Malaya Journal of Matematik, Vol. 9, No. 1, 319-328, 2021

https://doi.org/10.26637/MJM0901/0054

A survey on anti-phishing techniques: From conventional methods to machine learning

M. Noushad Rahim[1]* and K.P. Mohamed Basheer[2]

Abstract

Phishing is a deceptive technique to steal confidential information like user credentials and bank account details of web users. Employing technical and social engineering skills phishers make huge financial loss to web users and large organizations alike, and it has become one of the serious cybercrime today. This paper discusses different types of phishing techniques, their impacts, common indicators of phishing attacks, and analyses various anti-phishing solutions from conventional methods implementing blacklist, white list, heuristics, fuzzy logic, visual similarity, etc. to machine learning methods. The study provides gap analysis of conventional anti-phishing techniques, and points out the challenges facing machine learning based approaches including proper feature selection, diversity in data sets, imbalanced scenarios, and differences in evaluation metrics. This investigation outlines the need for serious researches in this area since there is no foolproof solution to phishing as phishers change their tactics very often to bypass anti-phishing detection systems.

Keywords

Cyber security, phishing, social engineering, feature selection, machine learning, deep learning.

AMS Subject Classification

97R40, 68T10, 60H30, 62H30, 03B52.

[1,2] *Department of Computer Science, Sullamussalam Science College, Areekode-676541, Kerala, India.*

***Corresponding author**: [1] mannayilnoushadrahim@gmail.com

Article History: Received **12** November **2020**; Accepted **23** January **2021**

Contents

1. Introduction

Phishing is an identity theft mechanism used to steal web user's personal and financial account credentials by employing social engineering and technical subterfuge skills. The phishers will lead web users to counterfeit web sites and they are being tricked to provide these details. Phishers normally send a baked email which may contain a URL that can lead the user to a spoofed website which may look very similar to an authentic site. Users may unknowingly enter their confidential details like username, password or credit card information into the page and cyber criminals may use these details to login to user's account. It can be used to blackmail the users or make financial damage to them.

1.1 Types of Phishing

Social engineering, link manipulation, spear phishing, clone phishing, voice phishing, etc. are used as phishing mechanisms. Spoofed websites, forged email messages, and phone calls are crafted to trap users into giving information about their credit card details or login details. Monetary gain in large amount is the main intention behind phishing attacks [21]. Some of the phishing types are listed in Figure. 1.

In email phishing, by spoofing the identity of an original enterprise, fake emails are sent to victims for stealing their private and sensitive information to perform fraud transactions. Deceptive phishing, spear phishing, whaling, etc. are

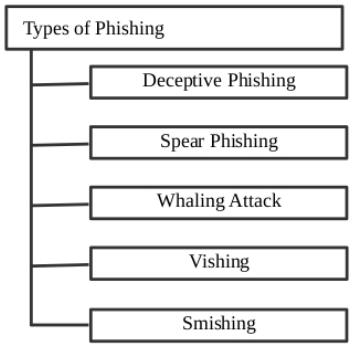

Figure 1. Types of phishing

examples of email phishing.

Deceptive phishing aims to trick victims to click a malicious attachment or URL of a spoofed web site to enter their sensitive information. Messages are sent to victims to verify account information. The reason for the loss of data can be claimed to be system failure which necessitates the victim to compulsorily re-enter all the personal details [14]. In deceptive phishing victims may not be targeted individually.

Spear Phishing is a targeted phishing attack. Spear phishing is a form of cyber attack attempting to infiltrate a system or organization for cybercrime or espionage purposes. Cyber attackers find inside information specifically relevant to users and craft fake email messages, usually impersonating well known companies, trusted relationships, or contexts. The users must take action for the attack to succeed. By clicking a link in an email message, for example, a malicious software could be installed on their system, or they might be prompted to provide personal information, such as a credit card number, username or password [7]. Social media sites like LinkedIn are the main platforms for spear phishing.

Whaling attack, which is similar to spear phishing where phisher targets a high profile individual like CEO of an organization. The attacker starts profiling the victim and steals his login details. He will use this compromised login details and masquerade as the CEO to target other members of the organization to steal their sensitive information for wire transfer [6, 24].

Vishing attack uses phone calls to steal sensitive information. It can be done by setting up a Voice over Internet Protocol(VoIP) server. Here the fraudster persuades the victim under pretences of a customer services executive to help him to do a financial transaction. The victims may lose their sensitive information and possibly his cash [15].

Smishing is an attack targeted to mobile devices in which the attacker sends text messages containing malicious links, phone numbers, or email ids to the victim and the attacker aims to steal sensitive user data like bank account details, passwords, user credentials, credit card details, etc. Through this message, the attacker prompts the user to click on the link or contact the phone number or email id provided in the SMS [17].

