International Journal of Scientific Research and Engineering Development-– Volume X Issue X, Year

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Phishing: A Common Cyber Threat To Combat

Syed Owais Umair*, Simran Banu H Shirahatti**, Soniya Devi T***, Syed Waseem Ahmed****, Pallavi K V*****

*(CSE, AMCEC, Bengaluru, syedowaisumair@gmail.com)
** (CSE, AMCEC, Bengaluru, simranbanuhs31@gmail.com)
***(CSE, AMCEC, Bengaluru, soniyachaudharyt17@gmail.com)
****(CSE, AMCEC, Bengaluru, syedwaseem1096@gmail.com)
*****(Asst. Prof CSE, AMCEC, Bengaluru, pallavi.vishwanath@gmail.com)

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

Abstract:

Phishing stands out as a prevalent cyber threat today, aiming to illicitly obtain sensitive personal information online by mimicking legitimate websites. This paper conducts a detailed study on phishing, emphasizing its various means and the mechanisms available to counter such attacks. Typically, perpetrators utilize emails, messages, or deceptive websites resembling trusted sources to deceive victims into revealing sensitive data. Detecting and shielding users from phished links remains crucial. Numerous industry solutions like email security gateways, endpoint protection, and URL filtering aid in identifying and mitigating phishing risks, primarily focusing on email phishing while SMS phishing demands more attention, particularly with the surge in mobile banking. This system assesses URLs for potential phishing threats by examining various indicators. It starts with a Phishtank Check to see if the link is flagged as phishing. Then, it analyzes the Website Status and Domain Rank, looking for signs of legitimacy based on response status and traffic volume. Domain Age is considered, with newer domains raising suspicion. Additionally, it scrutinizes URL characteristics such as Shortening, Length, and Depth, as well as checking for HSTS support and IP Presence. These checks collectively help identify potential phishing attempts by detecting anomalies in domain behavior, URL structure, and security features.

Keywords — Phishing, Smishing, Vishing, Cyber Threat, URL Filtering

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

I. INTRODUCTION

Phishing is the most common type of cyberthreat that is currently in use. This article's goal is to provide a quick phishing survey. Phishing is the practice of impersonating websites that a person visits in order to gather sensitive information about them online. This essay also covers the many defense tactics and technologies that can be used to thwart phishing efforts. Hackers frequently use messages, emails, or websites that appear to be from a reputable source to trick victims into divulging important information. Users' links must be recognized, and they must be protected from phishing links. To assist enterprises in identifying and mitigating phishing risks, a number of industry solutions and technologies are available for phishing link detection. gates for endpoint defense, URL security, and email security The term "phishing" is derived from the analogy of "fishing," when cybercriminals attempt to deceive unsuspecting individuals by casting a wide net. Phishing attacks usually concentrate on playing on people's trust and curiosity, often using social

