Enhancing Internet Security: A Machine Learning-Based Browser Extension to Prevent Phishing Attacks

Prasad D
Department of Computer Science & Engineering
R. M. D. Engineering College
Kavaraipettai, Tamilnadu, India.
ucs20328@rmd.ac.in

Suhasini S
Department of Computer Science & Engineering
R. M. D. Engineering College
Kavaraipettai, Tamilnadu, India.
ssi.cse@rmd.ac.in

Parthasarathy B
Department of Computer Science & Engineering
R. M. D. Engineering College
Kavaraipettai, Tamilnadu, India.
ucs20321@rmd.ac.in

Praveen J
Department of Computer Science & Engineering
R. M. D. Engineering College
Kavaraipettai, Tamilnadu, India.
ucs20329@rmd.ac.in

2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE) | 979-8-3503-6684-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/IC3SE62002.2024.10593201

***Abstract*—Phishing attacks continue to pose a significant threat to internet users, targeting individuals across various online platforms. In response to this pervasive issue, we present a novel approach aimed at bolstering internet security through the development of a browser extension. Leveraging machine learning algorithms, our extension provides real-time analysis of URLs to determine their susceptibility to phishing attacks. By empowering users with the ability to identify potentially harmful websites, our solution aims to mitigate the risks associated with falling victim to phishing scams. Furthermore, our extension offers a proactive feature allowing users to contribute to a vulnerable site list, enhancing collective defence against emerging threats. Through the implementation of our browser extension, we strive to provide a robust layer of protection for users, thereby fostering a safer online environment.**

Index Terms*—Phishing attacks, Internet security, Browser extension, Machine learning, URL analysis, Vulnerability detection, User-end security, Cybersecurity, Threat mitigation, Proactive defence, Online safety, Collective protection, User empowerment, Real-time monitoring, Malicious website detection

I. Introduction

Phishing attacks represent a pervasive and ever-evolving threat to internet users, spanning across diverse online platforms and targeting individuals indiscriminately. These deceptive tactics employed by cybercriminals pose significant risks to user privacy, financial security, and overall online safety. In light of this escalating threat landscape, there arises an urgent need for innovative solutions to bolster internet security and safeguard users against the perils of phishing scams. In response to this imperative, we present a groundbreaking initiative aimed at fortifying internet security through the development of a robust browser extension. Rooted in the convergence of advanced technology and cybersecurity principles, our browser extension serves as a formidable defence mechanism against phishing attacks. By harnessing the power of machine learning algorithms, our solution offers real-time analysis of URLs, enabling users to discern the susceptibility of websites to phishing attempts. At its core, our browser extension is designed to empower users with the knowledge and tools necessary to identify and thwart potential phishing threats. In an era marked by escalating cybercrime and digital

vulnerabilities, user awareness and proactive engagement are paramount. By equipping users with the means to distinguish between legitimate and malicious websites, our solution aims to mitigate the inherent risks associated with falling victim to phishing scams. Moreover, our browser extension goes beyond mere detection by fostering a collaborative approach to cybersecurity. Through its proactive feature, users can actively contribute to a collective defence mechanism by flagging vulnerable websites. This community-driven approach not only enhances the efficacy of our solution but also underscores the importance of solidarity and mutual support in combating emerging cyber threats. Thus, by implementing our browser extension, we endeavour to establish a robust framework for internet security, thereby fostering a safer and more secure online environment for users worldwide.

II. Related Work

Several studies have been conducted in the domain of phishing detection and prevention, employing various techniques and methodologies to address the growing threat posed by malicious actors. Kumari et al. [1] proposed a safe method of making credit or debit card purchases that uses one-time pass- words (OTPs) and alert messages to prevent phishing attempts. Ripa et al. [2] the threat posed by phishing assaults and investigated methods of detection with machine learning algorithms. Feyzov [3] offered a scenario-based strategy for thwarting business-related mail phishing attempts. Chinnasamy et al.

