Phishing Detection Using Machine Learning

Gauri Vishal Desale ,B.E Student
Department of Computer Science
Met, Institute of Engineering
Email- gauridesale321@gamil.com

Tejashree Subhash Agrahari , B.E Student
Department of Computer Science
Met, Institute of Engineering
Email- tejashreeagrahari@gmail.com

Darshan Ramdas Desale , B.E Student
Department of Computer Science
Met, Institute of Engineering
Email- darshandesale.official@gmail.com

Revati Pitambar Jagtap , B.E Student
Department of Computer Science
Met, Institute of Engineering
Email- jagtaprevati2002@gmail.com

Dr. Vijay. B. More , Project Guide
Department of Computer Science
Met, Institute of Engineering
Email- vbmore2005@rediffmail.com

---***---

Abstract - The exponential increase in phishing assaults has led to significant financial losses for both individuals and companies, and it has also raised questions regarding the security and privacy of their personal information. The notable annual increase in phishing attempts suggests that the existing detection methods are inadequate, underscoring the need for the development of more robust and effective phishing detection techniques. This paper introduces an extension model for phishing detection employing machine learning, with the aim of enhancing both efficacy and accuracy in identifying phishing attempts. The survey conducted in this paper delves into the current state-of-the-art techniques in phishing detection, highlighting their limitations. Furthermore, it puts forth a novel approach in the developmental stages, centered around feature extraction from malicious URLs using machine learning.

***Key Words*:** *Machine Learning, Phishing Detection Extension,*

1.INTRODUCTION

Phishing constitutes a form of deception wherein perpetrators, masquerading as reliable entities in electronic communication, seek to acquire sensitive information—such as usernames, passwords, and credit card details—with malicious intent. Phishing attacks are a major concern for security researchers because attackers can easily create fake websites that look nearly identical to legitimate ones. While experts can typically identify fake websites, many users cannot and fall victim to these attacks. Attackers aim to steal banking credentials, resulting in significant financial losses for businesses. Phishing attacks are successful because of the lack of user awareness, and it is challenging to mitigate them.

To make phishing websites look legitimate, attackers often use social engineering tactics to deceive users. For example, they may create emails that appear to come from a trusted source and ask the user to Access a hyperlink directing to a counterfeit website . The email may contain urgent or important-sounding language to prompt the user to act quickly without thinking carefully.

Perpetrators may employ spear phishing, a focused variant of phishing that entails thorough research on the target. This method involves crafting a personalized message that mimics communication from a familiar or trusted source, thereby deceiving the victim. By using this technique, attackers can increase the chances of the victim falling for the scam To combat phishing attacks, it's important to educate users about how to recognize and avoid them. This includes being cautious about clicking on links or downloading attachments in unsolicited emails, verifying the legitimacy of a website before entering sensitive information, and using two-factor authentication and password protection to make it more difficult for attackers to gain access to accounts.

2. PROBLEM STATEMENT

Detecting phishing URLs is a complex and dynamic challenge, influenced by various factors and criteria that exhibit inherent instability. The objective is to scrutinize the data associated with a URL and its respective websites or web pages. This involves extracting effective feature representations from URLs and constructing a prediction model using training data encompassing both malicious and benign URLs. Employing machine learning techniques aims to enhance the overall versatility of malicious URL detectors. The ultimate goal is to dynamically and accurately determine whether a new website is potentially phishing or legitimate.

3. PROPOSED SYSTEM

Phishing attacks can take different forms, including email phishing scams and spear phishing. Users should be cautious and not fully trust common security applications as they can be vulnerable to such attacks. Machine learning can be used as an effective technique to detect phishing and overcome the limitations of existing approaches. Machine learning is a branch of artificial intelligence that has the ability to learn without explicit programming. Some of the commonly used machine learning techniques include supervised learning, unsupervised learning, and reinforcement learning. ML can be used in the development of information security applications, providing optimization, classification, prediction, and decision support systems, which can benefit those responsible for information security. The project aims to explore the use-case of detecting phishing websites using machine learning.

The system is a comprehensive web extension designed to enhance online security through the effective detection and prevention of phishing threats. It combines sophisticated technology and user education to create a robust defense against cyberattacks.

1. Phishing Detection Engine: The core of the system, utilizing machine learning algorithms and heuristic analysis for real-time website assessment.

2. User Interface: A user-friendly interface with a browser toolbar icon, offering clear visual indications of a website's safety status.

