IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IN THE COVID-19 PANDEMIC PHISHING WEBSITE DETECTION AND PREVENTION

[1]Amit Mahajan, [2]Priyanka Gupta

[1]Student, [2]Assistant Professor

[1]School of Computer Science & Engineering,

[1]Lovely Professional University, Jalandhar, Punjab

Abstract: Criminal minds have figured out a means to steal personal information without having to meet the individual and with the least chance of getting discovered. This is referred to as phishing. We will examine the literature review of phishing attacks and give a taxonomy of different forms of phishing assaults in this paper. We've also spoken about the many types of phishing attempts and how the LinkGuard Algorithm can protect you from them. End-Host based Anti-Phishing Algorithm is another name for LinkGuard Algorithm. This algorithm is based on Hyperlink characteristics, and because it is character-based, it can identify and block not only known phishing attacks, but also undiscovered phishing assaults. Our tests showed that LinkGuard is capable of detecting phishing attacks with a low rate of false negatives.

Index Terms **– Phishing, URL classification, blacklists, whitelists, Security, LinkGuard Algorithm.**

I. INTRODUCTION

The primary goal of phishing is to steal data, money, personal information, or sensitive information by using a false or duplicate website. Detection and prevention of phishing assaults [1] is a difficult task, especially when the attacker strikes in such a way that he gets past the existing Anti-Phishing Techniques [2, 3]. This can happen to persons who are well-educated or have a lot of experience. In this case, the attackers create a fake webpage by copying material from a well-known firm, organisation, or bank and making alterations to the original sites, making it impossible for an internet user to distinguish between phishing and legitimate webpages [4]. To avoid phishing assaults, one successful option is to integrate security elements with the online browser, which can raise alarms when the user navigates to a phishing site. In general, web browsers offer protection against list-based solutions, such as White-list or Black-list. The list-based approach matches the provided domain with the domains contained in the Black-list or White-list [5] to make the right conclusion. In 1996, a gang of hackers used the term "phishing" for the first time on the Internet, stealing America Online (AOL) accounts by tricking ignorant AOL users into surrendering their passwords [6]. When a new domain has to be added to the list, either technical professionals or security software check it. To validate the identity, security software examines the webpage's numerous characteristics. According to the Anti-Phishing Working Group (APWG), between July and December 2014, 123,972 distinct phishing attempts were discovered [7]. The following is how the rest of the research paper is organized: The Literature Review of the Phishing Attack is presented in Section II. In Section III, we look at the features of hyperlinks utilized in phishing attempts. The LinkGuard Algorithm is then presented in Section IV. In Section V, we detail our LinkGuard System implementation, and in Section VI, we wrap up the research article.

II. LITERATURE REVIEW

"Phishing Detection Using Link Guard Algorithm" is the title of the study. Link Guard, according to Madhav Prasad Vijay Maheshwari, is not only effective for identifying phishing attempts, but it may also protect users from harmful or unwanted links in Web sites and Instant messaging [8]. V. Suganya gives an overview of numerous phishing attempts and several ways to safeguard information in his work "A Review on Phishing Attacks and Various Anti Phishing Techniques" [9]. Mallikka Rajalingam and Saleh Ali Alomari Putra Sumari's article "Prevention of Phishing Attacks Based on Discriminative Key Point Features of Webpage's" looked at a variety of approaches. The majority of techniques rely on text-based comparisons, which might fail to detect phishing attacks. Images can be found on this website. The most significant feature in a picture is the colour of the image on the page. As a result, they developed an image-based comparison approach in this article, which compares pictures based on colour values. This technique compares web pages based on colour values and produces precise results [10]. In the article "Intelligent Phishing Website Detection and Prevention System by Using Link Guard Algorithm," U. Naresh and U. Vidya Sagar used the generic features of hyperlinks in phishing assaults to build an end-host based anti-phishing algorithm called Link Guard. These features were developed by analyzing the Anti-Phishing Working Group's phishing data repository (APWG). Link Guard can identify both known and new phishing attempts since it is based on the basic features of phishing attacks. 195 of the 203 phishing assaults are successfully detected by Link Guard. Link Guard is likewise light weight, according to this study, and can identify and block phishing attempts in real time [11]. In the paper "E-Mail Security System Using for Phishing Attack-

