# Phishing Website Detection Based on Machine Learning: A Survey

Charu Singh
Department of Computer Science & Engineering
Madan Mohan Malaviya University of Technology
Gorakhpur, India
charusingh0011@gmail.com

Smt.Meenu
Department of Computer Science & Engineering
Madan Mohan Malaviya University of Technology
Gorakhpur, India

*Abstract*—**Phishing attacks are cybercrime in today's world which are done by social engineering and malware based. It is one of the most dangerous threats that every individuals and organization faced. URLs are known as web links are by which users locate information on the internet. The review creates awareness of phishing attacks, detection of phishing attacks and encourages the practice of phishing prevention among the readers. In phishing, phishers use email or message, as a weapon to target individual or organization by send URL link to target people and to deceive them. With the huge number of phishing emails or messages received every day, companies or individuals are not able to detect all of them. Here, different reviews give for detection of phishing attack, by using machine learning. Here it is used for detecting the web links, i.e., either phishing or legitimate.**

***Keywords—Social Engineering, Phishing, Legitimate, Machine Learning***

## I. Introduction

Due to rapidly growing technology internet has become an integral part of our daily life [1]. Lots of activities in our daily life are determined after the use of the internet. Social networking sites have rapidly increased over the last few years. Due to the regular use of the internet, the users have to undergo many threats; one of them is 'Phishing'.

The major problem is "phishing" is one of the today's world. Social engineering and malware based are the phishing attacks which contain malicious websites that are attached to E-mail, SMS or other communication method to deceive people. It is cybercrime or fraud that uses spam email as a weapon. Email spoofing or instant messaging carried out phishing. These emails and messages contain a URL link directs users to another website. It often directs users to enter personal information or sensitive information i.e., password, credits card details at a forged website which look like legitimate site.

## II. BACKGROUND AND OVERVIEW OF PHISHING

### *A. HISTORY*

In 1970's John Draper defined the term 'phishing'. For hacking telephone systems, he created infamous Blue Box that emitted audible tones [2]. Social engineering attacks are done in 1996, against America Online (AOL) accounts by online scammers [3].

### *B. PHISHING STATISTICS*

According to APWG, in 1Q 1,80,768 phishing sites was detected and unique phishing report was 112,393. HTTPS encryption protocol protects phishing sites. 58% of phishing sites were using SSL certificates. It was detected that 55% of SSL were used in phishing attacks in 2Q of 2019. SaaS & Webmail providers were counted as the most targeted sector with 36% of all phishing attacks recorded targeting its constitutions brands, according to APWG member MarkMonitor [4].

According to the RSA report, 2019, 29% of phishing attack has observed by RSA in 1Q. Fraud attacks from rogue mobile applications increased by 300% from 10,331 in 1Q. Card-not-present (CNP) fraud transactions increased 17% last quarter, and 56% of those originated from the mobile channel. The average value of a CNP fraud transaction in the U.S. was $403, nearly double that of the average genuine transaction of $213. 14.2 million unique compromised cards recovered over RSA in 1Q [5].

### *C. TYPES OF PHISHING ATTACKS*

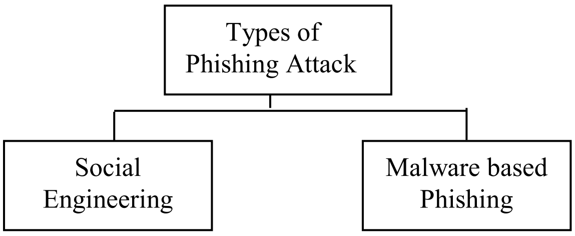
Phishing is a method by which attackers try to gain sensitive information from the users, to use it fraudulently [6]. Social engineering (deceptive phishing) and malware-based phishing are the types of phishing attacks.

Social engineering attacks usually exploit mind and susceptibility to manipulate users for obtaining confidential information. Malware based phishing refers to a scam in which malicious software or unnecessary programs run on the user's system. The malware uses a key logger, screen logger to record your keyboard strokes and sites that you visit on the internet. Key loggers, Session hijacking, DNS phishing, content-injection phishing, phone phishing, system

978-1-7281-5197-7/20/$31.00 ©2020 IEEE

reconfiguration, link manipulation is the classification of these attacks [7][8].

