The efficiency of AI for detecting and preventing Phishing Attacks

Shipra Prashantkumar Shah
shah571@uwindsor.ca

Riddhi Himanshu Thakkar
thakka95@uwindsor.ca

Janki Purohit
purohitj@uwindsor.ca

Nidhi Patel
patel8p5@uwindsor.ca

Nirali Amrutiya
amrutiyn@uwindsor.ca

Department of Computer Science - Master of Applied Computing
The University of Windsor,
Windsor, CA

Abstract– With the increase in the use of technology, there is a surge in cyber threats. Cyber-attacks are becoming one of the vital threats to all organizations. There are a number of attacks happening on data, however, phishing is one of them. All IT companies worldwide are looking for ways to control this cyber threat to protect their data. This document focuses on how AI helps detect phishing attacks and how it prevents phishing attacks using machine learning algorithms. The document discusses how the COVID pandemic brought a surge in phishing attacks, the classification of phishing attacks using AI systems approaches such as PhishHaven discussed in detail, and machine learning algorithms used by AI systems for detecting this attack.

Index Terms– Phishing attack, Detection, classification, artificial intelligence, cyber threat, machine learning algorithms.

I. Introduction

To retrieve the password and banking information of the individual, fraudulent messages through email, appear to be from some esteemed company and are addressed to be a phishing attack. There are many kinds of phishing attacks, spear phishing, vishing, and smishing [1].

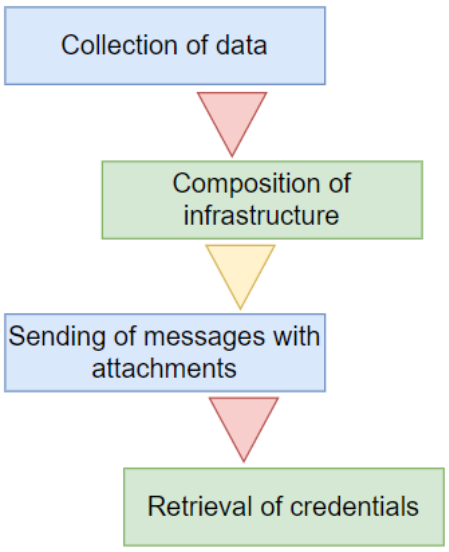

Fig 1. Steps for Phishing Attack

The above figure describes the steps that are performed by the attackers to perform the Phishing attack on the user accounts.

However, there is a solution to this problem, and that is Artificial intelligence which performs many kinds of tasks to prevent this kind of attack. Most of the detection of algorithms is done using machine learning methods and deep learning methods. There was a question about how phishing attacks are detected, so this survey paper describes three methodologies (1) URL-based approach, (2) HTML phishing detecting, and (3) Visual Similarity Based Approach. Along with that, also includes details of PhishHeavan- how to detect the machine-generated URL and human-generated URLs for detecting phishing attacks[1].

II. Effects of Phishing Attacks during COVID-19

According to the author Nithin Valiyaveedu, there were about 114,702 episodes of phishing attacks, whereas they got doubled with the number 241, 324 episodes during the COVID-19 pandemic[1]. There were multiple sectors that became victims of this cyber threat during covid-19, some of them are healthcare systems, financial institutes, government, and media channels. Financial Services and Institutes were under a serious threat because most of the transactions were shifted to online platforms because of shutdowns around the cities and countries as well. This situation was an advantage for cyber attackers. People become victims of social engineering, as hackers appear to be the honest person in acquiring confidential information [2]. Through the below graph, it can be observed how phishing website usage has escalated in the span of 1 month from Feb 2020 to March 2022. Almost it reached from 1000 to 7000.

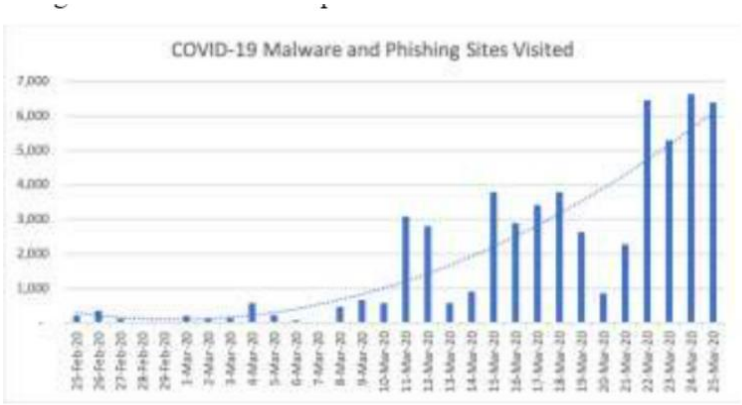

Fig 2. Malware and Phishing websites traversed [2]

