Exhaustive study on Detection of phishing practices and tactics

Rutvik Mehta [1], Dr. Keyur N Brahmbhatt[2]

[1] Research Scholar,
Gujarat Technological University, Ahmedabad
rutvik.mehta1990@gmail.com
[2] Associate Professor, Information Technology Department,
Birla Vishvakarma Mahavidyalaya, Vallabh Vidyanagar
keyur.brahmbhatt@bvmengineering.ac.in

Abstract*: *Due to the rapid development in the technologies related to the Internet, users have changed their preferences from conventional shop based shopping to online shopping, from office work to work from home and from personal meetings to web meetings. Along with the rapidly increasing number of users, Internet has also attracted many attackers, such as fraudsters, hackers, spammers and phishers, looking for their victims on the huge cyber space. Phishing is one of the basic cybercrimes, which uses anonymous structure of Internet and social engineering approach, to deceive users with the use of malicious phishing links to gather their private information and credentials. Identifying whether a web link used by the attacker is a legitimate or phishing link is a very challenging problem because of the semantics-based structure of the attack, used by attackers to trick users in to entering their personal information. There are a diverse range of algorithms with different methodologies that can be used to prevent these attacks. The efficiency of such systems may be influenced by a lack of proper choice of classifiers along with the types of feature sets. The purpose of this analysis is to understand the forms of phishing threats and the existing approaches used to deter them.

***Keywords*: Phishing crimes, Machine learning, Nature Inspired methods, Classification methods, Visual similarity based methods, List based methods**

(Article history: Received: 6th February 2021 and accepted 17th May 2021)

I. INTRODUCTION

Phishing attacks' metaphor is inferred from the term "fishing" for victims. It is also a tempting and enticing tactic for attackers who are launching a variety of fake websites which is of almost same visual appearance as some of the popular and legitimate websites on the Internet. Attackers are attracted by to this type of attacks, a crime-based practice, designed to persuade users to reveal their sensitivities data such as credit card details, authentication data, pin and passwords etc. The details obtained by phishing are used to financial access and participate in illegal activities such as downloading malware or spyware on the remote machines [1-2].

Attackers are able to carry forward this attack because of the following characteristics of Internet Users [10].

- Users have no detailed knowledge of Uniform Resource Locators (URLs).
- Users are not able to see the full URL of the web page due to the redirection to the hidden URLs.
- Users don't have much time to search the URL, or look for authenticity of the website.
- Users can not differentiate phishing web pages from legitimate web pages.

Because of the insufficient knowledge available with the users, attackers carry out this attack by targeting individual user's vulnerabilities and hence they are getting more popularity. Statistics shows that there is a hike in the phishing attack in recent times. The IC3 report [4] noted that there were 26,379 victims of "phishing/ vishing/ smishing/ pharming" in 2018, accounting for $48,241,748 in losses. One in ten URLs are malicious [6]. A report from Lookout [7] revealed that 56% of mobile device users received and tapped on a phishing URL. Out of all mobile threats, 33.6% cases in India were phishing threats. Phishing is now not limited to emails. 85% of phishing attacks seen on mobile devices take place outside of email. 5,707 phishing attacks were reported in India in 2018 [6]. Dark Caracal that focused on mobile phishing, compromised over 600 phones in over 21 countries [7]. FrozenCell, xRAT, ViperRAT, SocialPath, and Xsser/mRAT are all mobile threats that start with phishing [7]. Anti-Phishing system of kaspersky detected more than 130 million redirects to phishing sites [8].

