WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Waf Phishblocker: Preditive Attention Mechanism for Real Time Detection and Blocking of Phishing and Malicious Code

Ms. M. Sasikala[1], Mr. K. Yogeshwaran[2], Mr. R. Vimal[3], Mr. D. Srinivasan[4], Mr. K. P. Navrang[5]

[1]*Assistant Professor Computer Science and Engineering & Dhirajlal Gandhi College of Technology*
[2]*Computer Science and Engineering & Dhirajlal Gandhi College of Technology*
[3]*Computer Science and Engineering & Dhirajlal Gandhi College of Technology*
[4]*Computer Science and Engineering & Dhirajlal Gandhi College of Technology*
[5]*Computer Science and Engineering & Dhirajlal Gandhi College of Technology*

ABSTRACT

Web Application Firewall Phishing is defined as a cyber-attack which uses social engineering via digital means to persuade victims to disclose their personal information, such as their password or credit card number. In the end, the stolen personal information is used to defraud the trust of regular websites or financial institutions to obtain illegal benefits. Although different solutions have been exercised against phishing, phishing attacks have dramatically increased in the past few years. Some solutions are based on the features extracted by rules, and some of the features need to rely on third-party services, which will cause instability and time- consuming issues in the prediction service. This project proposes WAF PhishBlocker a deep learning framework that uses Predictive Attention Model with Recurrent Neural Network(RNN) to detect phishing links in a real-time web browsing environment using URL and HTML features.

Keywords : Web Application Firewall, Phishing, Detection, Malicious URL, Recurrent Neural Network(RNN)

I. INTRODUCTION

Web Application Firewall Phishing is a type of cybersecurity attack during which malicious actors send messages pretending to be a trusted person or entity. Phishing messages manipulate a user, causing them to perform actions like installing a malicious file, clicking a malicious link, or divulging sensitive information such as access credentials. Phishing is the most common type of social engineering, which is a general term describing attempts to manipulate or trick computer users. Social engineering is an increasingly common threat vector used in almost all security incidents. Social engineering attacks, like phishing, are often combined with other threats, such as malware, code injection, and network attacks. Phishing is the most common form of social engineering, the practice of deceiving, pressuring or manipulating people into sending information or assets to the wrong people. Social engineering attacks rely on human error and pressure tactics for success. Protocol Validation Protocol Validation is a basic passive defense against potential attacks that take advantage of atypical HTTP capabilities. Its purpose is to leave as little room for attackers as possible by limiting the request to special checks. First of all, this is checking HTTP headers for RFC compliance, but this is not enough, and manufacturers go further. Using restrictions on "best practices" and their own rules formulated during the analysis of possible vulnerabilities. The following restrictions generally apply: RFC requirements; length and number of headers, parameters; time frame; validation of JSON, XML entities; no invalid values.

II. LITERATURE SURVEY

Vishal *et al*. collected a dataset of 10,000 URLs, including 5,000 legitimate and 5,000 phishing URLs. The proposed framework uses features extracted from URLs and applies various machine learning algorithms to classify the URLs as phishing or legitimate. The framework combines various machine learning models such as decision tree, random forest, and support vector machine to increase the accuracy of the detection. The proposed hybrid framework achieved high accuracy of 99.1% in detecting phishing websites. The framework is also scalable and can be used in real-time environments.

Ahmed *et at.* found that the random forest algorithm achieved the highest accuracy in detecting phishing websites, with an accuracy of 98.46%. The study provides useful insights into the effectiveness of different machine learning algorithms in detecting phishing websites. Shih-Hao Hung *et al.* proposed method achieved an accuracy of 99.03% and outperformed several existing state-of-the-art methods for phishing website detection. The proposed method uses a deep neural network model consisting of a combination of convolutional neural network (CNN) and long short-term memory (LSTM) layers. The model is trained on a dataset of phishing and legitimate URLs.