engineering methods to create a false sense of urgency or legitimacy. Phishing can take many different forms. For example, spear phishing involves very targeted assaults against specific individuals or groups, while email phishing involves the sending of bogus emails that appear to be from reputable sources. A common phishing tactic involves making fake websites that seem a lot like the genuine ones in order to trick people into entering their passwords or other sensitive information. Phishers may also employ techniques like vishing, or phishing over phones, or smishing, or phishing over text messages, in order to take advantage of different communication channels. Phishing tactics change in step with technological advancements and cybersecurity protections. Because attackers are always refining their strategies, it can be challenging to identify and mitigate these threats. People and organizations need to apply best practices, remain cautious, and recognize potential red flags in order to avoid falling for these deceptive methods. Campaigns for awareness and education can help achieve these. Phishing link detection, which aims to identify and thwart malicious efforts to deceive users with bogus URLs, is an essential part of cybersecurity. Phishing is a popular cyberthreat in which attackers fabricate connections in an effort to This involves, among other things, looking at a link's structure, domain reputation, and content to distinguish between reliable and risky links. In order to enhance detection capabilities, artificial intelligence and machine learning algorithms are frequently employed. These tools enable systems to identify threats that were previously unidentified and adjust to evolving phishing techniques. In order to identify phishing links, heuristics, behavioral patterns, and real-time analysis are crucial elements. By continuously monitoring connection behavior and contextual data, detection systems are able to identify abnormalities that may indicate a phishing attempt. Additionally, integrating threat intelligence feeds and promoting collaborative data sharing help to create a dynamic security system that can quickly respond to fresh phishing attacks. User education is an essential component of phishing link detection systems since phishing attempts often take advantage of human vulnerabilities. Real-time notifications and meaningful feedback provide an additional layer of defense against phishing attempts by empowering users to make informed decisions about the reliability of links. The present introduction establishes the foundation for a comprehensive understanding of phishing within this framework. It emphasizes the need for proactive steps, technological safeguards, and user awareness to counteract the ever-evolving and enduring nature of phishing attacks in the digital realm. In response to the persistent threat of phishing attempts in this rapidly evolving environment, effective phishing link detection has been developed. The foundation for analyzing the state-of-the-art instruments, procedures, and collaborative tactics is laid forth in this introduction.

II. METHODOLOGY AND APPROACH

Your paper must be in two column format with a space of 4.22mm (0.17") between columns. With the help of SafeSurf, a phishing domain detection software, you can safely browse the website without having to visit it. constructed with Python. This can help you recognize dubious websites fast and defend against phishing scams. These are the features that users of SafeSurf can access. Anyone may easily traverse the website's simple interface and use it. Users don't need to visit the website to watch the preview. SafeSurf assigns a trust score to the URL, enabling the user to assess the legitimacy and trustworthiness of the site. To determine whether the URL is a reported phishing link, it is cross-referenced with a phishing database (PhishTank). SafeSurf offers important domain information (WHOIS, SSL, and general), which will aid the user in comprehending the URL in its entirety. The following external libraries are imported by the file: A tiny web framework called Flask is used to create web applications. requests: An HTTP request library used to contact external

sites. A package called BeautifulSoup is used to parse XML and HTML pages.A function called urljoin can be used to join relative and absolute URLs. Responds to POST and GET queries. Parses the input URL before sending it for evaluation to the `controller.main} function. Displays the evaluation result on the `index.html` template. Responds to POST and GET queries. Uses the requests library to retrieve the input URL's HTML content. Renders the HTML content in a prettier format for the `source_code.html` template. The primary controller for determining a URL's trustworthiness is the `controller.py} file. It coordinates a number of investigations and computations to ascertain a URL's trustworthiness. The code's operation is explained in full below:Worldwide Variables: {BASE_SCORE}: The default trust score (out of 100) for a URL is stored in this variable. The initial value is 50. {main(url)}: This function serves as the URL assessment's starting point. After receiving a URL as input, it carries out the following actions:1. **Input Validation**: Using the `include_protocol()` method from the `model} module, the URL is validated and prepared to include the protocol (HTTP or HTTPS). 2. **Default Data Initialization**: The domain name of the URL is obtained by extracting it using `tldextract}. The input URL and a 'SUCCESS' status are initialized in a default response dictionary. The base score is used to initialize the trust score. 3. **URL Evaluation**: Phishtank Check: This verifies whether the URL appears in the Phishtank database as a phishing link.Website Status: The `validate_url()` function from the `model} module is used to ascertain the website's response status.Domain Rank: The `get_domain_rank()` method from the `model} module is used to acquire the domain rank. Real websites usually receive a lot of traffic, which is a sign of popularity and dependability. Domain Age: It gets the age of the domain from WHOIS information and provides it in the answer. WHOIS information is used to calculate a domain's age. Websites that are under two years old could cause scrutiny.