This attack normally starts when the intruder or hacker sends a message or an email aimed at social engineering victim that appears to be original to the target. Attacker tricks them to update and check their details by clicking on the Uniform Resource Locator link in the email or private chat message, blogs, and forums as well as on banners which are representing legitimate sources. [3]. Attacker tricks user to

enter his/her personal information through building a similar user interface to other trusted websites, Users will enter

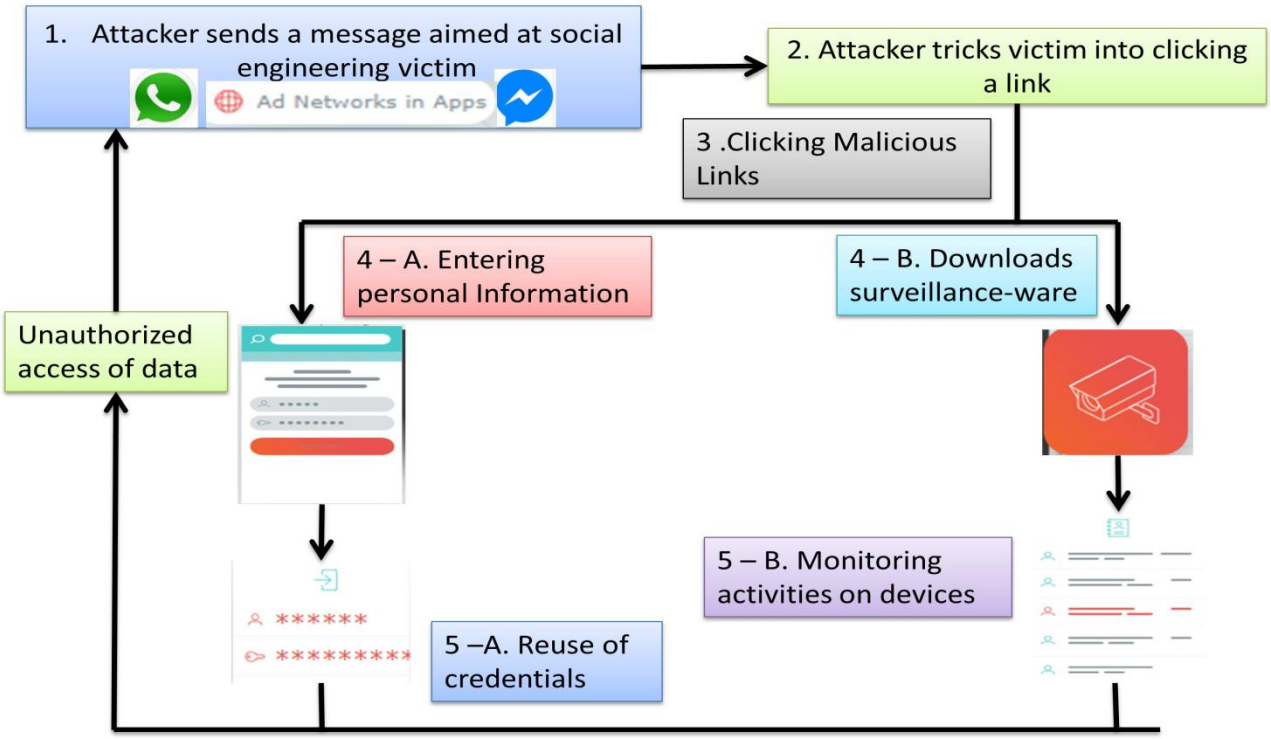

Fig. 1. Working of Phishing campaign in OSN

confidential information known as web-based phishing. Another way to perform phishing attack is to insert malicious code into a legitimate website and fake apps can be mounted on the user's device if users visit the website. This is known as malware-based phishing [9]. Mechanisms used by attackers to perform phishing attack are represented in figure -1.

In this presented work, different methods used for detecting the phishing attack have been investigated. Related work to detect phishing attack has been presented in section II. In section III, different parameters used for phishing detection have been discussed. Section IV discusses about different databases available for phishing detection. Section V concludes with future steps.

II. Related Work

Already, a range of studies have been suggested for detection of phishing attacks in literature and consumer products. The phishing detection systems are generally divided into two groups: User Awareness based and software based as illustrated in Figure - 2. As illustrated in earlier section, it is very difficult to detect phishing attacks based on user awareness because of lack of information with users and time involved in deciding authenticity of the page by the user.

A. List based phishing detection methods

List based phishing detection methods use two different sets, whitelist which contains legitimate websites and blacklist which contains phishing websites.