Fig.1: Types of Phishing Attacks



*D. MACHINE LEARNING MODELS*

Machine learning tends to predict whether the websites are phishing or legitimate. It learns the characteristics of phishing websites and predicts new phishing characteristics [7]. Prediction can be done by several algorithms such as Naïve Bayes, Decision Tree, Support Vector Machine, Random Forest, Artificial Neural Network, K-Nearest Neighbor, Bayesian Classification. Phishing detection accuracies vary from each other [7] [9].

## III.PHISHING DETECTION APPROACHES

There are various defense methods of phishing attacks which shown in table 1 [9].

Table1: Phishing Detection Approach

| S.No. | Approach | Technique |
|---|---|---|
| 1. | Heuristic based approach | Decision tree algorithm |
| 2. | Blacklist approach | Simhash algorithm |
| 3. | Fuzzy based approach | Fuzzy data mining algorithm |
| 4. | Machine Learning approach | Machine learning algorithm |
| 5. | CANTINA based approach | TF-IDF information retrieval algorithm |
| 6. | Image based approach | Web logo technique |

## IV. FEATURE EXTRACTION

There are various methods by which features can be detected. These are classified as follows:

a) Source code features

b) URL features

c) Image feature

a) Source code features

1. Tracking of login screen:

To check if it contains any text box to get information from the user, such as username, password, and PIN which is done by tracking of the login screen [10].

2. Disabling Right Click

Right-click is disabled by phishers so that the website code is not able to visualize to the users [10].

3. Pop Up

The legitimate sites do not ask them to enter their credentials while some messages appear in phishing sites to enter their details. [10].

b) URL features

Features of phishing URLs are identified by machine learning. Host-based features and lexical features are the types of features which is extracted from the URL [11] [12].

URLs are simply divided into sub parts, in which it contains, host name, a path, protocol or scheme. Based on any combination of these components, accuracy of site's legitimacy can access [11]. There are 30 phishing website features [1].

1.1. Address Based Features

1.1.1. IP Address

In phishing, at place of domain name of website it contains IP address.

1.1.2. URL length

In address bar phishers use long URL to hide suspicious part.

1.1.3. “Tiny URL”

On the “worldwide web” a method, URL shortening in which URL considered smaller in length. Phishers use “Tiny URL” for deceiving people in which long URL is used to connect the tiny URL [13].

1.1.4. URL’s containing “@” symbol

For deceiving people phishers use “@” symbol.

1.1.5. Redirecting using “//”

The user will be diverted to another website by using “//” in URL path. In legitimate, number of pages is less than 2, for suspicious it lies between 2 - 4, otherwise it is considered as phishing [13].

1.1.6. Adding (-) sign to the Domain

In legitimate URLs, a dash symbol is used for creating malicious URLs to deceive people [14].

**1**.1.7. Dots in URL

Number of dots counted in phishing or legitimate websites in URL. It is considered “Phishing” if number of dots is greater than legitimate [15].

1.1.8. HTTP with SSL

HTTPs is used by legitimate website for moving sensitive data. It requires certificate for using it with minimum age of 2 years.

1.1.9. Domain Registration Length

Phishing websites live for a short time period whereas trusted domains paid for several years in advance.

1.1.10. Favicon

In web page a graphic image is created which is known as favicon. If there is inconsistency between favicon of domain and URL, then it is considered as Phishing [14] [13].

1.1.11. Using Non-standard Port

On a specific server certain service is up and down. User data is in danger, if all the ports are opened. For controlling intrusion, it is advice to open only those ports which are necessary.

1.1.12. “HTTPS” Token in the Domain Part

In URL, HTTPs symbol is added to deceiving people.

1.2. Abnormal Based Features

1.2.1. Request URL

Request URL examines whether the contains on the website are loaded from the same website or from another website. URL with greater than 61% is considered as phishing, and if it lies between 22%-61%, then suspicious, otherwise legitimate [15] [13]

I.2.2. URL of Anchor

Tag <a> is defined as Anchor element which is treated as “Request URL”.

I.2.3. Links in <Meta>, <Script> and <Link> tags

Legitimate website uses <Meta> which display about HTML, <Script> create a client-side script and <Links>, and <Links> tag accept other web sources. % of <Meta>, <Script> and <Links> is less than 17 % then it is considered as legitimate, if it lies between 17%-81% then it is considered as suspicious, otherwise phishing [14][13].

1.2.4. Server form Handler

An empty string or blank is considered as phishing. If webpage is differed from the SFHs, then it is considered as suspicious, otherwise legitimate [14] [13].