3. Real-time Analysis: Continuous monitoring of websites visited by users, assessing website content, URLs, and other factors for potential phishing indicators.

4. Educational Resources: Providing users with guides, tips, and resources to enhance their awareness of phishing threats and the ability to recognize and avoid them.

5. Cross-Browser Compatibility: Designed to work seamlessly across popular web browsers, including Google Chrome, Mozilla Firefox, Microsoft Edge, and more.

How the System Works:

The system functions as a guardian of online security. When a user visits a website, its phishing detection engine instantly initiates a comprehensive analysis, examining website content, URLs and more for potential signs of phishing. In parallel, it continually monitors the user's browsing session, alerting them in real-time if any potential phishing threats are detected. Furthermore, the system prioritizes user education, offering valuable resources, guides, and tips to empower users to recognize and evade phishing attempts. To ensure it remains up-to-date and effective, the system regularly connects to external databases of known phishing websites, and its commitment to user privacy is unwavering, adhering to stringent data protection and encryption standards. Importantly, the system is accessible to a wide audience, as it is designed to work seamlessly across various web browsers, ensuring a broad user base can benefit from its phishing detection capabilities. Users play an active role in improving the system's performance by reporting any false positives or false negatives they encounter through the user feedback mechanism. In summary system is a robust and comprehensive solution, creating a safer online environment for users through advanced technology, user education, privacy protection, and engagement.

4. SYSTEM ARCHITECTURE

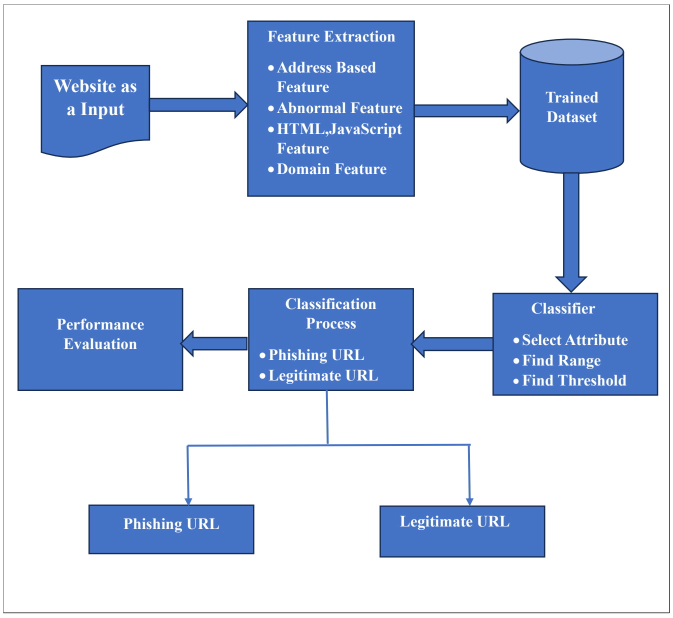

Fig -1: System Architecture

4. CONCLUSIONS AND FUTURE SCOPE

Websites are widely used in various fields for data entry and information processing applications. However, their popularity has also made them an attractive target for cyber threats such as phishing attacks. These attacks are typically designed to deceive users into providing their sensitive information or downloading malicious software by presenting a fake website that looks similar to the original one.

To combat phishing attacks, various detection methods and approaches have been developed and implemented. In this context, an application is proposed that uses Random Forest to classify and detect phishing URLs. The system aims to inform users about the presence of phishing URLs by identifying them as malicious even before the user visits the website.

The proposed system is based on a dataset obtained from the UCI website. The Random Forest Machine algorithm will be used to process the data and classify URLs as either phishing or benign. By using this approach, users can be alerted of potential phishing attacks and take appropriate measures to protect their sensitive information. The system can play a crucial role in mitigating the risks associated with phishing attacks and enhancing the security of web-based applications.

REFERENCES

1. Ahammad, S.K.H., Kale, S.D., Upadhye, G.D., Pande, S.D., Babu, E.V., Dhumane, A.V., Bahadur, M.D.K.J. (2022). "Phishing URL detection using machine learning methods."
2. https://www.researchgate.net/publication/362742019_Phishing_Detection_Using_Machine_Learning_Algorithm
3. https://www.hindawi.com/journals/jarn/2014/425731/(randomforest)
4. Ankit Kumar Jain and B.B. Gupta EURASIP Journal on Information Security (2016) 2016:9
5. Suleiman Y. Yerima, Mohammed K. Alzaylaee, High Accuracy Phishing Detection Based on Convolutional Neural Networks, IEEE Xplore