Google safe browsing service [13] enables client applications to search URL against Google's continuously updated list of malicious web services. This black list based phishing detection approach allows client API to send the URLs with the Hypertext Transfer Protocol (HTTP) based GET or POST requests, checked using the list of malware and phishing links used by Google. This service does not use hashing before sending the URL and does not limit on the response time taken by lookup server.

In [12], Jain and Gupta proposed a method which alerts users on the Internet with a whitelist of legal websites which is updated automatically. The process is divided in two modules. First module focuses on domain IP matching while later focused on the extraction of features from URL and source code. Experimental results of this approach showed it has 86.02 % true positive rate with a very less false positive rate of 1.48 %.

In [14], Rao and Ali, also used a white list based phishing detection method against the URL extracted from different pages along with the speeded up robust features (SURF) algorithm which uses computer vision technique to detect phishing pages. The technique of phishing detection was evaluated using many different threshold values for SURF algorithm, and a False Positive of 20.11% and False Negative rate of 15.23% were achieved respectively.

Many of the list based phishing detection methods are proposed and they have produced different results but these list based detection methods use an approximate matching algorithm to verify whether or not a suspect URL occurs in the blacklist/whitelist and these solutions need periodic changes in the lists. Moreover, the accelerated growth of the blacklist/whitelist needs an enormous amount of system resources.

B. Heuristic and rule based phishing detection algorithms

Heuristic and rule based phishing detection algorithms are based on the similarities between different phishing

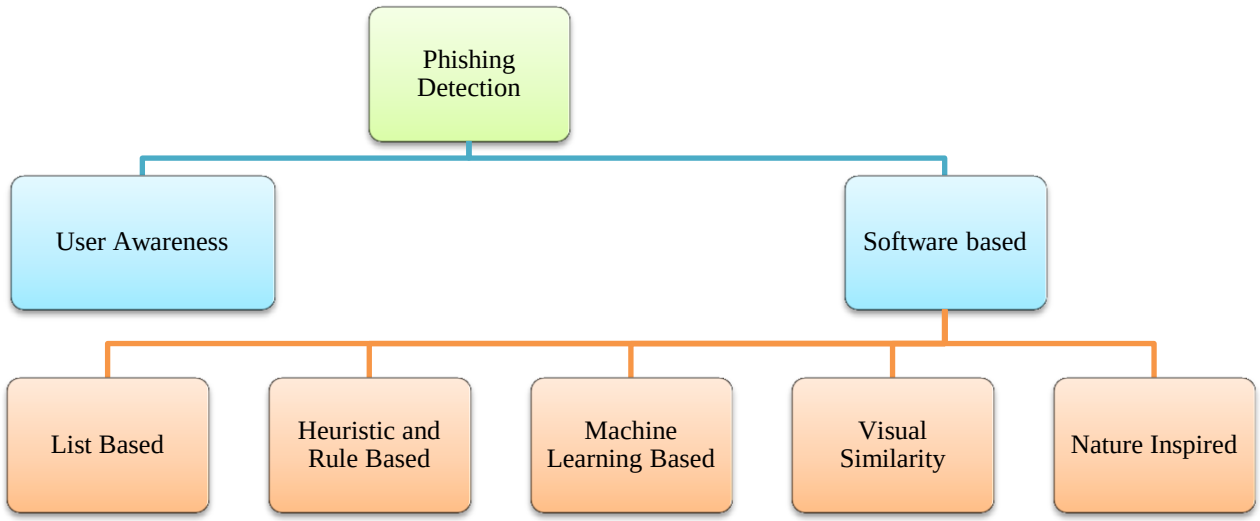

Fig. 2. Different methods of detecting phishing attacks

pages, the predictive statistical characteristics of the phishing pages and the knowledge available with experts prior to the attack. It extracts several features from phishing pages and generalizes them into series of heuristic features and set of rules. In [15], Sathish and Thirunavukarasu, proposed a model based on a fuzzy logic approach which consist of three stages data collection, rule creation and classification. First stage of this model was based on extraction of data like Google page rank, IP number in URLs, host name presence. Fuzzy rules were made in second stage and then it was supplied to the classification.