1.2.5. Submitting Information to Email

Submission of personal information is typically carried out by web services in which it redirects to the phisher mail. “mail()”, “mailto” function are used for server-side as a scripting language in PHP.

1.2.6. Abnormal URL

WHOIS database extract the feature. In URL, legitimate website is typically an identity part.

1.3. HTML and JavaScript-based Features

1.3.1. Website Forwarding

Distinguishes phishing websites from legitimate is based on redirection. Legitimate websites redirected one-time max whereas phishing websites redirected at least 4 times [13].

1.3.2. Status Bar Customization

Deceptive URLs are shown by phishers to deceive people with the help of Java Script. If source code of webpage is known, then any changes in the status bar can be done by “onMouseOver” event.

1.3.3. Disabling Right Click

Right-click function is disabled by phishers so that users are not able to see and save the web page source code [13].

1.3.4. Using Pop-up window

In pop-up window personal information is to ask by the user in phishing websites whereas legitimate website does not ask to submit [14].

1.3.5. IFrame Redirection

Phishers hide the webpage tag i.e., without frame border by ‘iframe’ and make it invisible to the users. So, for visual delineation phishers use frame border [14][13].

1.4. Domain Based Features

1.4.1 Age Domain

Age Domain checks the age of webpage. Phishing webpage remains for a shorter period of time whereas legitimate having minimum 6 months age. All these features are extracted from WHOIS database.[13]

1.4.2. DNS Record

In phishing website WHOIS database does not identified host name. In DNS record if it found empty, then it is considered as phishing otherwise legitimate. [13].

1.4.3. Page Rank

Weighted of a page rank lies between "0" to "1" where "0" indicates low page rank and "1" indicates higher page rank. Highest page rank is the most important for the webpage. Phishing webpage remains for a shorter period of time in which page rank does not exist. [14]

1.4.3. Website Traffic

A website is defined as phishing or legitimate is determined by its popularity i.e., page rank. Legitimate website ranked among top 100,000 whereas greater than 100,000 is considered as phishing.[14].

1.4.5. Google Index

Website is Google Index or not is based on indexing. Phishing webpage remains for shorter time period. When Google's indexed a site, it is displayed on search results.[14]

1.4.6. Links Pointing to Page

Website security depends on greater number of links. Website is phishing if number of links is 0, if it lies between 0-2 then considered as suspicious, otherwise legitimate [14]

1.4.7. Statistical Reports based Features

Statistical reports on phishing websites have been defined by Phish Tank and Stop Badware over a period of time [15].

c) Image Features

1. Grayscale:

0 or 1 value is contained by image. 0 is considered for black and 1 is for white, in which strength of the information is transmitted [10].

2. Color Histogram:

According to intensity of the colored image, pixels are categorized [10].

## V.RELATED WORK

Studied on various phishing detection methods have been studied. They are classified as Blacklist, Heuristic, Content Analysis, Machine Learning techniques.

Sonmez et. al [1] studied classification technique on 30 phishing websites features using Extreme Learning Machine. Whole problem is divided into a certain number of classes in classification. Different classification methods (Artificial Neural Network, Support Vector Machine, Naïve Bayes) have been applied. In this ELM achieved higher accuracy i.e., 95.34% by using 6 different activation functions.

Ram Basnet et. al. [16] proposed the detection of phishing attacks by using machine learning models. In this, 16 features of phishing are used on 6 different machine learning models for detecting, i.e., either phishing or legitimate. Support Vector Machine (SVM) gives the best results, whereas Biased SVM and Artificial Neural Network give the same accuracy i.e., 97.98%.

Kamal et. al. [17] proposed the use of machine learning for phishing detection with features extracted from the URL only. In 2014, according to APWG, increasing phishing attacks due to cheapness & freeness of domain name. Naïve Bayes algorithm is used for the classification of phishing websites on Weka Platform. Using the ensemble methodology an accuracy of 97.08% can be achieved using Stacking,
Bagging and Boosting along with the Naive Bayes, Decision Tree & Random Forest algorithm.

Baykara et. al. [18] proposed a software "Anti Phishing Simulator" for detecting phishing websites which contains malicious software and links and spam email by examining the mail contents. Also prevent serious threats like catches malicious email arriving at email addresses integrated into the system, providers a URL based control. As a result, Bayesian classification provides the weights of the words are calculated & spam word count are made.

Priya et. al. [10] proposed the ant colony optimization algorithm. Detection of phishing websites can be easily done by machine learning. In this proposed system, phishing websites features are extracted and to reduce features it is given to the ant colony optimization algorithm. Again, Naïve Bayes is used to reduce the features and classifies webpage.