This can be considered a result of people being nervous, frightened, and panicked about the Covid-19 pandemic. Everyone wanted to have a piece of news regarding what's happening around the world, what are the number of cases, what is the cause of this surge, and how WHO is taking action to get rid of Covid. For getting this information, citizens refer to various websites and open multiple links which was an advantage for hackers, they take advantage of this situation, and generate fake links, and when the user opens them analyzing the legit links. This results in a successful attempt by cyber attackers. [3] The below image shows what kind of messages and types of authenticated links hackers use to use during the pandemic for phishing attacks.

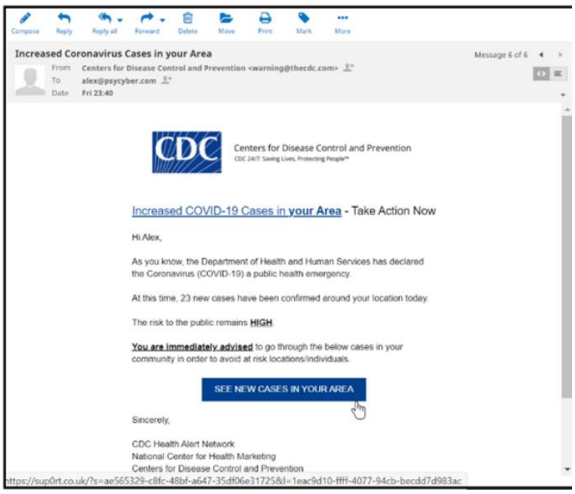

Fig 3. Sample Email for phishing during Covid-19 [4]

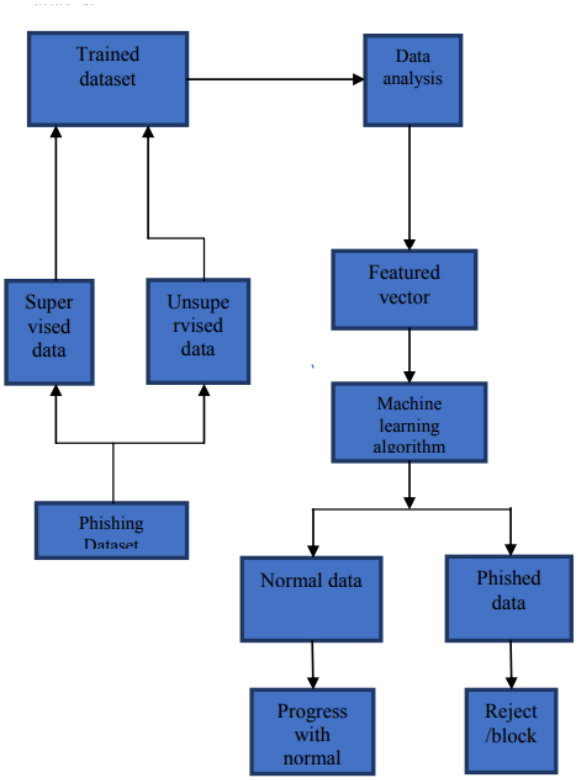

Fig 4. Approach for detecting and preventing phishing attacks [5]

From the above figure, supervised data is utilized to verify the authenticated data whereas unsupervised data is used to get alerted whether there is something suspicious. [5]

III. Actions are taken by AI for preventing Phishing attacks.

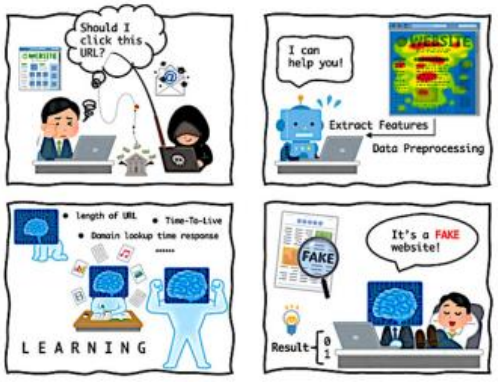

Fig 5. Detection of Phishing by implementing AI solutions [6]

The atop figure demonstrates how AI applies its extracting and data preprocessing features for checking whether the websites are fake or real.