Pharming attack can be carried out by modifying local host files or by DNS poisoning. The attacker sends emails containing an attachment to victims which can be a code to modify the 'hosts' file. The IP address in the hosts file is modified to direct to a spoofed web site. In DNS poisoning domain name system table is modified to direct the users to a spoofed web site [24].

1.2 Functions of Phishing

The first phase in phishing is to plan the attack. The attacker will identify the target person, company or the mass to send a phishing email. Once the target is identified a spoofed web site is created and the email address is forged. Then the bait is designed as malicious email in the disguise of a genuine one. Users are unknowingly trapped and their credentials are stolen. Later these stolen credentials are used for wire-transfer or other illegal activities [21]. Figure. 2 illustrates the sequence and functions of phishing attack.

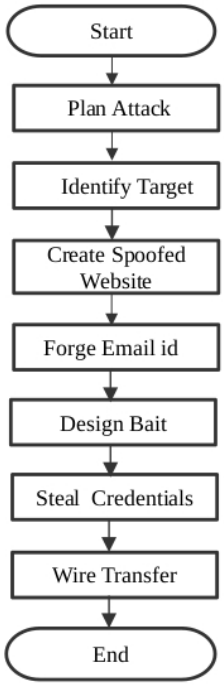

Figure 2. Functions of Phishing

Spoofed websites are used to steal the user's sensitive information. These sites ask for information like user names and passwords, bank account numbers, social security numbers, credit card numbers, PINs (Personal Identification Numbers), mother's maiden name, birthday, etc. Cybercriminals may use a compromised website or free web hosting sites for this purpose. Initially the page may not contain any phishing content so that they can escape anti-spam filter scams. Later the page is modified to include the phishing elements [9]. Some phishing sites even change domains very quickly to es-

cape from blacklisting. Phishers are remotely trying to logon to a targeted website using even harvested credentials [25]. Google Chrome's Safe Browsing detects thousands of new unsafe websites every day. Most of them are legitimate but compromised web sites used for phishing [12].

2. Motivation and Impact of Phishing

The main intention of a phishing attack is the monetary gain in general. Industrial espionage, malware distribution, identity theft, etc. are also some of the main targets of phishing attacks. It can also be used for destruction, revenge, ego, or even thrill. Free web site infrastructure is another attraction [19].

The impact of phishing is very huge and it involves the risks of identity theft and monetary losses. Proof Point, a leading cybersecurity company in its 2019 report says that 83% of survey respondents said they experienced phishing attacks, 49% experienced vishing or smishing, 4% USB based social engineering attacks, and 64% experienced spear phishing [10]. In its 2020 State of Phish Report it says 88% global respondents faced spear phishing in 2019. In the report another finding is that 39% of respondents do not know what phishing is, 69% do not know about ransomware, 34% do not know about malware 70% do not know about smishing, 75% do not know about vishing. About 55% of victims suffered the loss of data, 50% had a credential compromise,50% had ransomware infection and about 35% had a financial loss or wire transfer loss [26].

Phishing costs downtime hours of users, damages the reputation of organizations and even intellectual property loss. Companies normally suffer from phishing attacks. In 2019 Amazon had suffered an attack known as Amazon Prime Day Phishing Attack. The Company's customers were targeted in this attack. McAfee had a similar attack in 2018. Both attacks were carried out using a phishing kit known as 16Shop [5, 24]. In 2017 Google Doc users suffered an attack. This is done using a fake Google Doc invitation to share the document for collaboration. Once users accepted invitation they were encouraged to provide access to their email accounts [24, 27].

Chart 1. Time line of phishing attacks hosted on HTTPS websites

Anti-Phishing Working Group (APWG), a not-for-profit industry association focused on eliminating the identity theft and frauds, in its fourth-quarter report of 2019 identified webmail and Software-as-a-Service(SaaS) as the biggest target of phishing. Then comes payment systems and financial institutions. It also detected 162,155 unique phishing websites, 132,553 unique phishing reports, and 12,260 unique domain names used for phishing. Phishing targets social media and it grew every quarter of the year. Business Email Compromise (BEC) attacks are heavily used for gift, direct transfer cards, wire transfers, and payment diversion. In BEC attacks the criminal targets the employees in financial departments and trick them to wire transfer into the hacker's account. Another important finding is that 75% of all phishing websites use HTTPS protocol with SSL certificates. They use free certificates provided by the free web sites or abusing the certificates of hacked websites [3]. As per 2nd quarter, 2020 APWG's phishing activity trend report [4], the analysis is shown in Chart 1.