Priyanka et. al. [15] proposed phishing detection using feature extraction based on machine learning. They used the Adaline algorithm and Backpropagation algorithm along with SVM to enhance the detection rate and classification of web pages. For better result Adaline is compared with SVM with 99.14%. Minimal time taken by Adaline network when compared with the Backpropagation network with SVM.

Mustafa Kaytan et. al. [19] proposed an Extreme Learning Machine classification algorithm to detected phishing webpages. In this paper, the classification of phishing website features based on "Request URL" and "Website Forwarding". For evaluating performance 10-cross fold validation is used. The average classification accuracy was 95.05% and the best classification accuracy was 95.93%.

Amani et. al. [6] proposed the Random Forest algorithm to detect phishing websites. Random Forest technique is used for better performance as it gives high accuracy of 98.8% with the combination of 26 phishing websites features.

Xiang et. al. [20] proposed CANTINA+ which is the upgrade of CANTINA. False Positive (FP) and achieve runtime speedup reduces features. At CANTINA more features are applied and machine learning techniques are also applied in which 92% True Positive (TP) and0.4% (FP).

Weina Niu et. al. [21] proposed a model Cuckoo- Search SVM(CS-SVM) for the detection of phishing of email with high accuracy. It improves the classification accuracy. To construct hybrid classifier, 23 features are extracted by using CS-SVM. In hybrid classifier, Cuckoo-Search (CS) combined with Support Vector Machine to enhance the parameter selection of Radial Basis Function (RBF). In this, it calculates higher accuracy than the SVM classifier by using RBF. Using CS-SVM classifier accuracy of 99.52% is obtained.

Ishant et. al. [14] proposed various machine learning techniques that are applied to the URL to check whether the website is phishing or legitimate. 30 attributes of phishing websites are considered for detection using Python. For accuracy calculation, the Generalized Linear Model (GLM) and Generalised Additive Model (GAM) are used. To gain more accuracy it uses a Decision tree, Random Forest. An accuracy of 98.4% is obtained by using Random Forest.

Taware et al. [22] proposed an MCAC which differentiate phishing websites from legitimate websites in which result of MACC algorithm is better than other data mining algorithm.

Amirreza et.al [23] proposed the PhishMon framework for detecting phishing webpages. It distinguishes legitimate and phishing webpages with high accuracy. In this paper, PhishMon detects 95.4%.

Yadollahi et al [24] proposed a real-time anti-phishing system. Online and feature-rich machine learning technique distinguish webpages. Approaches are extracted from discriminative features. The solution is based on the client-side, there is no service from the third party.

Gutierrez et. al. [25] proposed SAFE-PC (Semi-Automated Feature generation for Phish Classification) for detecting whether the webpage is phishing or legitimate.

Yang et. al [26] proposed multidimensional feature phishing learning (MFPD). This threshold is fixed for reducing the time. The highest accuracy is 98.61% by using CNN-LSTM.

## VI.PERFORMANCE EVALUATION FOR PHISHING DETECTION

Detection of phishing websites is a binary classification problem. True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are categories in which webpage falls [27].

Table2: Classification Confusion Matrix

| | Classified as Phishing | Classified as Legitimate |
|---|---|---|
| Phishing | NP⟶P | NP⟶L |
| Legitimate | NL⟶P | NL⟶L |

NP ⟶ total number of phishing websites
NL ⟶total number of legitimate websites [28]

❖ Performance evolution parameters are as following:

- **True Positive (TP):** Number of correctly classification of phishing website [28]:
  TP= (NP→P)/NP

- **True Negative (TN):** Number of correctly classification of ham websites [28]:

  TN= (NL→L)/NL

- **False Positive (FP):** The number of ham websites wrongly classified [28]:

  FP= (NL→P)/NL

- **False Negative (FN):** Number of wrongly classification of phishing websites [28]:
  FN=(NP→L)/NP

❖ Measures used for classification of webpages:

- **Precision:** Percentage of correct positive predictions is defined in precision.

  Precision=TP/ (TP+FP)

- **Recall:** Percentage of positive prediction of positive labeled instances [29].

  Recall=TP/ (TP+FN)

- **Accuracy:** Percentage of correct prediction is defined in accuracy [29].