According to the study of Sameen and Hwang, PhishHaven is a real-time AI phishing URL detection system. This method uses lexical analysis for feature extraction, then URL HTML encoding is done to identify the URL. For detecting the small URLs, a special URL hit method is introduced by the mentioned authors. For this method, a voting system is used for the final classification of URLs. Also, this method works parallelly, so double-speed analysis can be done. This method allows estimating the future URLS, with an accuracy of 100% using lexical characteristics. Sameen and colleagues' studies analyzed 100,000 phishing and regular address links, which depicts PhishHaven has a 98% rate of accuracy. [7]

Another study was performed by Shruti and Nagapriya studied the Phishcatch algorithm.

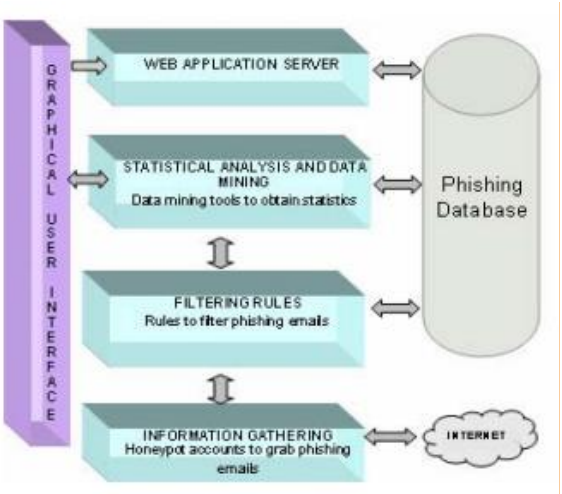

Fig 7. Implementation of the Phishcatch Algorithm [8]

The PhishCatch algorithm is made to identify phishing emails from SMTP and IMAP mail systems. New emails are first fetched from the SMTP server and divided into headers and content. An email's body is HTML-encoded, and the sort of encoding is noted in the "Content-type" field in the email header. The email is then checked for text filters, and the weight of each filter is added to a collection called Phish Tank. The email's received domain mismatch filter is given the proper weight if both entries do not have the same domain. The program then searches the email for all of the links it contains. A link is tested using the link characteristics () method after it is discovered to look for any potential misrepresentation and to determine whether the link is encoded. The link's length is also examined, and if it is found to be longer than a predetermined limit, the proper weight is set for the link filter. Additionally, the associated filter weights are established after checking the number of folders and subdomains in the hyperlinks. Finally, the email's source HTML code is used to retrieve the anchor tag for each hyperlink to compare it to that hyperlink's respective anchor tag and look for discrepancies between the visual link and the link itself. After all the filter weights have been determined, the PhishCatch guidelines are consulted, and the email is labeled as a phishing email. The detected phishing URL is saved, used to collect data, and cross-referenced with data from Phish Tank. By using phishing email, the algorithm gathers data about the phishing attack in general and the phishing link in specific. The kinds of phishing attacks that are currently occurring are analyzed using this data, and future attack types are predicted using this data. The email headers assemble data such as the email's size, its size at the sender's IP address, and the IP location of the phishing website. The alert module retrieves the IP address from the directory and plays a warning to the user not to click on any links in the email. The data center houses and keeps up with all of this information [8].

IV. Analyzing the Phishing attacks with its AI Detection

New technologies constantly run the potential of seriously jeopardizing current cybersecurity. While the intensity of cyberattacks has risen over time, the offensive-defensive technique that has been used is in an endless loop. [9] Even an unidentified attack leaves a trail of network events. New crimes occur in a variety of vulnerabilities to hack the targets. Once they have gained access to your base, they will act in a number of ways as they go through various files, servers, protocols, and systems to fulfill their objectives. And only an AI-driven big-data platform would be able to handle the volume of data required to find them, assess them, and lift the possible issue in a little period of time [9].

Phishing Attacks in Sensitive Areas Present Scenario

It's critical to comprehend how malicious assaults are organized and categorized in order to prevent them. Phishing attacks are a frequent type that frequently poses as reliable sources, tricking consumers with phony Links and character replacement in emails. Understanding these strategies can aid in identifying genuine phony emails.

