A survey. *Computers and Security*, *68*, 160–196.

[14]. Ali, W., & Malebary, S. (2020). Particle swarm optimization-based feature weighting for improving intelligent phishing website detec- tion. *IEEE Access*, *8*, 116766–116780.

[15]. Alsariera, Y. A., Adeyemo, V. E., Balogun, A. O., & Alazzawi,

A. K. (2020). Ai meta-learners and extra-trees algorithm for the detection of phishing websites. *IEEE Access*, *8*, 142532–142542.

[16]. Benavides, E., Fuertes, W., Sanchez, S., & Sanchez, M. (2020). Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review. In *Develop- ments and advances in defense and security* (pp. 51–64). Springer.

[17]. Cabaj, K., Domingos, D., Kotulski, Z., & Respício, A. (2018). Cybersecurity education: Evolution of the discipline and analysis of master programs. *Computers and Security*, *75*, 24–35.

[18]. Chen, Y. H., & Chen, J. L. (2019). Ai@ ntiphish—machine learn- ing mechanisms for cyber-phishing attack. *IEICE Transactions on Information and Systems*, *102*(5), 878–887.

[19]. Chiew, K. L., Yong, K. S. C., & Tan, C. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications*, *106*, 1–20.

[20]. Chiew, K. L., Tan, C. L., Wong, K., Yong, K. S., & Tiong, W.

[21]. K. (2019). A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. *Information Sciences*, *484*, 153–166.

[22]. Conklin, W. A., Cline, R. E., & Roosa, T. (2014). Re-engineering cybersecurity education in the us: An analysis of the critical factors. In *2014 47th Hawaii international conference on system sciences* (pp. 2006–2014). IEEE.

[23]. Curtis, S. R., Rajivan, P., Jones, D. N., & Gonzalez, C. (2018). Phishing attempts among the dark triad: Patterns of attack and vul- nerability. *Computers in Human Behavior*, *87*, 174–182.

[24]. El Aassal, A., Baki, S., Das, A., & Verma, R. M. (2020). An in- depth benchmarking and evaluation of phishing detection research for security needs. *IEEE Access*, *8*, 22170–22192.

[25]. Fatima, R., Yasin, A., Liu, L., & Wang, J. (2019). How persuasive is a phishing email? A phishing game for phishing awareness. *Journal of Computer Security*, *27*(6), 581–612.

[26]. Feng, Q., Tseng, K. K., Pan, J. S., Cheng, P., & Chen, C. (2011). New anti-phishing method with two types of passwords in openid system. In *2011 Fifth international conference on genetic and evo- lutionary computing* (pp. 69–72). IEEE.

[27]. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Informa- tion Security and Applications*, *50*, 102419.

[28]. Forecast. (2017). Global fraud and cybercrime forecast. https:// rsa.com/en-us/blog/2016-12/2017-global-fraud-cybercrime-fore cast. Accessed from 20 July 2020

[29]. Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: State of the art and future chal- lenges. *Neural Computing and Applications*, *28*(12), 3629–3654.

[30]. Gupta, B. B., Arachchilage, N. A., & Psannis, K. E. (2018). Defend- ing against phishing attacks: Taxonomy of methods, current issues and future directions. *Telecommunication Systems*, *67*(2), 247–

[31]. 267.

[32]. Hota, H., Shrivas, A., & Hota, R. (2018). An ensemble model for detecting phishing attack with proposed

remove-replace feature selection technique. *Procedia Computer Science*, *132*, 900–907.

[33]. Hulten, G. J., Rehfuss, P. S., Rounthwaite, R., Goodman, J. T., Seshadrinathan, G., Penta, A. P., Mishra, M., Deyo, R. C., Haber, E. J., & Snelling, D. A. W. et al. (2014). *Finding phishing sites*. US Patent 8,839,418.

[34]. Hutchinson, S., Zhang, Z., & Liu, Q. (2018). Detecting phish- ing websites with random forest. In *International conference on machine learning and intelligent communications* (pp. 470–479). Springer.

[35]. Iwendi, C., Jalil, Z., Javed, A. R., Reddy, T., Kaluri, R., Srivastava, G., et al. (2020). Keysplitwatermark: Zero watermarking algo- rithm for software protection against cyber-attacks. *IEEE Access*, *8*, 72650–72660.