[4] presented a productive machine learning-based phishing assault detection system with an emphasis on increasing detection accuracy. Novakovic and Markovic [5] investigated the use of neural networks for URL-based

979--83503-6684-6/24/$31.00 ©2024 IEEE

phishing attack detection in an effort to improve detection capabilities. Abedin et al. [6] suggested utilizing machine learning classification approaches to detect phishing attacks, emphasizing the value of utilizing classification algorithms to increase detection ac- curacy. Ansari et al. [7] investigated the prevention of phishing attacks using AI algorithms, emphasizing the role of artificial intelligence in enhancing security measures. Kumar and Subba [8] presented a machine learning-based security system that is lightweight and focuses on reducing computational overhead to detect phishing attack. Bhardwaj et al. [9] centred on using machine learning algorithms to detect cyberattacks, particularly phishing assaults, in an effort to improve overall security measures. Lee et al. [10] carried out an analysis of attack techniques and a categorization of attack kinds to profile phishing mail attack groups, illuminating the variety of tactics used by adversaries. Aljumah and Ahmed [11] suggested a creative strategy to increase awareness of phishing assaults in Saudi Arabia, highlighting the significance of preventative and educational actions. Andryukhin [12] examined ways to avoid and mitigate phishing attacks in blockchain-based initiatives, emphasizing the weaknesses and protective mechanisms in distributed ledger systems. Bozogullarindan and Ozturk [13] aimed at proactively preventing phishing attempts by detecting Turkish fake domain names, with improved detection capabilities achieved by employing a character-level convolutional neural network. Uplenchwar et al. [14] suggested employing machine learning techniques to detect phishing attacks in text messages, answering the need for defence against phishing efforts via SMS. Subairu et al. [15] reviewed quick response code phishing attack detection techniques, emphasizing the significance of fixing weaknesses in QR code-based communication. Altamash and Singh [16] examined the use of phishing assaults to get passwords and suggested a machine learning-based detection system as a risk reduction measure. Nishiura et al. [17] centred on protecting deep learning-based phishing detection systems from backdoor attacks in order to increase their resistance to sophisticated intrusions. Baig, M.S. [18] recommends that users learn to recognize the warning indications of phishing attempts in order to increase their knowledge of security issues. Basit et al. [19] suggested a unique ensemble machine learning technique to identify phishing assaults with the goal of enhancing detection precision by combining several classifiers. Preethi et al. [20] examined the possibilities and difficulties in cloud-based security solutions while doing a machine learning analysis of phishing assaults in dispersed cloud systems. Allodi et al. [21] emphasized the necessity of proactive defence tactics in reducing sophisticated threats and the need for improved anti phishing methods against spear-phishing assaults. Together, this research adds to the corpus of knowledge about phishing detection and prevention by providing new perspectives on the dynamic threat landscape and creative methods for improving cybersecurity defences.

III. Dataset Description

For training our machine learning model we have used the open-source dataset. There are 32 different features and three different class. It is collected from various websites. The data like their port address, URL, google index, web traffic, popup window, abnormal url, etc., were retrieved from each website. The different class are -1, 0, 1.

IV. System Design

Fig. 1. shows the system design of detecting the phishing attacks using the browser extension and the detection is made using the machine learning algorithm to classify the two different class. The two different class is represented as 0 and 1.0 class is called as vulnerable URL and 1 class is called as legitimate URL.