  Accuracy= (TP+TN)/(TP+TN+FP+FN)

- **F-Score:** It measure weighted average of true positive rate/recall and precision [29].
  F=2*(precision*recall/precision+recall)

## VII. CLASSIFICATION OF PROTECTION APPROACHES AGAINST PHISHING ATTACKS

### 1. Network level protection

Blacklist features which are also known as Network-level protection. Set of IP addresses or domain name to enter in a network does not allowed by network layer protection for implementation. It blocks the communication from those systems, which are identified as phishers. Example: Anti-Spam filters, DNS-Based filter [28].

### 2. Authentication

Authentication, in which it ensures or to verify the identification. It applies on both the user level and server/domain level to check whether the message is sent by trusted domain or not. Security communication is increased at both levels i.e., user and the domain level. Password is authenticated at the user level, but it can be easily broken by the phishers. Service providers ensured domain level. Domain level authentication examples are: Microsoft sender ID, Yahoo-based domain key [28].

### 3. Client-side tools

Client-side tools studied the phishing and attack initiate from the detecting phishing "Web browsers" directly. Another technique is domain check, URL examination, page content, etc. Blacklisting and whitelisting depend on these tools. Detection is failed for zero-day attack is the limitation of these methods.

### 4. Server-side filters and classifiers

It is considered to fighting zero-day attacks. By approval of statistical classifier for identification of weblinks, i.e., either phishing or legitimate is also trained on machine learning algorithm.

### 5. User Education

In today's generation phishing can be easily done by phishers due to lack of awareness about phishing. [30].

## VIII. ISSUES AND CHALLENGES

In previous work, for phishing attacks there are many solutions have been designed. No result is a "bullet of silver" for phishing [31]. Whenever researchers give solution for detecting and recovering phishing sites, then phishers breaks their solutions for fraudulent attempt. So, it is a rigid race between researcher and phisher.

Phishing attack is more successful due to lack of awareness about phishing. Therefore, one of the main challenges is the security, i.e., how to encourage users to protect themselves against phishing.

## IX. CONCLUSION

Internet is one of the most targeted phishing attacks, so the anti-phishing is used for protection. There is various defense technique for phishing. Better defense mechanism has adequate to identify phishing attack with low false positive (FP) [30]. It is a survey in which the machine learning algorithm was able to detect with approximate 99% accuracy by including a combination of 30 features.

Phishing detection techniques inform the users whether it is phishing, suspicious or legitimate websites.

## REFERENCES

[1] Y. Sonmez, Turker Tuncer, Huseyin Gokal & Engin Avci (2018). "Phishing web Sites Features Classification Based on Extreme Machine Learning". 6th International Symposium on Digital Forensic and Security (ISDFS).

[2] Kay, R (2004) Sidebar: The Origins of Phishing. [Online]. Available:http://www.computerworld.com/s/article/89097/Sidebar_The_Origins_of_Phishing.

[3]Mohmoud khonji, Youssef Iraqi and Andrew Jones(2013)."Phishing detection: A Literature Survey". IEEE communications Systems and Tutorials, pp (99):1-31.

[4]Anti-Phishing Working Group, "Phishing Activity Trends Report," 2019.

[5] RSA Online Fraud Report 2019.

[6]Eduardo Benavides, Walter Fuertes, Sandra Sanchez & Manuel Sanchez(2019)."Classification of Phishing Attack Solutions by Employing Deep Learning Techniques: A Systematic Literature Review'. Smart Innovation, Systems and Technologies, vol 152. Springer, Singapore.

[7]Alswailem, A., Alabdullah, B., Alrumayh, N., & Alsedrani, A. (2019)"Detecting Phishing Websites Using Machine Learning".2nd International Conference on Computer Applications & Information Security (ICCAIS)

[8]Himani Thakur & Supreet kaur(2016)."A Survey Paper On Phishing Detection". International Journal of Advanced Research in Computer Science(IJARCS). ISSN: 0976-5697.

[9]Kathrine, G. J. W., Praise, P. M., Rose, A. A., & Kalaivani, E. C. (2019). "Variants of phishing attacks and their detection techniques". 3rd International Conference on Trends in Electronics and Informatics.

[10] R.Priya (2016), "An Ideal Approach for Detection of Phishing Attacks using Naive Bayes Classifier". International Journal of Computer Trends and Technology(IJCTI). ISSN: 2231-2803.

