IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

PHISHWIPER: Real Time Scam Website Detection and Blocking using Predictive Attention Model

M. K. Siva Prakash[1] and A. Poongodi[2]
PG Student, Department of Computer Applications[1]
Professor, Department of Computer Applications[2]
Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India
mgmsivaprakash @vistas.ac.in and poongodimca1979@gmail.com

Abstract: *A data breach is a security event, where sensitive data is accessed without any permission from a website or an organization. An information breach will be considered as the purposeful or accidental gathering of secure or personal data from an organization. A breach can be an accession of a data without any permission, these kinds of regulations should be provided with safe and secured framework but this is not happening in many corporations. So, by analyzing the previous attempts (successful or unsuccessful attacks), the proposed model can be trained to adapt to new scenarios and predict the next breach. Further, this research work has designed a model by using machine learning to defend a website from security breaches. The primary aim of this research work is to create a machine learning model, which trains in Real-time and monitors the website or a system and trains from the state-of-art attacks. The proposed model has created a web application, which takes the data from multiple sources such as Amazon, Flipkart, Snapdeal, and Shop clues, which shows the data that is safe to obtain from the website.*

Keywords: Data breach, personal data, Real time, state-of-art attacks, framework

I. INTRODUCTION

Phishing is a type of cyber security 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. “Phishing” refers to an attempt to steal sensitive information, typically in the form of usernames, passwords, credit card numbers, bank account information or other important data in order to utilize or sell the stolen information. By masquerading as a reputable source with an enticing request, an attacker lures in the victim in order to trick them, similarly to how a fisherman uses bait to catch a fish. The most common examples of phishing
are used to support other malicious actions, such as on-path attack and cross-site scripting attacks. These attacks typically occur via email or instant message, and can be broken down into a few general categories. It’s useful to become familiar with a few of these different vectors of phishing attacks in order to spot them in the wild.

II. PROBLEM STATEMENT

Phishing URL and website detection can be challenging due to the constantly evolving tactics used by phishers to make their attacks more convincing and difficult to detect. Some of the problems that can be encountered include: Polymorphic URLs: Phishers can use a technique called polymorphic URLs, where they generate a unique URL for each target, making it harder to detect and block these URLs. False positives: URL and website detection tools can sometimes generate false positives, which means that legitimate URLs or websites are incorrectly flagged as phishing sites, leading to inconvenience and frustration for users. Zero- day attacks: Phishers can use previously unknown

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

vulnerabilities or exploits in popular websites or browsers to launch phishing attacks, making it harder to detect and prevent such attacks. Lack of user awareness: Despite the availability of advanced detection tools, many users are still not aware of the risks associated with phishing attacks and can fall prey to such scams. Phishing URL and HTML website detection using machine learning can face several challenges, including. The availability of high-quality training data is critical to the success of any machine learning model. However, for phishing website detection, there may be limited data available, especially for new or emerging types of phishing attacks. Identifying the relevant features to use in the machine learning model can be challenging. Some features may be more indicative of phishing websites than others, and selecting the wrong features can lead to poor performance. The number of legitimate websites is significantly higher than the number of phishing websites. As a result, the data may be imbalanced, which can lead to biased models that perform poorly on the minority class (phishing websites).

III. METHODOLOGY SECTION

Phish Wiper Web App

The design and development of the Phish Wiper website with Python Flask and MySQL modules:

- **Flask Framework:** Flask is a lightweight and flexible web framework written in Python. It provides a lot of features for building web applications, including routing, templates, and sessions. Flask is used in the Phish Wiper website to create web pages and handle HTTP requests and responses.
- **MySQL Database:** MySQL is a widely used open-source relational database management system. It is used in the Phish Wiper website to store user data, attack information, and other relevant data.
- **HTML/CSS/JavaScript:** HTML is used to create the structure of web pages, CSS is used for styling the web pages, and JavaScript is used for adding interactivity and functionality to the web pages. Recurrent Neural Network: The Phish Wiper website uses a recurrent neural network to predict and block phishing URLs. The RNN is trained on a dataset of phishing URLs and uses a predictive attention mechanism to make accurate predictions.
- **User Authentication:** User authentication is an important feature of the Phish Wiper website. It allows users to register, log in, and configure their systems to prevent phishing attacks.
- **Attack Information Storage:** The Phish Wiper website stores attack information in the user account, allowing users to view their attack history and take appropriate actions to prevent future attacks.
- **Model Training**: The Phish Wiper website allows the admin to train the model with new datasets to improve the accuracy of predictions.

End User Interface

The Phish Wiper end user interface consists of two modules, one for the admin and another for the user.