Classification

Phishing is a kind of cyberattack that targets specific people via email, text messaging, phone calls, social media, fax, and Wi-Fi. Spear-phishing is a type of phishing that targets people explicitly, whereas whaling targets senior executives. Vishing and smishing are two different methods of phishing that use text messages and phone calls, respectively. Social media phishing is used to deceive people into falling for scams

or to collect sensitive information from users. Phishing attempts try to deceive people into divulging private information that hackers could use to their own advantage.

Parties impacted and participants.

Customers, banks, financial institutions, online retailers, telecom and internet service providers, governmental organizations, mail client and web browser suppliers, and law enforcement are all impacted by phishing. To send phony emails from recipients' mailboxes or weak mail provider accounts, phishers exploit compromised computers. To unify players in the fight against phishing, regulatory entities like the Anti-Phishing Working Group and the Messaging Anti-Abuse Working Group exist. The Information Technology Regulations of 2013 must be followed by all service providers, data centers, and intermediaries.

Classification of AI-based Detection Technique

To identify and stop spam and phishing assaults, cybersecurity professionals are using artificial intelligence (AI). AI is a useful tool for detecting these threats due to its speed, precision, and capacity for detailed investigation. AI swiftly assesses an email's information, content, context, and user behavior to determine whether a phishing attempt has been made. This technology aids in safeguarding user data and preventing online threats.[10]

1. Deep Learning for Phishing Attack Detection: Uses unsupervised machine learning algorithms like CNN and LSTM to analyze the image, frame, and text information of a website for detecting phishing sites.

2. Machine Learning for Phishing Attack Detection: Utilizes machine learning algorithms such as logistic regression, decision trees, neural networks, and others to identify phishing websites by analyzing their content and elements.

3. Scenario-Based Phishing Attack Detection: Uses case-based reasoning technique to predict phishing by selecting the most related situation as a solution from past data and adding it to the dataset for future use.

4. Hybrid Learning (hl) Based Phishing Attack Detection: Recommends multiple approaches for detecting phishing attacks by using diverse computer programs or algorithms and experimental studies based on real-world examples. [10]

V. Machine Learning Algorithms

Phishing Detection using machine learning techniques articles by V Shahrivari gives an idea about the supervised method, where machines make verdicts by investigating attributes of phishing and datasets [11]. Web Page information, URLs, and navigation links are detected and extracted from supervised, unsupervised, and semi-supervised learning techniques[11],[12],[13].

3 datasets are being examined for phishing attacks, one is used by the University of California Irvine Machine Learning Repository, Mendeley_2018, and Mendeley_2020[6]. The Mendeley_2018 dataset has forty-eight features, with more than 4000 suspicions and real data [6][14]. For Mendeley_2020, Y. Sekiya states that this dataset contains about 111 features, however, they were divided into eight groups. And finally, UCI, for there are 30 features with 6157 legitimate websites and 4898 malicious websites [6].

Sr.No.	Classifier	Accuracy	Precision	Recall	FI score	Training Time Cost	Test Time Cost
1	Adoboost	93.53	90.73	90.51	90.62	7.373	0.292
2	GBDT	96.33	92.95	93.57	93.26	32.128	0.074
3	HGB	96.54	94.93	95.17	95.17	3.491	0.078
4	LightGBM	96.60	94.90	95.27	95.09	0.742	0.054
5	RF	96.64	95.24	95.83	95.49	7.229	0.462

Table 1. Performance Metrics of ML algorithm [6]

Dataset	Number of instances	Legitimate websites	Phishing websites	Number of features	Types of features	Features extracted from URL	Extra features
UCI_2015	11055	6157	4898	30	Boolean	12	18
Mendeley_2018	100000	5000	5000	48	Hybrid	25	23
Mendeley_2020	88647	58000	30647	111	Hybrid	96	14

Table 2. Comparison between 3 datasets [6]

VI. Future Work

In the future, most cyber-attacks will be controlled using AI techniques and predicted using ML algorithms. Different domains of cyber will be using AI in g the upcoming years. AI will be using automated processes, that will assist in integrating with semi-automate systems, for determining cyberattacks[15]. More anti-phishing approaches will be developed in the coming decade, that will be reliable, flexible, robust, and scalable. Also, they will be able to predict the procedure of phishing attacks [16],[17].

VII. Conclusion

The growing need for technology will lead to innumerable cyberattacks, particularly attacks like phishing because they are easy to perform and generate credentials. Therefore, through various research, it has been concluded that AI will be the most suitable and appropriate tool for detecting and preventing phishing attacks. Tools such as PhishHaven and Phishcatch together with machine learning algorithms will give an exceptional output for saving phishing attacks in the future. Therefore, AI has an inescapable role in resolving the challenges of cybersecurity in the near future with less hassle.