URL Shortening: Look for the usage of URL shortening services, as they may be used maliciously and mask the original URL. HSTS Support: This verifies whether HTTP Strict Transport Security (HSTS) is supported by the website. Check if HSTS and HTTPS are supported by the domain. Secure domains with HSTS support are frequently given priority. IP Presence: It determines whether an IP address connected to the domain is present. IP addresses may be used in place of domain names in phishing URLs. Dedicated domain names are usually associated with authentic domains. URL Redirects: It looks for redirects to other URLs. Hiding the original phishing link can be accomplished by redirecting users to different pages. Length of URL: Determines if the URL is too long. Phishing efforts may be indicated by URLs longer than 75 characters, as the attacker may attempt to hide questionable content in the address bar. URL Depth: This function determines if the URL is too deep. Serious websites tend to have simpler URL structures, so an excessive use of '/' in the URL structure makes one suspicious. IP Address and SSL Certificate: - It obtains and includes in the answer the IP address linked to the domain. The `get_certificate_details()` function from the `model` module is used to acquire SSL certificate details. Calculation of Trust Score: - The assessments' outcomes are used to determine the trust score. Response Generation: - The evaluation findings, including the trust score, are returned in a response JSON format.

Error Handling: An error response containing the specifics of the error is generated in the event that any exceptions are detected during the assessment process. By performing a number of calculations and tests, this function offers a thorough evaluation of the input URL's reliability, storing the results in a response dictionary. Reading Data from CSV: - The data is retrieved from the `top-1m.csv` file, which includes a ranking of the top 1 million websites. Installing Dictionary and Arrays: The two data structures are filled by it: {domain_data_array}:

The domain names taken out of the CSV file are listed in this list. {domain_data_dict}: This dictionary associates a rank with every domain name.
Sorting: To get a sorted list of domain names, the `domain_data_array} is sorted alphabetically. Clearing Existing Files: - It removes everything from the `domain-rank.json` and `sorted-top1million.txt` files that are currently there. The sorted domain names are written to the `sorted-top1million.txt` file in step five, Writing Data to Files. The domain-rank dictionary is written in JSON format to the `domain-rank.json` file.

Attackers that use phishing techniques use a variety of techniques to trick people and institutions into disclosing private information. These strategies are frequently distinguished by the use of social engineering methods, trust manipulation, and psychological manipulation. These are a few typical phishing techniques: Email-based fraud: Phishing emails: Cybercriminals send out emails that seem to be from reliable sources, such banks, governments, or well-known businesses. Frequently, the emails compel the recipients to click on links or submit sensitive information because they include urgent messages. Phishers may pose as coworkers, managers, or other reliable connections inside a company in order to fool staff members into disclosing private information or carrying out certain tasks. Piercing Phishing: Targeted Attacks: Spear phishing entails extremely focused and customized attacks, in contrast to generic phishing. Attackers increase their chances of success by obtaining knowledge on particular people or groups in order to develop messages that are convincing and targeted to their targets. Whaling: including spear phishing, whaling focuses on prominent figures in an organization, including CEOs or executives, in an attempt to obtain sensitive company data. Voice phishing, or vishing: Phone calls: Phishers pretend to be reputable organizations, such banks or government institutions, by using voice communication. They frequently instil a sense of urgency in victims, leading them to provide private information over the phone. SMS phishing, or smishing: Text Messages: Phishing message senders use short codes to convey links or instructions to click on links to websites that are dangerous or reveal personal information. They do this by impersonating reputable companies. Medicine: DNS spoofing: When a user enters the right website address, attackers employ DNS spoofing to trick visitors into visiting counterfeit websites. Because of this, people may think they are on a trustworthy website when, in reality, they are on a phishing website. Adverse Attachments: Infected Files: Phishers can use harmful files as attachments to emails, taking advantage of software flaws to install malware on the recipient's computer.