Using LinkGaurd Algorithm," Kalpana, Naveen Kuma, and Parul Saharavat present Link Guard algorithm Plus, which takes a step forward with the detection of phishing measures by introducing the part in which phishing websites are captured using text reading to find out the malicious and threatening words to be filtered and moved into spam [12]. Monali Deshmukh and Shraddha K. Popat reviewed the number of anti-phishing toolbars and presented a system model to combat phishing attacks in their article "Different Techniques for Detection of Phishing Attack." Different anti phishing strategies are addressed in this work, as well as the link guard algorithm [13]. The article " The Survey Analysis of Phishing Attack Strategies " by Oniyide Sakiru Adelakun, Fatai olawale Waheed, and Awotunde Joseph Bamidele examines the many methods that have been employed on phishing assaults in the future [14].

III. PHISHING ATTACK TAXONOMY

The techniques through which attackers are able to access the User's personal information details may be used to classify phishing attacks. Phisher either defrauds the User or employs harmful code to get access to the User's personal data. Social Engineering and Technical Subterfuge are two fundamental types of phishing assaults.

3.1 Social Engineering

The term "social engineering" refers to the process of luring someone into a malevolent hoax in order to achieve certain objectives that are harmful to the users. According to a global phishing study conducted in 2014 [15,16], Apple was the most targeted brand by phishers. Attackers sent out phishing emails asking users to upgrade their iCloud accounts by clicking on a link that sent them to a fraudulent website where they had to change their credentials [17].

Social engineering is further subdivided into the following categories:

3.1.1 Spoofed Emails

This is also known as Phishing Emails, and it is meant to persuade the user to give his or her personal information. Email phishing may be carried out in a variety of ways:

- Creating a false link by concatenating certain strings at the beginning or end of the lawful domain.
- To the user, the real and visible connections are not the same.
- Redirect a link to a malicious website using bugs.
- Replace specific characters in a lawful URL with comparable but non-identifiable ones.

There are Two forms of spoof email phishing:

3.1.1.1 Spear-Phishing

It is employed against any group of people in an organization that is working. This method will increase the attack success rate and strategies that an attacker may use to discover contextual information [18].

3.1.1.2 Whaling

It is a type of extortion that targets a company's top executives [19]. It's known as CEO Fraud, and it phishes people into handing over their bank passwords, employee information, and other personal information.

3.1.2 Fake Websites

These websites, which have a similar appearance to a real website, are designed to trick users into divulging important information; they are also known as phishing websites.

3.2 Technical Subterfuge

This is also a kind of fraud in which a phisher sends malicious code in the form of an attachment to an email, a website, or self-executable code.

As stated in [20],

[21-25] Technical Subterfuge is a subcategory of Technical Subterfuge.

3.2.1 Cross-site scripting (XSS)

When a dynamic webpage displayed input without adequately validating it, XSS occurred.

3.2.1 Session Hijacking

(Cookies snooping) In WLAN, it's a serious topic. The session key is taken via DoS attacks in order to steal the identity and get unauthorized access to the resources.

3.2.1 Malware Phishing

This is used to save the user's credentials and deliver them to the phisher (owner).

3.2.1 DNS Poisoning

In this case, the phisher creates a phony DNS server to entice the user to contact with it. Once the user connects, they are led to malicious Web pages or may be infected with malware.

IV. PHISHING ATTACK PROCEDURE AND PREVENTION METHODS

We assume that phishers utilize e-mail as their primary technique of phishing assaults in this research work. Our research and methodology, however, may be used to counteract assaults that employ other methods, such as instant messaging.