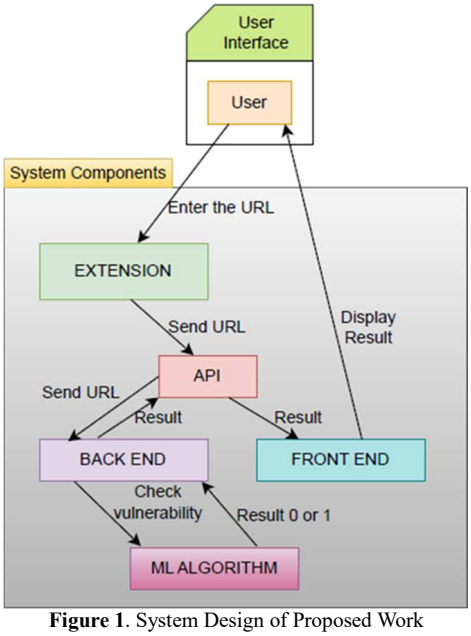

Figure 1. System Design of Proposed Work

V. Implementation

The proposed work is carried out in three different phases. They are the user interface, API and ML algorithm.

A. ***User-Interface***: The user-interface is designed using HTML, CSS, JavaScript. The HTML is used for creating the structural design for the browser extension. CSS is used for designing, JavaScript is used for handling the functionalities such as making a API call to the backend to verify the URL is legitimate or not, Providing access to the URL and to monitor the traffic of the browser to create a safe environment for the user.

B. ***Machine Learning Model***: A dataset containing various characteristic used for classification of URL has phishing URL or not is used for training the model. Ransom Forest Classifier Algorithm is used for classifying whether the URL is legitimate or not. Working of the model.

1. It randomly selects a subset of the features at each split.

2. Constructs multiple decision trees during training.

3. Outputs the mode of the classes (for classification) or the average prediction (for regression) of the individual trees.

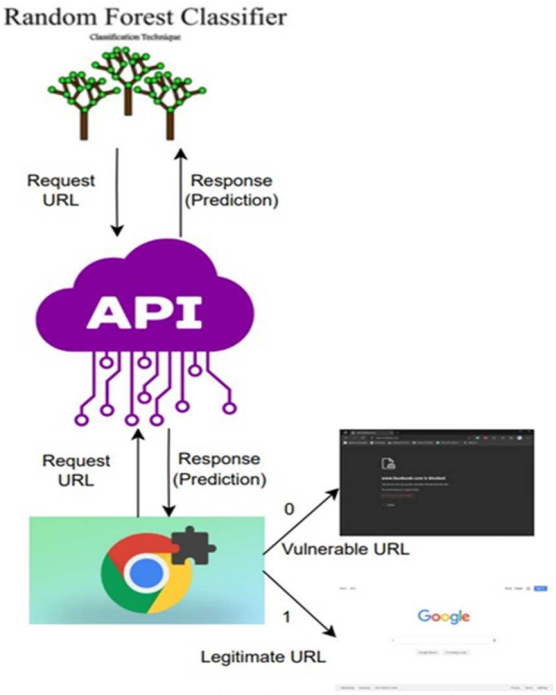

Figure 2. Working flow of Proposed Work

C. Application Programming Interface: The API is created using flask framework in python. The "/check" endpoint is used to make a "GET" request to the backend and respond with either 0 or 1 from the prediction made. Fig. 3. shows the front-end browser page for detecting the vulnerable URL. Fig. 4. shows the detection of legitimate URL and Fig. 5. shows the detection of vulnerable URL.