In [16], Sunil and Sardana, proposed a method, a well-known approach called as CANTINA+ based on Google toolbar rank and other URL related features. It uses simple forward linear model to classification algorithms which gives True positive rate of and false positive rate along with false negative rate of 0.98 and 0.02 respectively.

In [17], Nathezhtha, Sangeetha and Vaidehi, proposed Web Crawling based Phishing Attack Detection (WC-PAD), a Web crawling based phishing attack detection which uses Heuristic Analysis based on page rank, alexa reputation, URL analyzer, Web-content analyser. This method generates alerts once when either listing based approach detects phishing page or it is detected by heuristic approach. WC-PAD is able to detect zero day phishing attack with 98.9% accuracy.

In [30], Sonowal and Kuppusamy, made a browser extension for real time detection of phishing attack which also includes implementation for the visually impaired persons. phishing detection model with multi-filter approach (PhiDMA) uses a five layer filtering approach namely white-list filter, URL feature filter, Lexical signature filter, String matching filter and Accessibility filter. Every time scored obtained from each of the layer is compared with the threshold and if score is more than threshold URL was classified as phishing URL. Accessibility score was fetched using a third party tool called as Acheker. Model gave the accuracy of 92.72%

C. Visual Similarity based algorithms

Visual similarity based algorithms uses similarity measure between the phishing site and the already available database of resources such as Document Object Model (DOM) structures, Visual features, Cascading Style Sheets (CSS) features, pixel based features, visual perception, logos and favicons. In [22], Haruta et al., proposed a visual similarity approach for detecting phishing attack based on hue signatures of the pages. This approach generates a hue signature of the 100*100 size snapshot of the website and looks for the dominant color ratio and color combination with the signature database. It also updates the signature database if hue signature is similar to any legitimate websites. With the help of this auto-updating feature, this system can have tolerance to zero-day phishing attack.

In [23], Abdelnabi et al., proposed a visual similarity based approach with a new dataset of 155 websites with 9363 screenshots. This approach uses the triplet network paradigm with three shared convolutional networks. It fine-tunes the model weights by testing hard cases that were mistakenly identified by the last checkpoint according to the distance between the embedding. This will generalize new websites which are not included in the trusted list set.

In [24], Rao and pais, worked on a light weight visual similarity detection model which focuses on non-whitelisted legitimate sources. Instead of implementing whitelist based approach this technique implements black list based approach as a first level filter to detect the near duplicate phishing sites. For non-blacklisted phishing site this approach uses second level filter based on fuzzy similarity measures, simhash and perceptual hash and generates a fingerprint of the given websites. It uses ensemble model of

Random forest, extra tree and Xgboost to generate the accuracy of 98.72%

D. Naure Inspired algorithms

Nature inspired techniques, like Ant colony optimization, Particle Swarm Optimization and Artificial Immune system, inspired by the approach which nature adopts to solve the problems, can be used to detect phishing attacks also.

In [18], Fang et al., proposed an artificial immune system, a nature inspired mechanism to detect phishing pages which identifies phishing emails based on memory detectors and immature detectors. Memory detectors represent the earlier seen phishing emails while immature detectors are reproduced by using the mutation process. If signature of the new incoming email matches with signatures present in the memory detectors in generates the alert. This proposed mechanism uses immature detectors to identify new incoming phishing emails. With different Fire Alarm Rates values, researchers were able to produce True positive rate of 0.97 and false positive rate was 0.0375.

In [19], Adewumi and Akinyelu, proposed a natured inspired model, hybrid firefly in combination with Support vector machine (SVM), for detecting phishing emails. Here firefly approach uses the concept of difference in light intensity with distance between fireflies, and with the degree of absorption by the atmosphere to optimize the fitness function of phishing detection problem. This concept was tested along with the SVM known as support vector machine firefly algorithm (FFA_SVM), on dataset containing 4000 phishing emails and it produced accuracy of 99.94.