3. Current Anti-Phishing Solutions

Currently different strategies are used for protection from phishing. Some of these strategies are presented in Figure 3. One solution for phishing is the use of a good password manager. Instead of using the same passwords in different web sites, use strong and unique passwords in each, else hackers will try the hacked password on different sites. Use password managers and its auto-fill feature with the assumption that password managers won't autofill phishing site. But this can't be the case always since we have to use a password manager on allthe browsers we use. Also we have to trust fully the password manager companies that they won't misuse our passwords. And finally, most people are not aware of these things. Two-Factor Authentication is another solution. Along with password or PIN we can use a phone token authentication mechanism like OTP or biometric authentication. Big organizations can use powerful firewalls and gateways for email security and web security. Employee training on phishing techniques is a must in every organization to protect its confidential data [9].

Google Chrome tries to provide real-time phishing protection by checking the visited URLs against its blacklist, instead of locally saved lists in the user's computer. This happens every 30 minutes. Google says that it is not logging users visiting links as it is but an encrypted version of it to ensure that users are not being tracked. It initially checks the visited URL if it is in the locally stored blocked sites. If it is not present then it will be sent to Google's real-time server for verification. Google argues that it has a 30 percent increase in protections by warning the users [18]. Mozilla Firefox also uses a similar technique. It will first check the visited URL in the browser's downloaded list of blocked lists of websites. If it is not present in the list, it will ask Google's safe browsing service for verification.

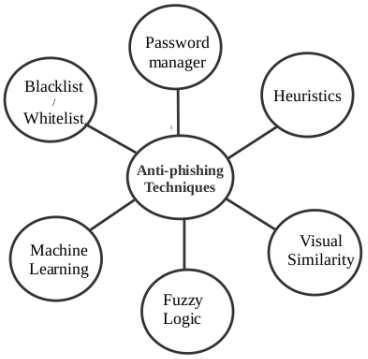

Figure 3. Anti-phishing techniques

3.1 Common Indicators of Phishing Attempts

Email header, links, attachments, grammatical errors in the content, etc. are common indicators of phishing emails. The email header can be analysed to identify the true origin of an email. This can be done using the facility provided in browsers. For example in Gmail we have 'show original' option for reading the header of an email. An attacker can forge an email address that will look very similar to an authentic email address. Links can also be used to mislead a web user. https: //mozilla.org/ is a legitimate link for mozilla.org, but it can be manipulated and another web address like https://mozi11a.org/ can be created. Here the letter l is replaced by digit 1 which a normal user won't be able to identify easily. Earlier HTTP protocol was an indicator of phishing websites. But these days phishers use HTTPS protocol for hosting their websites. Malicious attachments are very dangerous and they can contain programs written for bad purposes. PDF documents, MS Word/Excel documents, bash or .exe files, etc. can carry extremely dangerous virus programs. Another indicator of a phishing attack is badly written email content containing grammar mistakes. A professional organization will have professionals to draft emails that may not have grammatical errors in it normally. Phishing emails are found to be poorly written. Another indication is the urgency of action asked by the email for saving loss of data, cash, and credentials of the user. This is found in smishing and vishing also. A blank 'To address' and general salutations are also seen in phishing emails. Attackers do enough social engineering before carrying out the attack. They take advantage of current events like natural disasters, epidemics and health scares, major political events, economic concerns and even holidays.

To avoid being victimized to phishing, web users should be given proper awareness. We should be suspicious of unsolicited emails, text messages, and phone calls asking about personal and financial data. Employees in an organization should never reveal the internal structure of an organization to an unauthorized person. Never reveal financial or sensitive information in emails or social media. Analyze web address(URL) before logging in to any website, look for a closed padlock icon in the address bar.

3.2 Conventional Anti-Phishing Techniques

Phishers use diverse techniques to bypass the phishing detection systems and thus researchers have tried different tools and algorithms to tackle this menace. Possibilities of blacklisting, fuzzy logic, image processing, neuro-fuzzy approach, data mining, MD5 (Whois, URL), machine learning, etc. are extensively explored for the detection of phishing websites and emails.

All of them use different features of phishing websites and emails like URL length, unmatching URLs, visible links, invisible links, number of times websites visited by the user, age of the account, screenshots, WHOIS, IP Address, etc. [21]. But so far there is no complete single solution providing total immunity from phishing attacks. Since phishing can be treated as a classification problem, Machine Learning algorithms have also been explored to tackle the problem. Right feature set selection and suitable classification algorithm are the challenges to prevent the diverse phishing techniques and attacks.

In an earlier algorithm, LinkGuard [8], characteristics of links in phishing emails have been analysed. In its study it is observed that about 44% of phishing emails use faked DNS names, 42% uses doted decimal IP Addresses and 17% uses maliciously encoded hyperlinks. The algorithm extracts the DNS names of actual and visual hyperlinks. If there is a mismatch it alerts a possible phishing attack. It also alerts the user when the dotted IP address is given in actual DNS or actual or visual hyperlinks are encoded. The algorithm was implemented in the Windows XP environment and claimed 96% of detection of phishing attacks in real time.