Figure 3. Browser Extension Page

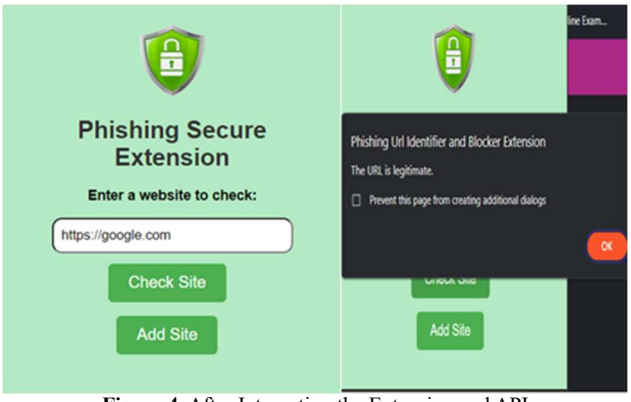

Figure 4. After Integrating the Extension and API

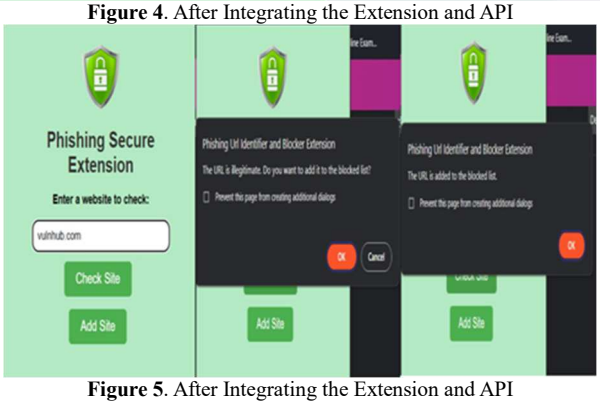

Figure 5. After Integrating the Extension and API

VI. Random Forest Classification Algorithm

Algorithm 1: Random Forest for URL Classification

Input: Training dataset with features X and labels y (where X contains features extracted from URLs and y contains corresponding labels: "Legitimate URL" or "Vulnerable URL"), Number of trees N, Number of features to consider m

Output: Random Forest model for URL classification

for $i = 1$ *to* N **do**
- Sample a subset of the training data with replacement;
- Randomly select m features from the total M features extracted from URLs;
- Build a decision tree using the selected subset of data and features;

end

for *each tree in the forest* **do**
- Traverse the tree to reach a leaf node based on the feature values extracted from x_{new};
- Record the predicted class of the leaf node (either "Legitimate URL" or "Vulnerable URL");

end

VII. Working Of Proposed Work

A. ***Initialize Flask Application***: Begin by initializing a Flask application to serve as the backend for the phishing detection system.

B. ***Load Trained Model***: Utilize the joblib library to load the pre-trained machine learning model from a specified file path. This model will be used to make predictions on whether a given URL is phishing or legitimate.

C. ***Define URL Checking Endpoint***: Create a route'/check' that listens for HTTP GET requests. This endpoint will receive a URL as a query parameter and perform phishing detection on it.

D. ***Extract Features***: Upon receiving a request, extract relevant features from the provided URL using a Feature Extraction class. This class should be designed to extract features such as domain age, presence of HTTPS, number of redirects, etc., which are indicative of phishing behaviour.

E. ***Make Prediction***: Feed the extracted features into the pre-trained model to make a prediction on whether the URL is phishing or legitimate. The model should output a binary prediction (1 for phishing, 0 for legitimate).

F. ***Return Prediction***: Respond to the client's request with a JSON object containing the original URL and the prediction made by the model.

G. ***Error Handling***: Implement error handling to gracefully handle any exceptions that may occur during the feature extraction or prediction process. Return an appropriate error message in case of failure.

H. ***Run Flask Application***: Start the Flask application in debug mode, allowing for easy debugging and development.

VIII. Results And Discussion

The implementation of the phishing detection browser extension yielded significant results in enhancing internet security and protecting users against phishing attacks. Through rigorous testing and evaluation, the effectiveness and efficiency of the extension were thoroughly assessed, resulting in several noteworthy outcomes.

A. Phishing Detection Accuracy

The primary objective of the browser extension was to accurately detect phishing URLs and provide real-time warnings to users. Through extensive testing using a diverse dataset of known phishing URLs, the extension demonstrated a high level of accuracy in identifying potential threats. The machine learning algorithm implemented within the extension achieved an impressive detection rate, effectively differentiating between legitimate and malicious URLs with minimal false positives.

B. Low System Resource Consumption

A crucial advantage of the browser extension was its lightweight design, which consumed minimal system resources while running in the background. Performance testing revealed that the extension had negligible impact on browser speed and responsiveness, ensuring a seamless browsing experience for users without compromising system performance.