In [20], Chen et al., presented back propagation(BP) neural network approach along with the Particle swarm optimization (PSO) to build phishing website detection systems. In this method PSO was used to search the optimal initial parameters of neural network are by the movement and update of particles. This combined approach overcomes the disadvantage of BP neural network to fall into the local optimum. This method achieved the accuracy of 0.9895 with false negative ratio of 0.0119.

In [21], Vrbancic et al., proposed a method which uses three different nature inspired algorithms like Bat algorithm, Hybrid Bat algorithm and firefly algorithm for the parameter setting for deep neural network (DNN) designed for phishing detection. The predictive performance of the resulting DNN, optimized using the swarm intelligence based methods, gives better result in detecting phishing websites when compared to the manually tuned neural network

E. Machine learning based algorithms

Machine learning based techniques are better than black list and white list based approaches as these techniques can detect zero-day phishing attacks which are not possible with list based approach as list based approach take time to update the list. They are better than Rule based approach also as rule based approaches can be bypassed by the attackers. Machine learning based techniques are better than even visual similarity based approaches as visual similarity based approaches require more computation resources and graphics processing unit (GPU)compared to the machine learning based resources.

In [25], Sahingoz et al. proposed a machine learning and Natural Language Processing (NLP) based phishing detection method which combines NLP features and word vector. This hybrid approach focuses on 1701 word features and 40 NLP features. They used seven different classification algorithms namely decision tree, Adaboost, kstar, kNN, Random Forest, Sequential minimal optimization (SMO), Naïve Bayes. NLP based features outperforms word vectors with the average rate of 10.86%. The use of NLP based features and word vectors together also increases the performance by 2.24% according to NLP based features and 13.14% according to word vectors.

In [26], Jain and Gupta, proposed a random forest based classification algorithm which focuses on 19 different client side features like URL features, Login from features, Hyperlink features, CSS features and web identity features. 8 features out 19 features are based on URL, 2 are based on Fake login form, 6 are based on hyperlink, 1 is based on CSS and 3 are based on web identity. These features along with five different machine learning algorithms namely Radom Forest, Support vactor machine, neural networks, Logistic regression and Naïve Bayes were used to perform testing. Random Forest algorithm outperformed all other classification algorithm with the accuracy of 99.09%.

In [27], Peng et al., proposed an approach to detect phishing emails based on text contained in the email. It uses semantics of the statements present in the email and determines if it exhibits malicious question or command, urgent tone, generic greeting and malicious URL link. It considers the reputation of the URL using Netcraft anti-phishing toolbar. They have used 1000 phishing emails Nazario phishing email set and 1000 non-phishing emails from the enron corpus to test their proposed approach. Multinomail Naïve Bayes algorithm was used for classification and provides a precision of 95%.

In [28], Rao and Pais, Considered three different feature set made up of 16 features: URL obfuscation, third-party, hyperlink for detection of phishing attack. This approach also checks whether the page is replace with an image or not. Eight different machine learning algorithms were used in training the model including logistic regression (LR), J48 tree, Random Forest (RF), multilayer perceptron (MLP), Bayesian network (BN), SVM, SMO and AdaBoostM1 (AM1). Out of these algorithms, RF performed well with the accuracy of 99.31%. This approach was carried forward with different orthogonal and oblique RF variants. Principal component analysis Random Forest (PCA-RF) gave the best results out of all possible oblique Random Forests (oRFs) with an accuracy of 99.55%.

In [29], Liew et al., used 11 classification features along with Random forest classifier to classify a phishing tweet in real time. Features used by them are URL length, Host length, Path length, Registrar, SSL connection, Hexadecimal, Alexa rank, Age of domain - Year, Equal, Digit in host and Number of dots in host. Experiments show that RF classifier along with this 11 features achieved accuracy of 94.75%.

