IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

PHISHGUARD PRO: PHISHING DETECTION AND PREVENTION

[1]Rushikesh Ghonmode, [2]Aditya Shimpi, [3]Priyam Shrivastav, [4]Dhanashree Wadhnere

Dept. Of Computer Science & Engineering, Sandip University Nashik, India

ABSTRACT

This study proposes an intelligent model for detecting phishing websites using Extreme Learning Machine (ELM). Phishing websites attempt to extract confidential data by mimicking legitimate sites. Our approach involves preprocessing a dataset of phishing and legitimate URLs, extracting features such as domain, address, abnormal attributes, HTML, and JavaScript features. Machine learning techniques, specifically Random Forest and Support Vector Machine (SVM), are employed for classification based on URL attributes. The system computes range and threshold values for classification, aiming to detect phishing instances effectively. By leveraging feature extraction and realtime detection, our model contributes to mitigating the risks associated with phishing attacks, enhancing user and organizational security.

I.INTRODUCTION

Phishing attacks pose a significant cybersecurity threat to individuals and organizations globally, exploiting unsuspecting users to divulge sensitive information. Conventional defenses like rulebased systems struggle to keep pace with the evolving sophistication of these attacks. Hence, there's a pressing need for advanced techniques. Machine learning (ML) offers promise in detecting phishing sites by learning patterns and anomalies associated with them. This project aims to develop an accurate and efficient MLbased system for phishing site detection. By training on labeled datasets, the model can distinguish between legitimate and phishing websites, improving detection performance and reducing false positives. The system empowers users to recognize potential threats, safeguarding sensitive data and organizational assets against phishing's damaging consequences. ThroughPhishguard Pro, we aim to contribute to a safer online environment by leveraging ML for intelligent defense against phishing.

II. METHODOLOGY

Problem definition : Phishing site detection is truly an unpredictable and element issue including numerous components and criteria that are not stable. To analyze the information of a URL and its corresponding websites or webpages, by extracting good feature representations of URLs, and training a prediction model on training data of both malicious and benign URLs. To improve the generality of malicious URL detectors, machine learning techniques. To make an accurate decision dynamically as to whether the new website is phishing or legitimate.

Use Case Diagram: A visual representation showcasing the interaction between users and the system, highlighting various user roles and associated use cases.

Class Diagram: Illustrates the structure of the application through main classes like User, URL, System, Alert, and Analyzer.

Activity Diagram: Depicts the system's behavior, demonstrating the flow of control from start to finish, including decision paths during execution.

Sequence Diagram: Illustrates how objects collaborate and interact over time, aiding in understanding system requirements and processes.

System Architecture: Outlined steps include Data Collection, Preprocessing, Feature Extraction, Training Data Preparation, Machine Learning Model Training, Model Evaluation, Real-time Phishing Detection, Alerting and Countermeasures, System Monitoring and Updates, and Reporting and Analysis.

Data Flow Diagrams: A visual representation mapping the flow of information within the system, ranging from simple overviews to detailed representations, facilitating analysis and comprehension for technical and non-technical audiences.

III. OVERVIEW OF PROJECT

Working Module : This module involves importing phishing datasets and legitimate URLs, preprocessing the data, and extracting URL features. Phishing detection is performed based on URL characteristics like domain-based, addressbased, abnormal-based, and HTML/JavaScript features. A browser extension is developed to analyze websites in real-time, classify them, and provide warnings to users.

Architectural Module : Defines the system's architecture and structure, facilitating the integration of various components and modules.

Feature Extraction Module : Transforms raw data into numerical features, enhancing machine learning performance. It automatically extracts features like URL structure, content analysis, SSL certificate status, domain age, IP reputation, HTML attributes, redirect chains, website popularity, and behavioral features.

Classification Module : Utilizes multiple classifier algorithms for phishing detection, including Logistic Regression, Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors (KNN), and Naive Bayes. Each algorithm's time complexity is analyzed to ensure efficiency.

Algorithm Details :

Logistic Regression : O(m * n)

Support Vector Machines (SVM) : O(m^2 * n)

Random Forest : O(m * n * log(n))

K-Nearest Neighbors (K-NN) : O(m * n * k)

Naive Bayes : O(m * n)

These algorithms are selected based on their efficiency and effectiveness in phishing detection.

IV. CONCLUSION

In a constantly evolving cyber threat landscape, the development of effective phishing detection systems is crucial. This project's clientside approach not only enhances detection speed but also prioritizes user privacy, making it a significant advancement in cybersecurity. Looking ahead, further improvements through advanced machine learning techniques, real-time threat intelligence integration, user behavior analysis, and automated response mechanisms are envisioned. Collaboration with security communities and continuous evaluation will ensure the system'seffectiveness in combating phishing attacks, ultimately contributing to safer online interactions and bolstering cybersecurity measures.

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] Rohmat Rose, M.A.S., Basir, N., Heng, N.F.N.R., Zaizi, N.J.M., Saudi, M.M. (2022). "Phishing Detection and Prevention using Chrome Extension."

[3] Yerima, S.Y., Alzaylaee, M.K. (2021). "Phishing Site Detection Using Similarity of Website Structure."

[4] Megha, N., Remesh, K.R., Babu, E.S. (2020). "An Intelligent System for Phishing Attack Detection and Prevention."

[5] Alkawaz, M.H., Steven, S.J., Hajamydeen, A.I. (2020). "Detecting Phishing Website Using Machine Learning."

[6] Yerima, S.Y., Alzaylaee, M.K. (2020). "High Accuracy Phishing Detection Based on Convolutional Neural Network."

[7] Satapathy, S.K., Mishra, S., Mallick, P.K., Badiginchala, L., Gudur, R.R., Guttha, S.C. (2019). "Phishing Detection Using Machine Learning."

[8] Jain, A.K., Gupta, B.B. (2016). "A novel approach to protect against phishing attacks at client-side using auto-updated white-list."

[9] Amani Alswailem, BashayrAlabdullah, Norah Alrumayh, Dr.AramAlsedrani, Detecting Phishing Websites Using Machine Learning 978-1-7281-0108-8/19/ 2019 IEEE