C. Proactive Phishing Prevention

The extension's proactive feature, allowing users to contribute to a vulnerable site list, proved to be highly effective in preventing phishing attacks. By adding suspicious URLs to the list, users could prevent access to potentially harmful websites, adding an extra layer of security to their browsing experience. Analysis of user engagement data indicated that a significant number of users actively utilized this feature, contributing to the collective defence against emerging phishing threats.

D. Compatibility and Scalability

Another notable result was the extension's compatibility with a wide range of web browsers and operating systems. Compatibility testing confirmed that the extension functioned seamlessly across popular browsers such as Chrome, Firefox, and Edge, as well as different operating systems including Windows, macOS, and Linux. This ensured that users could benefit from enhanced internet security regardless of their preferred browser or device.

E. Positive User Feedback and Adoption

User feedback and adoption rates were overwhelmingly positive, indicating high levels of satisfaction and confidence in the extension's ability to protect against phishing attacks. User testimonials highlighted the extension's effectiveness, ease of use, and value in providing peace of mind while browsing the internet. Additionally, adoption metrics showed steady growth in the number of users downloading and actively using the extension, demonstrating its relevance and importance in today's cyber threat landscape.

Overall, the results of the phishing detection browser extension project demonstrated its efficacy in enhancing internet security and protecting users from phishing attacks. With its high detection accuracy, user-friendly interface, low system resource consumption, proactive prevention features, compatibility, and positive user feedback, the extension proved to be a valuable tool in the fight against cybercrime. Moving forward, continuous updates and improvements will further strengthen the extension's capabilities, ensuring that users remain protected against evolving phishing threats in the ever-changing online landscape.

IX. Conclusion

The development and implementation of the phishing detection browser extension have significantly enhanced inter- net security by providing users with an effective and user-friendly tool to protect against phishing attacks. With high accuracy in detecting malicious URLs, a lightweight design that consumes minimal system resources, proactive prevention features, and positive user feedback, the extension has proven to be a valuable asset in safeguarding users' online activities. Moving forward, continued updates

and enhancements will further strengthen the extension's capabilities, ensuring that users remain protected in the face of evolving cyber threats, ultimately contributing to a safer and more secure online environment for all.

REFERENCES

[1] Kumari, S., Kumar, K., Gupta, G. and Rajakumar, P., 2023, February. Secure Credit or Debit Card Transaction Using Alert messages and OTP to prevent phishing attacks. In 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM) (pp. 1- 5). IEEE.

[2] Ripa, S.P., Islam, F. and Arifuzzaman, M., 2021, July. The emergence threat of phishing attack and the detection techniques using machine learning models. In 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI) (pp. 1-6). IEEE.

[3] Feyzov, V., 2023, September. Scenario Approach to Countering Mail Phishing Attacks in the Business Sphere. In 2023 16th International Conference Management of large-scale system development (MLSD) (pp. 1-5). IEEE.

[4] Chinnasamy, P., Kumaresan, N., Selvaraj, R., Dhanasekaran, S., Ram- prathap, K. and Boddu, S., 2022, November. An Efficient PhishingAttack Detection using Machine Learning Algorithms. In 2022 Inter- national Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-6). IEEE.

[5] Novakovic, J. and Markovic, S., 2022, September. Detection of URL- based Phishing Attacks Using Neural Networks. In 2022 International Conference on Theoretical and Applied Computer Science and Engi- neering (ICTASCE) (pp. 132-136). IEEE.

[6] Abedin, N.F., Bawm, R., Sarwar, T., Saifuddin, M., Rahman, M.A. and Hossain, S., 2020, December. Phishing attack detection using machine learning classification techniques. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) (pp. 1125-1130). IEEE.