III. FEATURES USED FOR PHISHING DETECTION

Many different phishing classification algorithms are represented in the study and many other are already proposed by researchers and used in commercial products. These

algorithms identifies phishing attack based on different types of features like page rank and accessibility scores, body and content, visual similarities, Image based, WHOIS, URL

TABLE I. FEATURES USED FOR PHISHING DETECTION

Feature type	Feature	References
1. URL based Features	1.1Having IP address	12, 14, 15, 16, 17, 26, 28, 29, 30, 31
	1.2 Shorten URL	
	1.3 Having @ symbol as a prefix. URL resolvers ignores @ in the prefix	
	1.4 URL Length	
	1.5 Double slash redirection	
	1.6 Use of Non-standard ports	
	1.7 Use of http or https in domain part of the URL	
	1.8 More dots in the URL representing more subdomain	
	1.9 Use of prefix suffix with dash (-)	
	1.10 Multiple Top level Domains	
	1.11 Position of Top Level Domains	
	1.12 URL hostname containing words such as: “www”, “signin”, “blog”	
	1.13 Data URI	
	1.14 Brandname inURL	
2. WHOIs – Features	2.1 Domain Name registration	17, 29,30
	2.2 Age of domain	
3. DOM Features	3.1 Foreign Hyperlinks	26, 28
	3.2 Empty Hyperlinks	
	3.3 External css with foreign domain names	
	3.4 No Hyperlink	
	3.5 Error in hyperlinks	
	3.6 Common page detection ration (ratio most common anchor links and total links)	
4. Login form based features	4.1 redirection to the login pages using script (hyperlink rediects to login page)	26
5.Lexical and content semantic based features	5.1 Term frequency	11, 17, 27, 30
	5.2 Semantic based malicious question or command, urgent tone, generic greeting	
	5.3 word vectors	
	5.4 presence of brand names	
6. Social Network features (Applicable to social networks only	6.1 Number of followers	29
	6.2 Number of followees	
	6.3 Ratio of followers and followees	
	6.4 Age of account	
	6.5 Number of posts	
	6.6 Presence of personal details and about information	
7. Page rank and reputation based	7.1 Alexa reputation	28, 29, 30
	7.2 google page rank	
	7.3 Traffic rates	
8. Visual content and Image based features	8.1 Hue based representation	22, 23, 24, 26
	8.2 Presence of logos of legitimate websites	
	8.3 favicon	
	8.4 Web page replaced with images	

-based features, DOM based features, Social network related features, login form based features, CSS based, web identity based, pixel based and visual perception based etc. These features are mentioned in detail in Table - I.

IV. DATASET USED FOR PHISHING DETECTION

Many of the different datasets are used in different phishing detection mechanism. This section gives details about those datasets. Phistank [33], Spamassasian [35], Anti-phishing working group [34] and UCI [32] data are widely used Dataset for phishing detection algorithms. Other datasets like one at [36] contains 96,018 URLs. Out of these URLS, 48,009 legitimate URLs and 48,009 phishing URLs. Dataset at [37] is updated every 90 minutes which provides data in Comma Separated Values (CSV) and JavaScript Object Notation (JSON) formats. In [38], Snapshot of the websites which were used to perform phishing attack is available which actually look visually similar to the legitimate websites. Many of the research do not contain details about the dataset used because of confidentiality issues.

V. CONCLUSION AND FUTURE DIRECTIONS

In this study, different mechanisms for detection of phishing attack have been studied along with the feature set used for the classification. Many techniques are focused on email based phishing attacks [18, 19, and 27] only but phishing is carried by attackers not only using emails but various platforms like mobile phishing and social network.

List-based solutions have the fast access time but they suffer from the low detection rate especially for the zero-day attacks. Some algorithms are dependent on third party services like alexa rankings and WHOis servers [17, 28, 29, 30]. These services violate user privacy and their searching history. Access to external resources is unstable and unpredictable. Ex. API for Google page rank deprecated in 2016. Performance of such services is network dependent and the format used by these services to return the data may take more time in processing the data. Rule based approaches are based on different rules which can be easily bypassed by attackers who are having knowledge of such approaches. Visual similarity based approaches [22, 23, 24] requires image processing which in turn requires more computing and storage resources. Many of the phishing detection algorithms were trained using small and limited dataset which are neither robust nor generalize the phishing patterns. Approaches which analyse the DOM features [26, 28] are not efficient alone as newer attacks closely mimic legitimate websites. Many algorithms are based on content of the webpages, term frequencies and semantics of the content [11, 17, 27, 30]. These algorithms are based on English only but many of the web pages' content is written in different regional languages which means these algorithms cannot identify phishing pages which contains content in other languages.