Cao, Y et al. proposed an anti-phishing technique based on Automated Individual White List(AIWL) [6]. This is opposite to the conventional approach of a blacklist based detection system. The system maintains a white list of every user familiar Login User Interfaces (LUI). Whenever the user submits his credentials in LUI the system will verify if it is present in the white list and if not, it will notify the user of a possible phishing trap. The LUI is defined based on URL Address, page feature, and DNS-IP mapping. The page feature is represented using the hash code of web page's certificate and that of its source code. To maintain the white list upto-date and to reduce false positives Naive Bayesian classifier is used. The classifier learns the LUI properties of both legitimate and phishing sites in the training period. In real-time if an unfamiliar LUI is opened, the classifier notifies the user and if the user submits the credentials irrespective of the warning, then it will be updated in the white list. The study claims that the solution is better than existing anti-phishing tools like Web

Wallet, Spoof Guard etc. in terms of LUI authentication and anti-pharming.

Joshi, Y. et al. designed a browser plugin called PhishGuard [13]. The algorithm submits random credentials to login to the site before submitting the actual user credentials as a means to identify whether the site is phished one. The response from the site is analyzed and if the random credentials are accepted by the server the user is warned of a definite phishing attack. At the first sight this looks a good idea, but the problem is in the way the server responds. If the server always responds success then the solution works very well, but if the server response is always failure then the algorithm tries to send actual user credentials to ensure if it is a phished one. This will at the end reveal the actual user credentials to the phisher and his account will be compromised.

Prakash, P. et al. proposed a system, PhishNet [20], which enhances a blacklisting based phishing detection. From the existing blacklisted URLs of phishing websites taken from PhishTank and SpamScatter it predicted new malicious URLs. The new URLs are verified by DNS lookup. If the URLs are found existing, then by using an HTTP GET request the content of the page is extracted. The page is then undergone a similarity test against the parent page and based on the degree of resemblance, say above 90%, the generated URL is considered to be a malicious phishing site and finally added to the existing blacklist. It also uses an approximate matching data structure to identify the piece-wise similarity between generated and existing URLs. The system claims that it has very few false positives (less than 5%) and false negatives (less than 5%) taking malicious URLs from PhishTank and SpamScatter, and legitimate URLs from Yahoo and DMOZ. The system found that many predicted URLs were present in Google's blacklist, which they say is an indication of the significance of their work.

Aburrous, M. et al. have proposed a fuzzy data mining intelligent system for phishing detection especially for e-banking sites [2]. Fuzzy logic can process vaguely defined variables and can categorize such variables. This can be utilized to identify how the features of a web site can determine how much it belongs to a legitimate or phishing site. Data mining helps researchers focus on the most significant features in their data archive. Their approach utilizes fuzzy logic along with data mining algorithms to identify e-banking phishing attempts using 27 features pertaining to forged websites. These linguistic variables represent important phishing characteristics. In the fuzzification step descriptors like low, medium, and high are assigned to different characteristics like the length of the URL address, anomalous SSL certificate, spelling errors, etc. based on the degree of belongings of the values. Membership functions are applied to each phishing characteristics to classify the site as very phishy, phishy, suspecious, legitimate, and very legitimate. The system incorperated data mining classification and association algorithms like RIPPER, JRip, Prism, C4.5, etc. on different features to learn the phishing probabilities. The model used PhishTank and APWG datasets. This is a three layer system, each layer applies rules on different features like URL and domain identity, security and encryption criteria, page style and contents criteria, etc. The data mining algorithms by analysing website details help to identify significant features and generate rules for classification, fuzzy rule engine identifies the membership belongings of the features and finally categorizes them into different classes by defuzzification. The system has been implemented in Matlab. The main findings are that URL and Domain Identity are important indicators of a phishy website. Also, that based on a few characteristics it is difficult to classify whether a site is phishy or legitimate. They also indicated that feature selection is an important aspect of better prediction by the model.

Mao, J. et al. have come up with another solution called BaitAlarm [16], which analyses the visual similarity features of the suspected and the victim page. The visual layout similarity is calculated by extracting the CSS structure of the pages. This is because phishers started creating visual similar pages by replacing text contents with images. This issue can be resolved using CSS based similarity detection. Once the CSS structure of the page is extracted and converted into a normalized model, the similarity score is computed. If the score breaks certain threshold the URLs of both the pages are compared to determine if it is a phishing page. The system analysed about 300 phishing pages of Hotmail, ASB Bank, Google, etc. from PhishTank and claims a detection rate of 100% , and 0% of the false-negative rate.