III. APPLICATIONS

Variety of industries, assisting businesses and individuals in safeguarding private data and reducing the dangers of phishing scams. These are a few important uses: Email Protection: Inbound Email Filtering: Email security systems can incorporate phishing link detection to check incoming emails for harmful links. This helps stop consumers from clicking on phishing emails' misleading URLs. Web Browsing Protection: Browser Extensions: When browsing, users can install plugins or extensions for their browsers that use phishing link detection to alert them to potentially dangerous websites. These technologies provide an additional line of defense against phishing scams. Endpoint Security: Anti-Malware Products: To find and stop malicious URLs that might be a part of phishing campaigns, phishing link detection is frequently incorporated into anti-malware and endpoint security products. Phishing link detection is included into network security architecture to evaluate and filter out dangerous URLs, preventing them from reaching internal systems. Network security is facilitated by firewalls and intrusion detection systems. Platforms for Information Sharing and Collaborative Threat Intelligence: Businesses can participate in and gain from these platforms, which allow industry peers to

share knowledge into phishing link detection. By using a collective defensive strategy, the entire cybersecurity posture is improved. Cloud Security: Cloud-based Email Security: When it comes to cloud-based email platforms in particular, cloud email security services can use phishing link detection to shield users from dangerous links in emails. Mobile Security: Mobile Security Apps: To safeguard users against phishing attempts via text messages (smishing) or other mobile communication channels, mobile security applications employ phishing link detection to locate and stop malicious URLs. Systems of Authentication: Phishing link detection is a useful addition to multi-factor authentication (MFA) systems as it stops hackers from deceiving users into divulging their login credentials by means of phony websites. Teaching Resources: Security Awareness Training: To help consumers identify and steer clear of phishing efforts, security awareness training programs include phishing link detection. monetary services Online Banking Security: In order to shield consumers from fraudulent websites trying to get login credentials or personal data, phishing link detection is essential in the financial industry. Social Networking Sites: Account Security: To safeguard users against phishing attempts that stem from social engineering techniques, social media platforms employ phishing link detection to locate and stop fraudulent links published on their networks. By integrating phishing link detection into all of these apps, you can protect confidential data, build a strong defence against phishing attempts, and lower your chance of suffering financial losses, data breaches, and other negative outcomes.

IV. FUTURE RESEARCH DIRECTION

Further investigation into phishing link detection is required in order to stay abreast of evolving cyberthreats and enhance the effectiveness of security measures. Some potential study directions are as follows: Deep Learning and Neural Networks: Analyze how deep learning techniques and neural networks are used to identify phishing links. Investigate how more complex neural network topologies could improve the accuracy of distinguishing legitimate from fraudulent URLs. Explainable AI's Application to Phishing Detection Give methods for making AI-driven phishing link detection more understandable. Understanding how AI models decide can boost trust and make it simpler to identify false positives or negatives. Behavior analysis and user profile creation: Look through state-of-the-art techniques for behavioral analysis to identify subtle patterns in user behavior that may indicate phishing schemes. Build user profile models to gain a better understanding of typical behavior and quickly identify irregularities. Human-Cantered Approaches: Examine human-centric approaches to phishing link detection, incorporating concepts from cognitive science and psychology to build systems that consider how users perceive and interact with online information. Dangers posed by Zero-Day Phishing: Keep an eye out for zero-day phishing attempts, which prey on undiscovered vulnerabilities. Analyze methods for spotting and thwarting novel phishing techniques before they become popular. Cross-Platform Lookup: Investigate approaches for identifying phishing links across many platforms, ensuring that detection strategies function on social media, messaging apps, and email. Machine Learning with Adversaries: Investigate ways to fortify phishing detection systems against malicious attacks. Analyze methods for identifying and foiling hostile attempts to alter or evade machine learning models. Examining Shifting Approaches: Evaluate and adapt to phishers' evolving tactics frequently. To quickly identify and stop new phishing attempts, research methods such altering the structure of URLs, obfuscation techniques, and social engineering tactics are required. Examine the potential applications of blockchain technology in URL database security and threat intelligence sharing. Consider how blockchain technology could improve the data integrity and transparency in joint military projects. Impact of Quantum Computing: Investigate whether quantum computing could have