III. SYTEM IMPLEMENTATION

A. Existing System

There are several existing machine learning systems for phishing website detection and web attacks,

1. OpenPhish - a free, open-source machine learning system that analyzes URLs to identify potential phishing websites.

2. PhishTank - a community-driven website that uses machine learning algorithms to detect and block phishing websites.

3. Google Safe Browsing - a machine learning system that detects phishing and malware websites by analyzing URLs and web content.

These systems use a variety of machine learning techniques, including supervised and unsupervised learning, deep learning, and natural language processing, to identify and block phishing websites. They typically analyze website content, URLs, and other characteristics to determine whether a website is a phishing site or not.

There are several machine learning algorithms that have been used for phishing website detection. Here are a few examples:

Decision Trees: Decision trees are commonly used in phishing detection systems because they can handle both numerical and categorical data. They work by recursively splitting the data into subsets based on the most informative features until a stopping criterion is met.

Support Vector Machines (SVMs): SVMs are a type of binary classifier that separate data into different classes based on a hyperplane. They are useful for phishing detection because they can handle large feature sets and are effective at detecting outliers. These are just a few examples of the machine learning algorithms that are commonly used for phishing detection. The effectiveness of each algorithm depends on several factors, such as the quality and size of the dataset, the feature selection process, and the chosen performance metrics.

B. PROPOSED SYSTEM

WAF Phishing attacks are one of the most common cyber threats nowadays. To prevent these attacks, we propose a system called "PhishBlocker" that uses a predictive attention mechanism with a recurrent neural network to detect and block phishing websites in real-time. The proposed system consists of several stages, including dataset collection, preprocessing, feature extraction (URL and HTML), classification, model building and training, and performance evaluation. The dataset is collected from various sources and preprocessed to extract relevant features such as domain age, IP reputation, and SSL certificate information. The URL and HTML features are extracted using various techniques, such as domain analysis, HTML parsing, and content analysis. The classification stage involves training a recurrent neural network model using the extracted features to classify a website as either legitimate or phishing. The model is trained and evaluated using various performance metrics such as accuracy, precision, recall, and F1 score. The system is developed using Python Flask and MySQL. The admin is responsible for training the model and updating the database with new phishing information. Users can open their browser and input the URL into PhishBlocker, which predicts and blocks the URL if it is identified as a phishing website. The system also stores attack information in the user's account on the PhishBlocker website, enabling users to track their activity and take necessary measures.

B.1 COLLECTION OF DATASET

Data Collection: The first step in the system design involves the collection of data. The data can be collected from various sources like online databases, APIs, and web crawlers. The collected data includes both legitimate and phishing URLs.

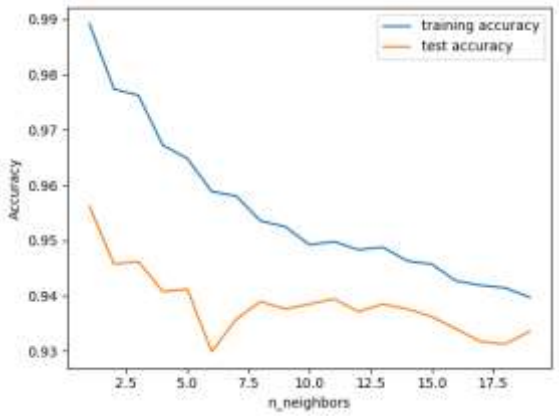
B.2 SELECTION OFATTRIBUTES

Attribute or feature selection is an important step in building a machine learning model as it involves identifying the most relevant and informative attributes that can be used to predict the target variable. In the case of a phishing detcetion, various attributes such as gender, URL, ip,parameters commonly used as input features

B.3 PRE-PROCESSING OF DATA

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre- processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model.

B.4 MODEL SELECTION



Accuracy

Select a suitable machine learning algorithm to train the model, such as logistic regression, decision tree, or random forest. Validate the model using the testing data and tune the hyperparameters to improve the model's accuracy.