Rao, R. S. et al. have discussed a combination of white list and computer vision technique to defend phishing attacks. They used SURF(Speed Up Robust Features) detector which is a computer vision tool [22]. It can extract discriminative key point features using square-shaped filters. These extracted features are used to compute the visual similarity between legitimate and suspicious web pages. In order to eliminate legitimate pages to undergo SURF processing, the system uses a white list of legitimate websites. When users open a URI it is first searched in the white list, if not found SURF algorithm is applied on the page. If the degree of similarity is less than the threshold it is added to the white list. Otherwise, the user is warned of a suspicious web page. The system can possibly detect real-time phishing attacks but can be improved using CSS similarity of the pages.

3.3 The Machine Learning Approaches

Researchers have started experimenting with Machine Learning to prevent phishing. The problem of phishing actually involves automatic grouping of websites into a predefined set of classes (phishy, legitimate) based on feature variables, and thus it can be taken to be a classification problem. Machine Learning can reveal concealed information about a new event

using a model that is built on a related dataset. The model can be set up by carefully extracting features of legitimate and phishing websites/emails from datasets like PhishTank, and can predict with higher accuracy whether a web site or an email is legitimate or phishy. To provide better result proper feature selection and classifier selection is very important. The model should be trained and tested by splitting the dataset properly. This model can be integrated into a web browser that can communicate to the end-user the outcome in real time [1].

Zhu, E. et al. has proposed an OFS-NN (Optimal Feature Selection and Neural Network) [29], amachine learning model to detect phishing websites. This is an enhancement work on neural network-based anti-phishing technique. To avoid over-fitting problem in NN models because of the selection of small influence features, the system uses a new index called Feature Validity Value(FVV Index), to select optimal features of phishing and legitimate websites. Using these features an NN classifier is constructed to detect phishing attacks. The system classifies features primarily into address bar features like IP address, length of URL, abnormal features like WHOIS information, number of hyperlinks in the page, presence of meta, script, link tags, javascript features like status bar customization, disable right click, popup windows, etc. and domain features like DNS record, website traffic, google index, etc. The algorithm initially calculates the FVV Index of all features of the URL inputs to the system and based on a threshold the relevant features are selected. The FVV index is calculated using positive and negative values. A positive value of a feature indicates that the site can be a phishy and a negative value indicates the site may be legitimate. Features that are not relevant are thus eliminated from the modeling process. The system uses a seven-layer fully connected neural network classifier based on performance analysis on the accuracy, precision, F1-Score, etc. The system also integrates blacklist and whitelist to avoid unwanted computations. The authors claim that OFS-NN reduces about 0.17s average time cost compared to an NN classifier. The system also eliminates over-fitting problems and provides 0.993 accuracy, 0.969 precision and 0.964 F1-score. They used Alexa/PhishTank datasets.

Researchers have to extract optimal features from a dataset for classifiers to provide better accuracy and precision in predictions and the researchers' expertise on the domain is a major issue in a dynamically changing environment like phishing attacks. Deep learning has been used in such a scenario by researchers to solve the problem since it can extract the optimal features from datasets automatically.

Xiao X et al. have come up with a solution based on deep learning called Convolutional Neural Network and Multi-Head Self-Attention combined approach (CNN-MHSA) [28]. This is an integrated solution of both CNN and MHSA for providing more precision than applying any one of them separately. The possibility of CNN for Natural Language Processing and its application in the classification of sentences are explored in the research. Also that CNN can learn features of URLs automatically without any human intervention. Google has proposed that MHSA can discover the inner relationship between the characters of a single sentence and it is far better than the Long Short Term Memory (LSTM) technique. This analysis is utilized in the system for analyzing the URLs and generating a weight matrix that can give different importance to different features in it. The system inputs URL string to a mature CNN model and extracts its features. Meantime the URL is also fed to MHSA to identify the relationship of the characters in the URL. Features' weights are calculated by MHSA and features are learned by the CNN. Outputs of both these subsystems are fed to the output layer to compute the classification result. CNN-MHSA claimes an accuracy of 99.84% which is an outstanding figure.