4.1 The Procedure of Phishing Attacks:

Phishing attacks are typically carried out in the following four steps:

1. Phishers create a fake website that looks identical to the real website, including setting up the web server, applying the DNS server name, and building web pages that seem similar to the target website, among other things.
2. Send a huge number of fake e-mails to targeted individuals in the name of real firms and organisations, attempting to persuade potential victims to visit their websites.
3. Recipients get the e-mail, open it, and provide the needed information by clicking the faked hyperlink in the e-mail.
4. Phishers take personal information from victims and use it to commit fraud, such as transferring money from their accounts.

4.2 Approaches to Prevent Phishing Attacks:

Phishing attacks can be prevented in a variety of methods (technical and non-technical):

1. train users to recognize phishing attempts and to be on the lookout for phishing-alike e-mails;
2. Punish phishing attackers using legal means;
3. thwart phishing attackers through technical means. We will just look at the third one in this study report. Technically, we can effectively block a phishing attempt if we can cut off one or more of the stages required by the assault. We'll go through these techniques briefly in the sections that follow.

4.2.1 Detect and block the phishing Web sites in time

We can ban phishing Web sites and prevent phishing assaults if we can detect phishing Web sites in time. It is very simple to establish whether a site is a phishing site or not (manually), however it is tough to locate such phishing sites in a timely manner. Here are two techniques for detecting phishing sites. 1) A lawful Web site's Web master searches the root DNS for suspicious sites on a regular basis (e.g. www.1cbc.com.cn vs. www.icbc.com.cn).

Because the phisher must reproduce the target site's content, he must utilize tools to (automatically) obtain the target site's Webpages. As a result, this type of download may be detected at the Web server and traced back to the phisher. Both strategies have drawbacks. DNS scanning adds to the cost of DNS systems and may create problems with routine DNS requests; moreover, many phishing attempts do not require a DNS name. To get beyond phishing download detection, smart phishers may easily develop programs that mimic the behaviors of human individuals.

4.2.2 Enhance the security of web sites

The company's operations Bank websites, for example, can utilize innovative techniques to ensure the security of their consumers' personal information. Using hardware devices is one way to improve security. The Barclays bank, for example, equips consumers with a hand-held card reader. Before purchasing on the internet, customers must first insert their credit card into the card reader and enter their (personal identification number) PIN code, after which the card reader will generate a one-time security password, which can only be used once. Another option is to utilize biometrics (such as voice, fingerprint, iris, and so on) to authenticate users. PayPal, for example, attempted to replace the single password verification with voice recognition in order to improve the Website's security. Even after obtaining a portion of the victims' information, phishers are unable to complete their goals using these tactics. However, all of these approaches need the purchase of extra hardware in order to provide user-to-website authentication, which will raise the cost and cause some difficulty. As a result, widespread adoption of these approaches will take time.

4.2.3 Block the phishing e-mails by various spam filters

E-mails are commonly used by phishers as "bait" to entice potential victims. The SMTP (Simple Mail Transfer Protocol) protocol is used to send and receive e-mails over the Internet. It's a very basic protocol that doesn't have the essential authentication methods. In SMTP, information about the sender, such as the sender's name and email address, the message's route, and so on, can be faked. As a result, the attackers may send out a huge number of faked e-mails that appear to be from real companies.

Because phishers conceal their identities while sending faked e-mails, phishing attempts will be drastically reduced if anti-spam systems can detect whether an e-mail is sent by the declared sender (Am I Whom I Say I Am?). From here, phishing attempts may be effectively defeated using methods that prohibit senders from forging their Send ID (e.g., Microsoft's SIDF).

SIDF is a mix of Microsoft's Caller ID for E-mail and Meng Weng Wong's SPF (Sender Policy Framework). Both Caller ID and SPF examine the domain name of the e-mail sender to see whether the e-mail is being sent from a server that is permitted to send e-mails for that domain, and if the e-mail is using a faked e-mail address. If it is falsified, the Internet service provider can identify whether or not the email is spam.