[11]Arun Kulkarni, Leonard L.. "Phishing Websites Detection using Machine Learning", International Journal of Advanced Computer Science and Applications, 2019

[12] Aron Blam, Brad Wardman, Thamar Solorio and Gary Warner (2010)," Lexical feature based phishing URL detection using online learning", 3rd ACM Workshop on Security and Artificial Intelligence.

[13]Rami M. Mohammad, Fadi Thabtah & Lee McCluskey "Phishing Websites Features".(2014)

[14]Tyagi, I., Shad, J., Sharma, S., Gaur, S., & Kaur, G. (2018)."A Novel Machine Learning Approach to Detect Phishing Websites". 5th International Conference on Signal Processing and Integrated Networks (SPIN).

[15] Singh, P., Maravi, Y. P. S., & Sharma, S. (2015). "Phishing websites detection through supervised learning networks". 2015 International Conference on Computing and Communications Technologies (ICCCT)

[16]Basnet, R., Mukkamala, S., & Sung, A. H. (n.d.). Detection of Phishing Attacks: A Machine Learning Approach. Studies in Fuzziness and Soft Computing, 373–383.

[17] Gyan Kamal and Monotosh Manna, Detection of Phishing Websites Using Naive Bayes Algorithm, Proceeding of International Journal of Recent Research and Review, Vol. XI, Issue 4 December 2018, ISSN 2277-8322.

[18]Baykara, M., & Gurel, Z. Z. mm(2018). Detection of phishing attacks. 2018 6th International Symposium on Digital Forensic and Security 355389(ISDFS).

[19] M. Kaytan and D. Hanbay "Effective classification of Phishing Webpages Based on New Rules by Using Extreme Machine Learning" Anatolian Journal of Computer Sciences, AJCS 17, pp: 15-36, ISSN: 2548-1304, 2017.

[20] Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011). CANTINA+A Feature-Rich Machine Learning Framework for Detecting Phishing Web Sites. ACM Transactions on Information and System Security, 14(2), 1–28.

[21] Niu, W., Zhang, X., Yang, G., Ma, Z., & Zhuo, Z. (2017). Phishing Emails Detection Using CS-SVM. 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC).

[22] Wa'el Hadi, Faisal Aburub, Samer Alhawari. "A new fast associative classification algorithm for detecting phishing websites", Applied Soft Computing, 2016.

[23] Niakanlahiji, A., Chu, B.-T., & Al-Shaer, E. (2018). PhishMon: A Machine Learning Framework for Detecting Phishing Webpages. 2018 IEEE International Conference on Intelligence and Security Informatics (ISI).

[24] Yadollahi, M. M., Shoeleh, F., Serkani, E., Madani, A., & Gharaee, H. (2019). An Adaptive Machine Learning Based Approach for Phishing Detection Using Hybrid Features. 2019 5th International Conference on Web Research (ICWR).

[25] Gutierrez, C., Kim, T., Della Corte, R., Avery, J., Cinque, M., Goldwasser, D., & Bagchi, S. (2018). Learning from the Ones that Got Away: Detecting New Forms of Phishing Attacks. IEEE Transactions on Dependable and Secure Computing.

[26] Yang, P., Zhao, G., & Zeng, P. (2019). Phishing Website Detection based on Multidimensional Features driven by Deep Learning. IEEE Access.

[27] Khonji, M., Iraqi, Y., & Jones, A. (2013). "Phishing Detection: A Literature Survey". IEEE Communications Surveys & Tutorials, 15(4), 2091–2121.

[28] Almomani, A., Gupta, B. B., Atawneh, S., Meulenberg, A., & Almomani, E. (2013). "A Survey of Phishing Email Filtering Techniques". IEEE Communications Surveys & Tutorials, 15(4), 2070–2090.

[29] Asha S Manek, D K Shamini, Veena H Bhat, P Deepa Shenoy, M. Chandra Mohan, K R Venugopal, L M Patnaik. "ReP-ETD: A Repetitive Preprocessing technique for Embedded Text Detection from images in spam emails", 2014 IEEE International Advance Computing Conference (IACC), 2014

[30] A.MahaLakshmi, N.Swapna Goud, G.Vishnu Murthy (2018)."A Survey on Phishing And It's detection technique Based on support Vector Machine (SVM) and Software Defined Networking (SDN)". International Journal of Engineering and Advanced Technology (IJEAT). ISSN: 2249-8958.

[31] Gupta, B. B., Aakanksha Tewari, Ankit Kumar Jain, and Dharma P. Agrawal. "Fighting against phishing attacks: state of the art and future challenges", Neural Computing and Applications, 2016.