VIII. References:

[1] M. A. Ivanov, B. V. Kliuchnikova, I. V. Chugunkov, and A. M. Plaksina, "Phishing Attacks and Protection Against Them," *IEEE Xplore*, Jan. 01, 2021. https://ieeexplore.ieee.org/document/9396693 (accessed Apr. 23, 2022).

[2] N. A. Khan, S. N. Brohi, and N. Zaman, "Ten Deadly Cyber Security Threats Amid COVID-19 Pandemic," *figshare*, May 12, 2020. https://www.techrxiv.org/articles/preprint/Ten_Deadly_Cyber_Security_Threats_Amid_COVID-19_Pandemic/12278792/1

[3] L. Tawalbeh, F. Muheidat, M. Tawalbeh, M. Quwaider, and G. Saldamli, "Predicting and Preventing Cyber Attacks During COVID-19 Time Using Data Analysis and Proposed Secure IoT layered Model," *IEEE Xplore*, Oct. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9264301 (accessed Mar. 31, 2021).

[4] H. Abroshan, J. Devos, G. Poels, and E. Laermans, "COVID-19 and Phishing: effects of human emotions, behaviour, and demographics on the success of phishing attempts during the pandemic," *IEEE Access*, pp. 1–1, 2021, doi: https://doi.org/10.1109/access.2021.3109091.

[5] G. J. W. Kathrine, P. M. Praise, A. A. Rose, and E. C. Kalaivani, "Variants of phishing attacks and their detection techniques," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2019, doi: https://doi.org/10.1109/icoei.2019.8862697

[6] Y. Wei and Y. Sekiya, "Sufficiency of Ensemble Machine Learning Methods for Phishing Websites Detection," *IEEE Access*, pp. 1–1, 2022, doi: https://doi.org/10.1109/access.2022.3224781.

[7] M. Sameen, K. Han, and S. O. Hwang, "PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System," *IEEE Access*, vol. 8, pp. 83425–83443, 2020, doi: https://doi.org/10.1109/access.2020.2991403.

[8] W. D. Yu, S. Nargundkar, and N. Tiruthani, "PhishCatch - A Phishing Detection Tool," *2009 33rd Annual IEEE International Computer Software and Applications Conference*, 2009, doi: https://doi.org/10.1109/compsac.2009.175.

[9] S. Gupta, A. Singhal, and A. Kapoor, "A literature survey on social engineering attacks: Phishing attack," *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Apr. 2016, doi: https://doi.org/10.1109/ccaa.2016.7813778

[10]M. Khonji, Y. Iraqi and A. Jones, "Phishing Detection: A Literature Survey," in IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091-2121, Fourth Quarter 2013, doi: 10.1109/SURV.2013.032213.00009.

[11]V. Shahrivari, M. M. Darabi, and M. Izadi, "Phishing Detection Using Machine Learning Techniques," *arXiv:2009.11116 [cs, stat]*, Sep. 2020, Available: https://arxiv.org/abs/2009.11116

[12] S. Lodha, A. Singh, and Harshal, "Everything Is in the Name – A URL Based Approach for Phishing Detection," May 2019.

[13] A. Butnaru, A. Mylonas, and N. Pitropakis, "Towards Lightweight URL-Based Phishing Detection," *Future*

Internet, vol. 13, no. 6, p. 154, Jun. 2021, doi: https://doi.org/10.3390/fi13060154.

[14] K. L. Chiew, C. L. Tan, K. Wong, K. S. C. Yong, and W. K. Tiong, "A new hybrid ensemble feature selection framework for machine learning-based phishing detection system," *Information Sciences*, vol. 484, pp. 153–166, May 2019, doi: https://doi.org/10.1016/j.ins.2019.01.064.

[15] K. Morovat and B. Panda, "A Survey of Artificial Intelligence in Cybersecurity," *IEEE Xplore*, Dec. 01, 2020. https://ieeexplore.ieee.org/document/9458190 (accessed Jan. 11, 2022).

[16] K. Morovat and B. Panda, "A survey of AI in Cybersecurity," 2020.

[17] M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: A Literature Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013, doi: https://doi.org/10.1109/surv.2013.032213.00009.