Phishers' faked e-mails are an example of spam e-mails. Spam filters can likewise be used to screen phishing e-mails from this perspective. Blacklists, whitelists, keyword filters, Bayesian filters with self-learning abilities, and E-Mail Stamp, for example, can all be applied to e-mail server or client systems. The majority of these anti-spam approaches filter e-mails at the receiving end by scanning their contents and addresses. And, as we'll see below, they all have advantages and disadvantages. If the names of the spammers aren't known ahead of time, blacklisting and whitelisting won't function. Keyword and Bayesian filters can detect spam based on content, allowing for the detection of unknown spam. They can, however, lead to both false positives and false negatives. Furthermore, spam filters are meant to screen generic spam e-mails and may not be particularly effective at filtering phishing e-mails since they do not take into account the unique characteristics of phishing assaults.

4.2.4 Install online anti-phishing software in user's computers

Despite all of the foregoing steps, users are still able to access faked Web sites. Users can install anti-phishing software on their PCs as a final line of defense. There are two types of ant phishing technologies in use today: blacklist/whitelist based and rule-based.

- Category I: When a user accesses a Web site, the ant phishing program checks the site's address against a database blacklist. The anti-phishing application notifies users whether the visited site is on the list. Tools in this area include EarthLink's Scam Blocker, PhishGuard, and Netcraft, among others. Despite the fact that the creators of these tools have all said that they would be able to update the blacklist in a timely manner, they will not be able to prevent assaults from newly discovered (unknown) phishing sites.
- Category II: This group of tools has software that has specific rules and checks the security of a website against those standards. Spoof Guard, created by Stanford, and Trust Watch, produced by GeoTrust, are examples of this sort of technology. SpoofGuard examines a Web site's domain name and URL (which includes the port number), as well as if the browser is led to the current URL via links in e-mail messages. Spoof Guard will alert users if the domain name of the visited Web site is similar to a well-known domain name or if they are not utilizing the usual port. The security of a Web site is assessed in TrustWatch by whether it has been reviewed by a trustworthy third-party organization. SpoofGuard and TrustWatch both provide a browser toolbar that informs users if a website is validated and trusted. It is clear that all of the aforementioned defensive techniques are valuable and complimentary to one another, but none of them is ideal at this time. We focus on end-host based approaches in the rest of the research article and present an end-host based LinkGuard algorithm for phishing detection and prevention. Our work varies from others in the following ways:

1) LinkGuard is based on a thorough examination of the features of phishing hyperlinks, whereas Spoof Guard is more of a framework;
2) LinkGuard has a proven very low false negative rate for unknown phishing assaults, but Spoof Guard's false negative rate is unknown. In the next part, we'll look into the features of hyperlinks in phishing emails before proposing the LinkGuard algorithm.

V. LINKGUARD ALGORITHM

Phishers typically try to persuade potential users to click the hyperlink contained in the phishing e-mail by unlawfully collecting useful information from them. The LinkGuard Algorithm operates by evaluating the discrepancies between the visible link and the actual link. This Algorithm also calculates the URI's similarity to a known reliable site [26].

```
vLink: visual link;
aLink: actualLink;
vDNS: visual DNS name;
aDNS: actual DNS name; senderDNS: sender‟sDNS name.
int LinkGuard(vLink, aLink} {
1 vDNS = GetDNSName (vLink);
2 aDNS = GetDNSName (aLink);
3 if ((vDNS and aDNS are not
4 empty) and (vDNS != aDNS))
5 return PHISHING;
6 if (aDNS is dotted decimal)
7 return POSSIBLE_PHISHING;
8 if (aLink or vLink is encoded)
9 {
10 vLink2 = decode (vLink);
11 aLink2 = decode (aLink);
12 return LinkGuard(vLink2, aLink2);     13 }
14 /* analyze the domain name for
15 possible phishing */
16 if(vDNS is NULL)
17 return AnalyzeDNS (aLink);
 }
int AnalyzeDNS (actual link)
{ /* Analyze the actual DNS name according to the blackList and whiteList*/
18 if (actualDNSin blackList) \
19 return PHISHING;
20 if (actualDNSin whiteList)
21 return NOTPHISHING;
22 return PatternMatching (actualLink);
}
int PatternMatching(actualLink) {
23 if (senderDNS and actualDNSare different)
24 return POSSIBLE_PHISHING;
25 for (each item prevDNS in seed_set)
26 {
27 bv = Similarity(prevDNS, actualLink);
```

```
28 if (bv == true)
29 return POSSIBLE_PHISHING;
30 }
31 return NO_PHISHING;
}
float Similarity (str, actualLink) {
32 if (str is part of actualLink)
33 return true;
34 int maxlen = the maximum string
35 lengths of str and actual_dns;
36 int minchange = the minimum number of
37 changes needed to transform str
38 to actualDNS(or vice versa);
39 if (thresh<(maxlen-minchange)/maxlen<1)
40 return true
41 return false;
 }
```