[36]. Parekh, S., Parikh, D., Kotak, S., & Sankhe, S. (2018). A new method for detection of phishing websites: Url detection. In *2018 Second international conference on inventive communication and computational technologies (ICICCT)* (pp. 949–952). IEEE.

[37]. Parsons, K., Butavicius, M., Delfabbro, P., & Lillie, M. (2019). Predicting susceptibility to social influence in phishing emails. *International Journal of Human-Computer Studies*, *128*, 17–26.

[38]. Patil, V., Thakkar, P., Shah, C., Bhat, T., & Godse, S. (2018). Detec- tion and prevention of phishing websites using machine learning approach. In *2018 Fourth international conference on comput- ing communication control and automation (ICCUBEA)* (pp. 1–5). IEEE.

[39]. Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from urls. *Expert Systems with Applications*, *117*, 345–357.

[40]. Shie, E. W. S. (2020). *Critical analysis of current research aimed at improving detection of phishing attacks*. Selected computing research papers, p. 45.

[41]. Subasi, A., & Kremic, E. (2020). Comparison of adaboost with multiboosting for phishing website detection. *Procedia Computer Science*, *168*, 272–278.

[42]. Subasi, A., Molah, E., Almkallawi, F., & Chaudhery, T. J. (2017). Intelligent phishing website detection using random forest classi- fier. In *2017 International conference on electrical and computing technologies and applications (ICECTA)* (pp. 1–5). IEEE.

[43]. Tyagi, I., Shad, J., Sharma, S., Gaur, S., & Kaur, G. (2018). A novel machine learning approach to detect phishing websites. In *2018 5th International conference on signal processing and integrated networks (SPIN)* (pp. 425–430). IEEE.

[44]. Ubing, A. A., Jasmi, S. K. B., Abdullah, A., Jhanjhi, N., & Supra- maniam, M. (2019). Phishing website detection: An improved accuracy through feature selection and ensemble learning. *Interna- tional Journal of Advanced Computer Science and Applications*, *10*(1), 252–257.

[45]. Volkamer, M., Renaud, K., Reinheimer, B., & Kunz, A. (2017). User experiences of torpedo: Tooltip-powered phishing email detection. *Computers and Security*, *71*, 100–113.

[46]. Vrbancˇicˇ, G., Fister Jr, I., & Podgorelec, V. (2018). Swarm intel- ligence approaches for parameter setting of deep learning neural network: Case study on phishing websites classification. In *Pro- ceedings of the 8th international conference on web intelligence, mining and semantics* (pp. 1–8).

[47]. Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring sus- ceptibility to phishing in the workplace. *International Journal of Human-Computer Studies*, *120*, 1–13.

[48]. Yao, W., Ding Y., & Li, X. (2018). Logophish: A new two- dimensional code phishing attack detection method. In *2018 IEEE international conference on parallel and distributed pro- cessing with applications, ubiquitous computing and commu- nications, big data and cloud computing, social computing and networking, sustainable computing and communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* (pp. 231–236). IEEE.

[49]. Yasin, A., Fatima, R., Liu, L., Yasin, A., & Wang, J. (2019). Contemplating social engineering studies and attack scenarios: A review study. *Security and Privacy*, *2*(4), e73.

[50]. Zamir, A., Khan, H. U., Iqbal, T., Yousaf, N., Aslam, F., Anjum, A., et al. (2020). Phishing web site detection using diverse machine learning algorithms. *The Electronic Library*.

[51]. Jagadeesan, S., Chaturvedi, A., & Kumar, S. (2018). Url phishing analysis using random forest. *International Journal of Pure and Applied Mathematics*, *118*(20), 4159–4163.

[52]. Jain, A. K., & Gupta, B. B. (2018). Towards detection of phishing websites on client-side using machine learning based approach. *Telecommunication Systems*, *68*(4), 687–700.

[53]. Jain, A. K., Parashar, S., Katare, P., & Sharma, I. (2020). Phish- skape: A content based approach to escape phishing attacks. *Procedia Computer Science*, *171*, 1102–1109.

[54]. James, J., Sandhya, L., & Thomas, C. (2013). Detection of phish- ing urls using machine learning techniques. In *2013 International conference on control communication and computing (ICCC)* (pp. 304–309). IEEE.

[55]. Javed, A. R., Jalil, Z., Moqurrab, S. A., Abbas, S., & Liu, X. (2020). Ensemble adaboost classifier for accurate and fast detection of botnet attacks in connected vehicles. *Transactions on Emerging Telecommunications Technologies*.