Sahingoz, O. K. et al. have proposed a real-time anti-phishing system, which uses Natural Language Processing(NLP) for feature extraction and tried seven machine learning algorithms to classify the website [23]. The system focuses on the URL analysis of phishy and legitimate websites. They have created their own dataset containing 73,575 URLs out of which there are 36400 legitimate URLs and 37175 phishing URLs. Phishers use techniques like cybersquatting, typosquatting, random characters, combined word usage, etc. for creating URLs of forged websites. In the proposed system, during data preprocessing such words are identified from the URLs using Natural Language Processing. Three different features named NLP features, Word vectors, and Hybrid features are selected from the analyzed words. NLP Features include number words in an URL, brand name check for the domain, number of keywords in a URL, random domain, etc. Word features likeverification, configuration, billing, services are extracted and converted into word vectors to use in ML algorithms. To increase the performance of the system NLP features and word vectors are combined to form a hybrid model and a feature reduction mechanism is applied to select prominent 104 features for classification. The features are fed to different machine learning algorithms and analysed their performance. The algorithms include Naive Bayes, Random Forest, kNN(n=3), Decision Trees, etc. Random forest algorithm on NLP features provided the best performance of 97.98%. The advantages of the system are language independence, huge dataset, real-time execution, feature-rich classifiers, etc.

However there are multiple questions yet to be adequately addressed in machine learning-based research on phishing detection. Most of the solutions provided behave differently in different scenarios. The claims of research papers regarding the accuracy and other metrics are not always realistic. There are real challenges to address these issues. Phishing attackers are highly proactive in learning defensive mechanisms against the attacks. Other issues include diversity in datasets used

in researches, imbalanced scenarios in phishing attacks, use of different metrics for evaluation, near real-time detection, etc. The style of attacks are changing every day and doing research on the old dataset cannot solve the problem. Even features selected from old dataset may not be able to detect the attacks.

El Aassal et al. have come up with a benchmarking and evaluation phishing detection tool called Phishbench [11]. This a benchmarking framework for researchers working on a machine learning-based phishing detection system. The researchers can test their classification algorithms systematically and compare the result on common datasets. The input module of the system can load datasets into its memory and extract features from it. Features extracted can be syntactic, semantic or pragmatic. An example of a syntactic feature can be a TLD position in a URL or format of the text of an email. Semantic feature targets the interpretation of the content of email, URL or website. It can be the presence of hidden elements in a web page, presence of @ symbol in a link, or blacklisted words in an email text.

Disabling right-click on a page to hide source code of a page, details of web page registration or age of a website, etc. comes under pragmatic features. Phishbench facilitates code for extracting about 200 features. The system also has different ranking mechanisms for features selected including Information Gain(IG), Gini Index, Recursive Feature Elimination (RFE), Chi-Square Metric (Chi-2), etc. Once features are extracted, ranked, and normalized, they are fed to the classifier algorithm. PhishBench supports about 30 different classifiers including Support Vector Machines (SVM), Random Forest (RF), Decision Tree (DT), Gaussian & Multinomial Naive Bayes, Logistic Regression (LR) and K Nearest Neighbors (KNN). The system works on diverse datasets. It has been tested on legitimate and phishing datasets like Alexa, Alexa Login, PhishTank, OpenPhish, APWG (all URL datasets), Wikileaks, Enron datasets, Nazario, SpamAssasin(all email datasets), etc.

4. Gap Analysis of Existing Anti-Phishing Techniques

Even if so many researches have been done in tackling phishing attacks, users are still suffering loss of their personal and financial data due to phishing attacks. This indicates that traditional methods are not enough to tackle this menace.

Blacklisting and whitelisting approaches cannot solve zero-day attacks. Comparing anchor-text and URI as proposed by LinkGuard algorithm can result in high false positives, as opposed to its claim since these heuristics can be tricked by new phishers. The Automated Individual White List uses Naïve Bayesian classifier to determine entries to the white list, since the list is maintained locally, user credentials can be stolen during the training period. Another issue is that user has to stick to the same browser and same machine to connect to internet. The change rate of IP is also an issue in AIWL. PhishGuard algorithm submits random credentials to check whether the site accepts every value without verifying the validity of the data. To identify whether the system rejects all credentials submitted, it sends the actual credential to the server. This has the problem of losing the actual credential of the user. In case login attempts are limited the system may become ineffective. PhishNet algorithm predicts possibly new phishy URLs from existing the blacklist. The heuristics used to predict new URLs may not be able to detect new attacks always. This can be enhanced by predicting new phishy URLs to populate the blacklist from legitimate URLs which are normally targeted by phishers.

The Fuzzy-Data Mining approach faces the challenge of identifying proper features to classify the URLs. The Latest machine learning algorithms can outperform this solution. BaitAlarm which used CSS-based similarity detection expects that phishers copy content of legitimate web pages as it is to create a phishy page. But the same CSS layout can be implemented in different ways and phishers can make use of the same to fool the detection system. Computer Vision Technique with Whitelist identify scaling and rotation changes in phishing sites, but it cannot identify if major changes are present on the website. This is summarised in Table 1.