5.1 Steps of LinkGuard Algorithm

1. To begin, go to the hyperlink and extract the DNS name from the real and visible links, then compare the actual and visible DNS names. If they differ, it is phishing.
2. It is also a phishing attempt if a dotted decimal IP address is directly used in genuine DNS
3. If the real or visible link is encoded, we decode it first and then call the LinkGuard recursively to get a result.
4. LinkGuard runs Examine DNS to analyse the true DNS when there is no destination information in the visible link.
5. If the real DNS name is in the blacklist, it is phishing; otherwise, it is not a phishing attack; if the actual DNS name is not in either the blacklist or the whitelist, pattern matching is used. Pattern matching is a technique that involves categorising hyperlinks from past assaults in order to find new ones. The LinkGuard algorithm is efficient, lightweight, and capable of detecting up to 96% of undiscovered phishing assaults in real time.

VI. Conclusion

The taxonomy of phishing assaults, i.e., Social Engineering phishing, which is based on faked email attacks and phony websites, has been studied and addressed in this study. Another subcategory is Technical Subfurtuge, which is further divided into XSS, Session hijacking, and so forth. We also looked into the characteristics of the URLs included in phishing emails. After that, we created the LinkGuard Algorithms phase, which is based on derived characteristics. The LinkGuard Algorithm will be extended further in the future.

References

[1] "Online Detection and Prevention of Phishing Attacks," Chuanxiong Guo, https://www.researchgate.net, July 16, 2015.

[2] Jyoti Chikara et. al, “The crime of identity theft and the fight against crime of identity theft Strategies: Case Study”, JARCSSE, Volume 3, Issue 5, May 2013.

[3] "Anti-Phishing Techniques: A Review," IJERA, Volume 2, Issue 2, March-April 2012, pp. 350-355. Gaurav, Madhuresh Mishra, and Anurag Jain, "Anti-Phishing Techniques: A Review," IJERA, Volume 2, Issue 2, March-April 2012, pp. 350-355.

[4] “Intelligent Phishing Possibility Phishing Detector,” Rajeev Kumar Shah, Md. Atlab Hossin, and Asif Khan, International Journal of Computer Application (0975-8887), Volume: 148, No. 7, August 2016.

[5] "E-Mail Security System Using for Phishing Attack- Using Link Gaurd Algorithm 2018 IJSRCSEIT | Volume 3 | Issue 6 | ISSN: 2456-3307." Kalpana, Naveen Kumar, Parul Saharavat "E-Mail Security System Using for Phishing Attack- Using Link Gaurd Algorithm 2018 IJSRCSEIT | Volume 3 | Issue 6 | ISSN: 2456-3307.

[6] Understanding and Preventing Phishing Attacks with the Phishing Guide Gunter Ollmann, IBM Internet Security Systems' Director of Security Strategy, 2007

[7] The Anti-Phishing Working Collection is a group of people that work together to combat phishing Antiphishing.org (http://www.antiphishing.org/).

[8] Madhav Prasad (Madhav Prasad) is a "Phising Detection Using Link Guard Algorithm," IJRITS, Vol. 1, Issue 8, July 2014. Vijay Maheshawari, "Phising Detection Using Link Guard Algorithm," IJRITS, Vol. 1, Issue 8, July 2014.

[9] IJCA, Volume-139, No-1, April 2016, V.Suganya, “A Review on Phishing Attacks and Various Anti-Phishing Techniques,” IJCA, Volume-139, No-1, April 2016.