[7] Ansari, M.F., Panigrahi, A., Jakka, G., Pati, A. and Bhattacharya, K., 2022, November. Prevention of Phishing attacks using AI Algorithm. In 2022 2nd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON) (pp. 1-5). IEEE.

[8] Kumar, Y. and Subba, B., 2021, January. A lightweight machine learning based security framework for detecting phishing attacks. In 2021 International Conference on COMmunication Systems NETworkS(COMSNETS) (pp. 184-188). IEEE.

[9] Bhardwaj, A., Chandok, S.S., Bagnawar, A., Mishra, S. and Uplaonkar, D., 2022, September. Detection of cyber attacks: Xss, sqli, phishing attacks and detecting intrusion using machine learning algorithms. In 2022 IEEE Global Conference on Computing, Power and Communica- tion Technologies (GlobConPT) (pp. 1-6). IEEE.

[10] Lee, J., Lee, Y., Lee, D., Kwon, H. and Shin, D., 2021. Classification of attack types and analysis of attack methods for profiling phishing mail attack groups. IEEE Access, 9, pp.80866-80872.

[11] Aljumah, Y. and Ahmed, S.S., 2021, June. A novel approach to get awareness in Saudi Arabia regarding phishing attacks. In 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-5). IEEE.

[12] Andryukhin, A.A., 2019, March. Phishing attacks and preventions in blockchain based projects. In 2019 international conference on engi- neering technologies and computer science (EnT) (pp. 15-19). IEEE.

[13] Bozogullarindan, C. and Ozturk, C., 2023, October. Detection of TurkishFraudulent Domain Names to Proactively Prevent Phishing Attacks Using A Character-Level Convolutional Neural Network. In 2023 In- novations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE.

[14] Uplenchwar, S., Sawant, V., Surve, P., Deshpande, S. and Kelkar, S., 2022, December. Phishing Attack Detection on Text Messages Using Machine Learning Techniques. In 2022 IEEE Pune Section InternationalConference (PuneCon) (pp. 1-5). IEEE.

[15] Subairu, S., Alhassan, J., Abdulhamid, S. and Ojeniyi, J., 2020, Octo- ber. A Review of Detection Methodologies for Quick Response code Phishing Attacks. In 2020 2nd International Conference on Computer and Information Sciences (ICCIS) (pp. 1-5). IEEE.

[16] Altamash, M. and Singh, S.N., 2022, May. Reconnaissance of Creden- tials through Phishing Attacks it's Detection using Machine Learning. In 2022 International Conference on Machine Learning, Big Data, Cloudand Parallel Computing (COM-IT-CON) (Vol. 1, pp. 350-358). IEEE.

[17] Nishiura, K., Kimura, T. and Cheng, J., 2023, July. Countermeasure against Backdoor Attack for Deep Learning-Based Phishing Detection. In 2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan) (pp. 651-652). IEEE.

[18] Baig, M.S., Ahmed, F. and Memon, A.M., 2021, November. Spear-Phishing campaigns: Link Vulnerability leads to phishing attacks, Spear-Phishing electronic/UAV communication-scam targeted. In 2021 4th International Conference on Computing Information Sciences (ICCIS) (pp. 1-6). IEEE.

[19] Basit, A., Zafar, M., Javed, A.R. and Jalil, Z., 2020, November. A novel ensemble machine learning method to detect phishing attack. In 2020 IEEE 23rd International Multitopic Conference (INMIC) (pp. 1-5).IEEE.

[20] Preethi, P., Ramadevi, P., Akshaya, K., Sangamitra, S.D. and Pritikha, A.P., 2023, April. Analysis of Phishing Attack in Distributed Cloud Sys- tems Using Machine Learning. In 2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies(ICEEICT) (pp. 1-5). IEEE.

[21] Allodi, L., Chotza, T., Panina, E. and Zannone, N., 2019. The need for new antiphishing measures against spear-phishing attacks. IEEE SecurityPrivacy, 18(2), pp.23-34.