While research on detecting a phishing has a long history, Detection of phishing in the context of online social network data is a relatively new research field. Research can be conducted on detecting social media posts which contains phishing URLS based on various post related features and social network features. Machine learning techniques are dependent on some features which can be easily bypassed by different attack patterns. Deep learning methods can be used build improved knowledge on different attack patterns which can be used by attackers. Deep learning models can be used for visual similarity based approaches for better results. Nature inspired algorithms are not fully explored for detecting phishing attacks. These algorithms will be able to handle the changes in phishing attack patterns because of their nature to adapt current situations. Ensemble learning approaches might require more time to produce the final class for a website but can be used to combine results from many different classifiers which can be used to increase the accuracy of classification. Accuracy of these algorithms is based on feature selection also. Different algorithms can be used for feature selection and parameter optimization. With these approaches researchers might be able to tackle changing attack patterns for phishing.

REFERENCES

[1] H. Y. Abutair, A. Belghith(2017), Using case-based reasoning for phishing detection. Procedia Computer Science, Vol. 109, pp. 281–288

[2] S. Rathore, P. Sharma, V. Loia, Y. Jeong, J. H. Park (2017), Social network security: Issues, challenges, threats, and solutions, Information Sciences, Elsevier, vol. 421 pp. 43-69.

[3] H. Shirazi, B. Bezawada, I. Ray (2018), Kn0w Thy Doma1n Name: unbiased phishing detection using domain name based features, Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, pp. 69–75.

[4] IC3 Internet Crime Report(2019) "https://pdf.ic3.gov/2019_IC3Report.pdf"

[5] APWG - Phishing Activity Trends Report - 1st quater (2020), "https://docs.apwg.org/reports/apwg_trends_report_q1_2020.pdf"

[6] Symantec, Internet Security Threat Report(2019), "https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf"

[7] Mobile phishing: Myths and facts facing every modern enterprise today(2019),"https://info.lookout.com/rs/051-ESQ-475/images/Lookout-Phishing-wp-us.pdf"

[8] Spam and phishing in Q2 (2019), "https://securelist.com/spam-and-phishing-in-q2-2019/92379/"

[9] Z. Dong, A. Kapadia, J. Blythe, L.J. Camp (2015), Beyond the lock icon: real-time detection of phishing websites using public key certificates, 2015 APWG Symposium on Electronic Crime Research (eCrime), pp. 1–12, IEEE.

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

[11] O. K. Sahingoz, E. Buber, O. Demir, B. Diri (2019), Machine learning based phishing detection from URLs, Expert Systems With Applications, Elsevier, Vol. 117, pp. 345-357.

[12] A. K. Jain , B. B. Gupta (2016), A novel approach to protect against phishing attacks at client side using auto-updated white-list, EURASIP Journal on Information Security Article number: 9, pp. 1-11.

[13] https://safebrowsing.google.com/

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

[15] S. Sathish, A. Thirunavukarasu (2015), Phishing webpage detection for secure online transactions, International Journal of Computer Science and Network Security, vol. 15 Issue 3, pp. 86–90.

[16] A. N. V. Sunil, A. Sardana (2012), A PageRank Based Detection Technique for Phishing Web Sites, IEEE Symposium on Computers & Informatics , pp. 58-63

[17] Nathezhtha, D. sangeetha, V. Vaidehi (2019), WC-PAD: Web Crawling based Phishing Attack Detection, International Carnahan Conference on Security Technology (ICCST), pp. 1-6.