Machine learning approaches have challenges like diversity in datasets, imbalanced scenarios, the use of different metrics for evaluation, etc. The diversity of datasets and classifier performance is strongly related as shown in Table. 3 and Table. 2. Here we can find that Support Vector Machine provides an accuracy of 99.24 when it is tested on OpenPhish and AlexaLogin datasets. But the same classifier drops the result to 96.75 accuracies on PhishTank and Alexa datasets. Decision Tree provides 97.93 accuracies on OpenPhish and AlexaLogin, but it drops to 95.37 on PhishTank and Alexa datasets. Similarly deep learning shows good results on OpenPhish and AlexaLogin but it fails to do so on PhishTank and Alexa.

Classifiers of balanced data set will not provide expected performance on an imbalanced problem. For phishing detection in different scenarios, no single classifier is a complete solution. Proper feature selection is the key to success for classification performance. Since phishers are very proactive and change their techniques often to bypass existing anti-phishing techniques, features in the old dataset will not help us to detect newer zero-day attacks. Automated feature selection and an up-to-date dataset is the real challenge in the machine learning approach. Also it is evident that historical data analysis alone cannot solve phishing attacks.

Table 1. Summary of Analysis of Various Anti-Phishing Algorithms

SlNo	Algorithm Name	Detection Technique	Significance/Limitations
1	Blacklisting	Checks ULRs against blacklist	Cannot detect zero day phishing attacks.
2	Link Guard	Compare anchor text aginst URIs.	Claims 96% of accuracy, Phishers can trick and bypass these heuristics
3	Automated Individual White List (AIWL)	Keep a white list of individual user's Login User Interface at client side	Solves false negative issue of black list based solution, User need to use same browser on same machine always.
4	PhishGuard	Submit random credentials to severs to check if it is accepted	Depends on the server response of success/failure Possibile of loss of user credentials.
5	PhishNet	Predict new malicious URLs from the black listed URLs	Less than 5% of false positives and false negatives
6	Fuzzy Data Mining Intelligent System	Used data mining to identify significant features and Fuzzy logic to classify	Identified feature selection is an important aspect, and URL and domain identity are important indicators. Machine learning can be used for better classification.
7	Bait Alarm	CSS based visual similarity analysis	Claims the detection rate of 100% and 0% of false negatives Phishers can create same look and feel pages with different CSS.
8	Hybrid system of White List and Computer Vision	White List and Speed Up Robust Features Based Computer Vision Tool	Provides Real time detection Proposes improvement using CSS. Changes in page may result in false negatives.

Table 2. Performance of classifiers on OpenPhish and Alexa Login [11]

Classifier	F1 Score	Accuracy
Random Forests	99.23	99.62
Decision Tree	95.78	97.93
Linear Regression	97.33	98.68
SVM	98.46	99.24
Deep Learning	94.48	97.38
5NN	93.63	96.80

Table 3. Performance of classifiers on PhishTank and Alexa [11]

Classifier	F1 Score	Accuracy
Random Forests	96.22	96.87
Decision Tree	94.38	95.37
Linear Regression	95.98	96.67
SVM	96.07	96.75
Deep Learning	95.45	96.29
5NN	94.64	95.53

5. Conclusion

Phishing makes huge damage to web users and organizations alike. It has a huge impact including downtime hours of users, damages to the reputation of organizations, and even intellectual property loss. To tackle phishing attacks different techniques have been explored by researchers. It includes conventional methods like blacklisting, white list based solutions, fuzzy logic, heuristic methods, computer vision techniques, etc. Since phishing detection is a classification problem researchers have analysed possibilities of machine learning techniques including deep learning. From the analysis of these methods, we have found both conventional and machine learning methods contain gaps that help phishers to continue their attacks. Conventional methods have challenges like zero-day phishing attacks, false positives and false negatives, possible-loss of user credentials, the change rate of IP of websites, inadequate heuristics to prevent newer methods of attacks, and many more. The Machine learning approach also has challenges like diversity in datasets, imbal-

anced scenarios, use of different metrics for evaluation, etc. Proper feature selection is the key to success for classification performance. Machine learning based on historical data analysis cannot alone solve this issue. There is a lot to be done in the anti-phishing research area both in conventional and machine learning methods. The possibility of a hybrid of both the approaches can also be explored.

References

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

[2] Aburrous, M., Hossain, M. A., Dahal, K., and Thabtah, F. (2010). Intelligent phishing detection system for e-banking using fuzzy data mining. *Expert systems with applications*, 37(12):7913–7921.

[3] APWG. Phishing activity trends report. Technical report, Anti-Phishing Working Group, 2019 4th Quarter Report.

[4] APWG. Phishing activity trends report. Technical report, Anti-Phishing Working Group, 2020 2nd Quarter Report.