B.5 PHISHING DETECTION

Various machine learning algorithms like SVM,Naive Bayes, Decision Tree, Random Tree, Logistic Regression, MLP are used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for Phishing prediction. Once the model is trained and validated, deploy it in a web application or mobile app that can be used by users predict the malicious url or code and block it.

IV. SYSTEM ARCHITECTURE

An architecture for heart disease prediction using machine learning was developed. The first step involved collecting data on risk factors for phishing detection, such as url, IP, Parameters. The data was preprocessed to remove missing values and normalize the features.

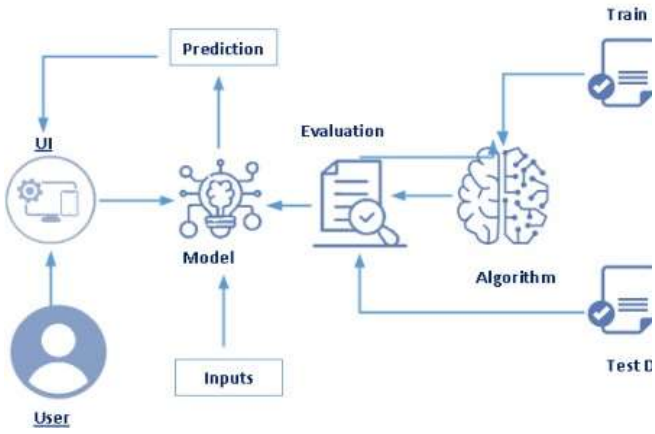


Fig: System architecture

V. CONCLUSION

"PhishBlocker" is a sophisticated web application that utilizes a predictive attention mechanism using recurrent neural networks to detect and block phishing websites in real-time. The system is designed with a comprehensive dataset collection, pre-processing, and feature extraction of URLs and HTML, followed by classification and model training. The performance evaluation of the model is measured with precision, recall, F1-score, and accuracy. The system also includes an alert or notification module, a track history module, and a user account to store attack information. Through the feasibility study and software testing, the system has demonstrated its ability to accurately detect and block phishing websites, making it a valuable tool for internet users to protect themselves from phishing attacks. The software testing also highlighted the compatibility of the system with various web browsers and operating systems. Overall, "PhishBlocker" provides a reliable and effective solution to protect against phishing attacks, which remain a

significant threat to internet users. However, further improvements can still be made to the system, including expanding the dataset, improving the feature extraction process, and integrating additional security measures. Overall, PhishBlocker is a useful tool in the fight against phishing attacks and can help users stay safe online.

VI. REFERENCES

[1]. Shostack, Adam. Threat modeling: designing for security. Wiley, 2014.

[2]. Jajodia, Sushil, et al. Handbook of database security: Applications and trends. Springer, 2007.

[3]. Clarke, Nathan J., et al. "Phishing attacks and countermeasures." ACM Computing Surveys (CSUR) 48.2 (2015): 1-33.

[4]. Kumar, Anish, et al. "Phishing detection using machine learning: a review." International Journal of Advanced Research in Computer Science and Software Engineering 8.2 (2018): 373-379.

[5]. Alazab, Mamoun, and Sitalakshmi Venkatraman. "Phishing websites detection based on machine learning techniques." International Journal of Computer Applications 179.24 (2020): 6-12.

[6]. Bacciu, Davide, et al. "A comprehensive review of computational intelligence techniques applied to phishing detection." Journal of Network and Computer Applications 100 (2017): 1-24.

[7]. Xu, Tian, and Zheng Yan. "Detecting Phishing Websites Using Machine Learning Techniques." Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques (2021): 404-419.

[8]. Chan, David WK, et al. "Effective phishing website detection using machine learning." Expert Systems with Applications 41.10 (2014): 4974 4985.

[9]. Ye, Jinyu, et al. "A multi-feature based machine learning approach for phishing website detection." IEEE Transactions on Information Forensics and Security 12.6 (2017): 1287-13