[18] X. Fang, N. Koceja, J. Zhan, G. Dozier, D. Dipankar(2012), An artificial immune system for phishing detection, IEEE Congress on Evolutionary Computation, pp. 1-7,

[19] O. A. Adewumi , A. A. Akinyelu(2016), A hybrid firefly and support vector machine classifier for phishing email detection, Kybernetes, Vol. 45 Issue 6, pp. pp. 977 – 994

[20] W. Chen, X. A. Wang, W. Zhang, C. Xu(2018), Phishing Detection Research Based on PSO-BP Neural Network, Advances in Internet, Data & Web Technologies, pp. 990–998

[21] G. Vrbancic , Iztok Fister, Jr., V. Podgorelec(2019), Parameter Setting for Deep Neural Networks Using Swarm Intelligence on Phishing Websites Classification, International Journal on Artificial Intelligence Tools, Vol. 28, No. 6, pp. 1-28

[22] S. Haruta, H Asahina, F Yamazaki, Iwao Sasase (2019), Hue Signature Auto Update System for Visual Similarity-Based Phishing Detection with Tolerance to Zero-Day Attack, IEICE TRANSACTIONS on Information and Systems, Vol. E-102 – D No.12, pp.2461-2471

[23] S. Abdelnabi ¸ K. Krombholz, M. Fritz(2020), VisualPhishNet: Zero-Day Phishing Website Detection by Visual Similarity, arXiv, pp. 1-18

[24] R. S. Rao, A. R. Pais (2019), Two level filtering mechanism to detect phishing sites using lightweight visual similarity approach, Journal of Ambient Intelligence and Humanized Computing, pp. 1-20

[25] O. K. Sahingoz, E. Buber, O. Demir, B. Diri (2019), Machine learning based phishing detection from URLs, Expert Systems With Applications, Elsevier, Vol. 117, pp. 345-357.

[26] A. K. Jain, B. B. Gupta(2018), Towards detection of phishing websites on client-side using machine learning based approach, Telecommunication Systems, Vol. 68, pp. 687-800.

[27] T. Peng, I. G. Harris, Y. Sawa(2018),Detecting Phishing Attacks Using Natural Language processing and machine learning, 12th IEEE International Conference on Semantic Computing, pp. 300 – 301.

[28] R.S. Rao, A. R. Pais (2019), Detection of phishing websites using an efficient feature-based machine learning framework, Neural Computing and Applications, Springer, Vol. 31, 3851–3873.

[29] S. W. Liew, N. F. M. Sani, Mohd. T. Abdullah, R. Yakob, M. Yunus Sharum (2019), An effective security alert mechanism for real-time phishing tweet detection on Twitter, Computers & Security, Elsevier, Vol. 83, pp. 201-207.

[30] G. Sonowal, K. S. Kuppusamy (2020), PhiDMA – A phishing detection model with multi-filter approach, Journal of King Saud University –Computer and Information Sciences, Elsevier, Vol. 32, pp. 99-112.

[31] M. Babagoli, M. P. Ahgababa, V. Solouk (2019), Heuristic nonlinear regression strategy for detecting phishing websites, Vol. 23 , pp. 4315–4327.

[32] https://archive.ics.uci.edu/ml/datasets/Phishing+Websites

[33] https://www.phishtank.com/developer_info.php

[34] https://apwg.org/ecx/

[35] https://spamassassin.apache.org/old/publiccorpus/

[36] https://research.aalto.fi/en/datasets/phishstorm-phishing-legitimate-url-dataset

[37] https://phishstats.info/

[38] https://www.circl.lu/opendata/datasets/circl-phishing-dataset-01/Clean_phishing/

AUTHOR PROFILE

Rutvik Mehta

He is pursuing his PhD degree from Gujarat Technological University. He has received his bachelor degree from Gujarat University and Master of Engineering Degree from Gujarat Technological University. His research interests include machine learning and cyber security. He has published and presented more than 15 research papers at National, International Conferences, and Journals.

Dr. Keyur Brahmbhatt,

He holds a bachelor Degree in Information technology from North Gujarat University. He obtained his Master of Engineering degree in Computer Engineering from Sardar Patel University. He has completed his Doctor of Philosophy (PhD) degree in Computer Engineering from Charusat University. At present, he has a total experience of 13 years in the field of Education and Research. He has published and presented more than 30 research papers at National, International Conferences, and Journals. He has received an Academic Excellence Award at the college level for the year 2017-18. He is also the author of the book named “Image Fusion.”