[10] "Prevention of Phishing Attacks Using Discriminative Key Point Features of WebPages," International Journal of Computer Science and Security (IJCSS), Volume (6), issue (1), 2012. Mallikka Rajalingam, Saleh Ali Alomari Putra Sumari, "Prevention of Phishing Attacks Using Discriminative Key Point Features of WebPages," International Journal of Computer Science and Security (IJCSS), Volume (6), issue (1), 2012.

[11] "Intelligent Phishing Website Detection and Prevention System by Using Link Guard Algorithm," IOSR-JCE, Volume 14, Sep-Oct 2013. U.Naresh, U.Vidya Sagar, and C.V.Madhusudan Reddy, "Intelligent Phishing Website Detection and Prevention System by Using Link Guard Algorithm," IOSR-JCE, Volume 14, Sep-Oct 2013.

[12] Parul Saharavat, Kalpana, Naveen Kuma, Kalpana, Naveen Kuma, Kalpana, Naveen Kuma, Kalpana, Na IJSRCS, Volume 3, Issue 6, 2018, published a paper titled "E-Mail Security System for Phishing Attacks- Using Link Gaurd Algorithm."

[13] "Different Techniques for Detection of Phishing Attack," IJESC, Volume 7, Issue No. 4, 2017. Monali Deshmukh, Shraddha K.Popat, "Different Techniques for Detection of Phishing Attack," IJESC, Volume 7, Issue No. 4, 2017.

[14] "The Survey Analysis of Phishing Attack Methods," IJERT, Volume:3, Issue 4, April-2014. Oniyide Sakiru Adelakun, Fatai olawale Waheed, Awotunde Joseph Bamidele, "The Survey Analysis of Phishing Attack Methods," IJERT, Volume:3, Issue 4, April-2014.

[15] Response to a security threat Symnatec. https://connect.symantec.com/blogs/apple-ids-targeted-kelihos-botnet-phishing – campaign On December 14, 2014, I was able to get a hold of this information.

[16] J. Li, J. Li, X. Chen, C. Jia, and W. Lou (2015). In cloud computing, identity-based encryption with outsourced revocation is used. 425–437 in IEEE Transactions on Computers.

[17] APWG is an acronym for the Association of Professional Women's Group (2016). http://www.antiphishing.org/resources/apwg-reports/. Phishing activity trends report.

[18] B. Almomani, B. Gupta, T. Wan, and others (2013). Phishing dynamic evolving neural fuzzy framework for "zero-day" phishing email online detection 3960–3964 in Indian Journal of Science and Technology, vol. 6, no. 1.

[19] B. Srivastava, B. Gupta, A. Tyagi, A. Shamn, and A. Mishra. Srivastava, B., Gupta, B., Tyagi, A., Shamn, A., and Mishra, A. DDoS assaults and defence measures were the subject of a recent survey. Communications in computer and information science, Vol. 203, pp. 570–580, Advances in parallel distributed computing.

[20] M. Khonji, Y. Iraqi, and A. Jones (2013). "A literature survey" to identify phishing. 2091–2121, IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091–2121.

[21] "Intelligent phishing website detection system utilising fuzzy techniques," IEEE conference, Damascus, Syria, pp. 1–6. Aburrous, M. et al.

[22] M. Aburrous et al (2010). Using classification mining methods and experimental case studies, we were able to predict phishing websites. In Seventh International Conference on Information Technology (pp. 176–181), an IEEE conference (pp. 176–181). Las Vegas, Nevada is a city in the state of Nevada.

[23] A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani (2013). A look at phishing email screening methods. 2070–2090 in IEEE Communications Surveys & Tutorials, 15(4).

[24] J. Hong (2012). The current condition of phishing scams. ACM Communications, vol. 55, no. 1, pp. 74–81.

[25] Chuenchujit, T." Ataxonomy Criminal Investigations. University of Illinois at Urbana-Champaign, doctoral degree. "2016.

[26] "An Optimal Approach for Detection and Prevention of Phishing Attacks," Narendra Shekokar, Procedia Computer Science, December 2015.