[5] bankinfosecurity.com (Last accessed November 29, 2020). Phishing campaign tied to amazon prime day. *https://www.bankinfosecurity.com/phishing-campaign-tied-to-amazon-prime-day-a-12782*.

[6] Cao, Y., Han, W., and Le, Y. (2008). Anti-phishing based on automated individual white-list. In *Proceedings of the 4th ACM workshop on Digital identity management*, pages 51–60.

[7] Caputo, D. D., Pfleeger, S. L., Freeman, J. D., and Johnson, M. E. (2013). Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, 12(1):28–38.

[8] Chen, J. and Guo, C. (2006). Online detection and prevention of phishing attacks. In *2006 First International Conference on Communications and Networking in China*, pages 1–7. IEEE.

[9] Cyren (2018). The phishing issue from targeted attacks to high-velocity phishing. Technical report, Cyber Threat Report.

[10] Egan, G. (Last accessed November 29, 2020). State of the phish report: Attack rates rise, account compromise soars. proofpoint,threat protection. *https://www.proofpoint.com/us/corporate-blog/post/2019-state-phish-report-attack-rates-rise-account-compromise-soars*.

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

[12] Google.com (Last accessed November 29, 2020). Safe browsing: Malware and phishing. *https://transparencyreport.google.com/safe-browsing/overview*.

[13] Joshi, Y., Saklikar, S., Das, D., and Saha, S. (2008). Phishguard: a browser plug-in for protection from phishing. In *2008 2nd International Conference on Internet Multimedia Services Architecture and Applications*, pages 1–6. IEEE.

[14] Kathrine, G. J. W., Praise, P. M., Rose, A. A., and Kalaivani, E. C. (2019). Variants of phishing attacks and their detection techniques. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 255–259. IEEE.

[15] Kedem, O., Turgeman, A., Novick, I., Zaloum, A. B., Karabchevsky, L., Mintz, S., and Maor, R. U. (2019). Device, system, and method of detecting vishing attacks. US Patent App. 16/188,312.

[16] Mao, J., Li, P., Li, K., Wei, T., and Liang, Z. (2013). Baitalarm: detecting phishing sites using similarity in fundamental visual features. In *2013 5th International Conference on Intelligent Networking and Collaborative Systems*, pages 790–795. IEEE.

[17] Mishra, S. and Soni, D. (2019). Sms phishing and mitigation approaches. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pages 1–5. IEEE.

[18] Nepper, P. and Nair, K. C. (Last accessed November 29, 2020). Better password protections in chrome - how it works. *https://security.googleblog.com/2019/12/better-password-protections-in-chrome.html*.

[19] PhishLabs (Last accessed November 29, 2020). Growing social engineering threats. Technical report, Phishing Trends And Intelligence Report.

[20] Prakash, P., Kumar, M., Kompella, R. R., and Gupta, M. (2010). Phishnet: predictive blacklisting to detect phishing attacks. In *2010 Proceedings IEEE INFOCOM*, pages 1–5. IEEE.

[21] Prasad, R. and Rohokale, V. (2020). *Cyber Security: The Lifeline of Information and Communication Technology*. Springer.

[22] Rao, R. S. and Ali, S. T. (2015). A computer vision technique to detect phishing attacks. In *2015 Fifth International Conference on Communication Systems and Network Technologies*, pages 596–601. IEEE.

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

[24] Shankar, A. and Shetty, R. (2019). A review on phishing attacks. *International Journal of Applied Engineering Research*, 14(9):2171–2175.

[25] www.f5.com (Last accessed November 29, 2020). 2020 phishing and fraud report. *https://www.f5.com/labs/articles/threat-intelligence/2020-phishing-and-fraud-report*.

[26] www.proofpoint.com (Last accessed March 27, 2020). State of phish : An in-depth look at user awareness,vulnerability and resilience. Technical report, ProofPoint Annual Report.

[27] www.us cert.gov (Last accessed November 29, 2020). Google docs phishing campaign. *https://www.us-*

cert.gov/ncas/current-activity/2017/05/04/Google-Docs-Phishing-Campaign.

[28] Xiao, X., Zhang, D., Hu, G., Jiang, Y., and Xia, S. (2020). Cnn-mhsa: A convolutional neural network and multi-head self-attention combined approach for detecting phishing websites. *Neural Networks*.

[29] Zhu, E., Chen, Y., Ye, C., Li, X., and Liu, F. (2019). Ofs-nn: An effective phishing websites detection model based on optimal feature selection and neural network. *IEEE Access*, 7:73271–73284.

⋆⋆⋆⋆⋆⋆⋆⋆⋆
ISSN(P):2319 – 3786
Malaya Journal of Matematik
ISSN(O):2321 – 5666
⋆⋆⋆⋆⋆⋆⋆⋆⋆