an impact on the phishing detection systems. Look at developing cryptography methods that can withstand quantum advancements to protect detection protocols against future computer breakthroughs. Methods for Safeguarding Privacy: To safeguard user privacy and effectively identify and neutralize phishing risks, provide privacy-preserving techniques for identifying phishing links. IoT Safety: To reduce potential security vulnerabilities in Internet of Things (IoT) networks and devices, investigate ways to integrate phishing link detection algorithms into IoT security frameworks. Human-in-the-Loop Techniques: Analyze the effectiveness of human-in-the-loop methods, which combine human experience with machine learning algorithms to improve the overall accuracy of phishing link detection. Further study in these areas will eventually improve the cybersecurity landscape by helping to create phishing link detection systems that are more resilient.

V. CONCLUSION

In conclusion, phishing link detection shields individuals and businesses from the crafty tactics employed by bad actors, making it a vital tool in the ongoing battle against cyber threats. Because phishing attack methods are always evolving along with the digital landscape, it is imperative that detection systems be both intelligent and adaptable. The continuous advancement of phishing link detection technologies is essential for navigating the dynamic and intricate realm of cybersecurity. Building up defenses and countering the pervasive threat of phishing attacks necessitates a holistic approach that includes collaborative efforts, user education, and technological innovation. As researchers and practitioners work together to address these problems, more resilient and adaptable solutions seem to be in store for the ongoing fight against cyberthreats. Consider the following tips to protect yourself from phishing scams: Have Doubts: Avoid responding to unexpected emails, especially if they request sensitive or immediate action. Verify the Communications: Use the company's official methods to contact them and confirm the authenticity of any email or correspondence that seems suspicious. Check the URLs: Hover your cursor over links in emails to see the complete URL before clicking. Verify that the domain names match the official website. Use multi-factor authentication, or MFA: Enable MFA whenever you want to add an extra degree of security to your accounts. Stay Up to Date: Stay informed on the latest threats and common phishing techniques. Businesses and individuals can reduce their susceptibility to phishing efforts by implementing robust cybersecurity protocols and exercising caution.

REFERENCES

[1] Identification of Phishing Attacks using Machine learning algorithms Dinesh P, Mukesh M, Navaneethan B E3S web Conf. Volume 399,2023 International Conference on Newer Engineering Concepts and Technology (ICONNECT-2023)

[2] Targets of phishing attacks: A bigger fish to fry Mirjana Pejic Bach, Tanja Kamenjarska, Bersilav Zmuk International Conference on industry sciences and computer sciences innovation. Procedia Computer Science 204(2022) 448-455

[3] A systematic literature review on phishing website detection techniques Asadullah Safi, Satwinder Singh Journal of King Saud University- Computer and Information Sciences 35(2023) 590-611

[4] Buber, E., Demir, Ö., & Sahingoz, O. K. (2017). "Feature selections for the machine learning based detection of phishing websites". In 2017 international artificial intelligence and data processing symposium (IDAP) (pp. 1-5). IEEE.

[5] Creese, S. W. H. Dutton, P. Esteve-González & R. Shillair (2021) "Cybersecurity capacity-building: cross-national benefits and international divides", Journal of Cyber Policy 6 (2): 214-235.

[6] Eurostat (2017). "Community survey on ICT usage in households and by individuals". Available at: https://ec.europa.eu/eurostat/statistics explained/index.php?title=Glossary:Community_survey_on_ICT_usage_in_households_and_by_individuals Farhadi, R. I. (2012). "Information and Communication Technology Use and Economic Growth. PLoS ONE, 7(11).

[7] Filippini, M, and Lester C. H.. (2011) "Energy demand and energy efficiency in the OECD countries: a stochastic demand frontier approach." Energy Journal 32 (2): 59–80.

[8] Filippini, M., and Lester C. H. (2012) "US residential energy demand and energy efficiency: A stochastic demand frontier approach." Energy Economics 34 (5): 1484–1491.