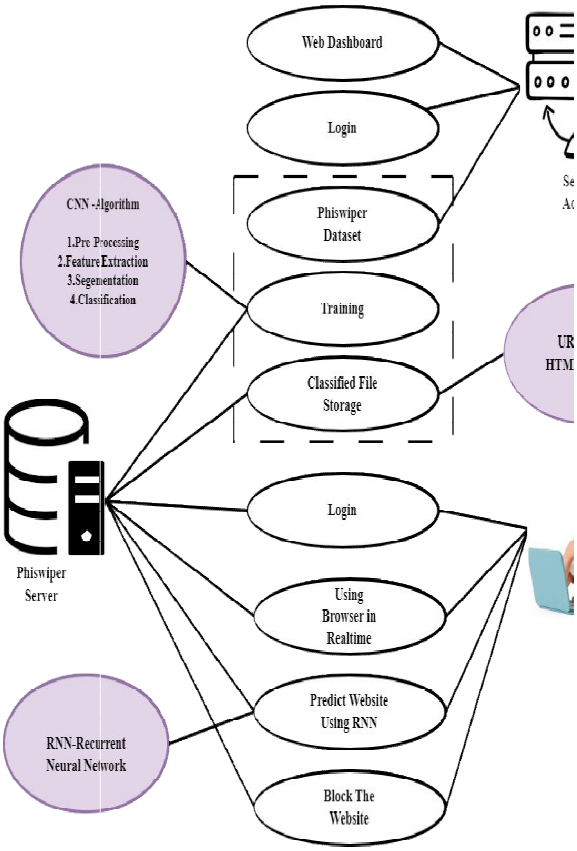
Admin Interface Module

The admin interface module allows the admin to login to the Phish Wiper website with their credentials. Once logged in, the admin can train the model with new phishing URLs and HTML pages, which will be used for real-time detection and blocking of phishing websites. The admin can view and manage the trained models, as well as view the attack history and analytics.

User Interface Module

The user interface module is designed for end-users to configure their system to prevent phishing attacks. The user needs to register on the Phish Wiper website to get login credentials. Once logged in, the user can configure their system by providing necessary details such as the browser they use, the operating system, and other security-related settings. The user can also view the history of detected phishing attempts and their status.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

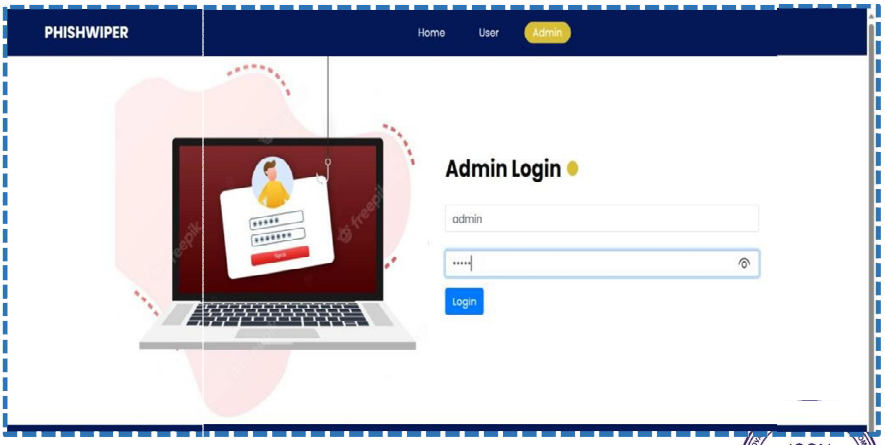


Use case Diagram

IV. EXPERIMENTAL RESULT

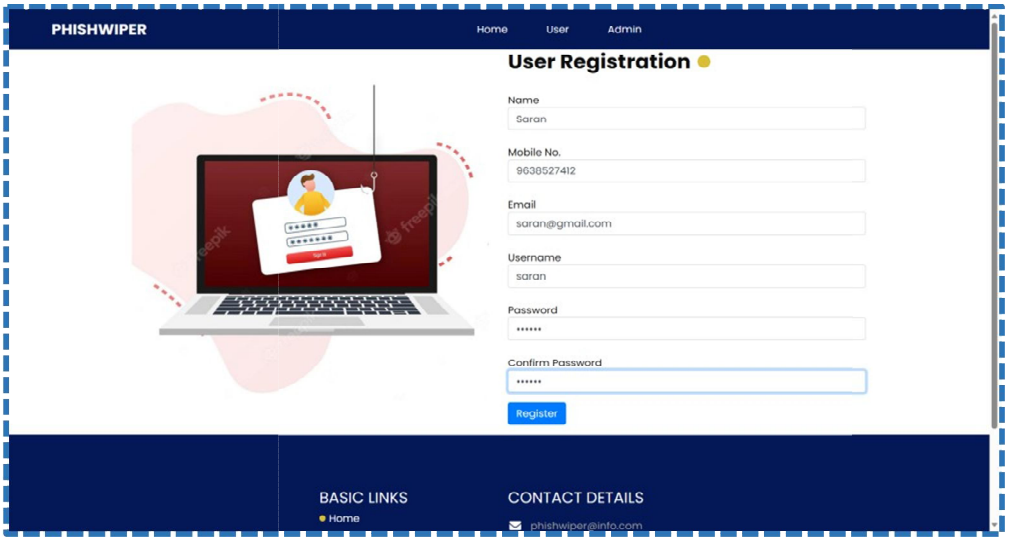
Phish Wiper is a web application developed using Python Flask and MySQL. The application aims to provide a real-time detection and blocking solution for phishing websites and URLs. The project uses an RNN with a predictive attention mechanism for accurately classifying and training the phishing websites and URLs. The interface allows users to register, login, and configure their system to use the Phish Wiper service. When a user opens a web page in their browser, Phish Wiper predicts whether the page is a phishing website or not using the trained RNN model. If the page is identified as a phishing website, Phish Wiper blocks the page and notifies the user about the attack. Phish Wiper also stores information about the attack in the user's account, allowing the user to view their attack history and take appropriate actions to prevent future attacks.

Admin Login:

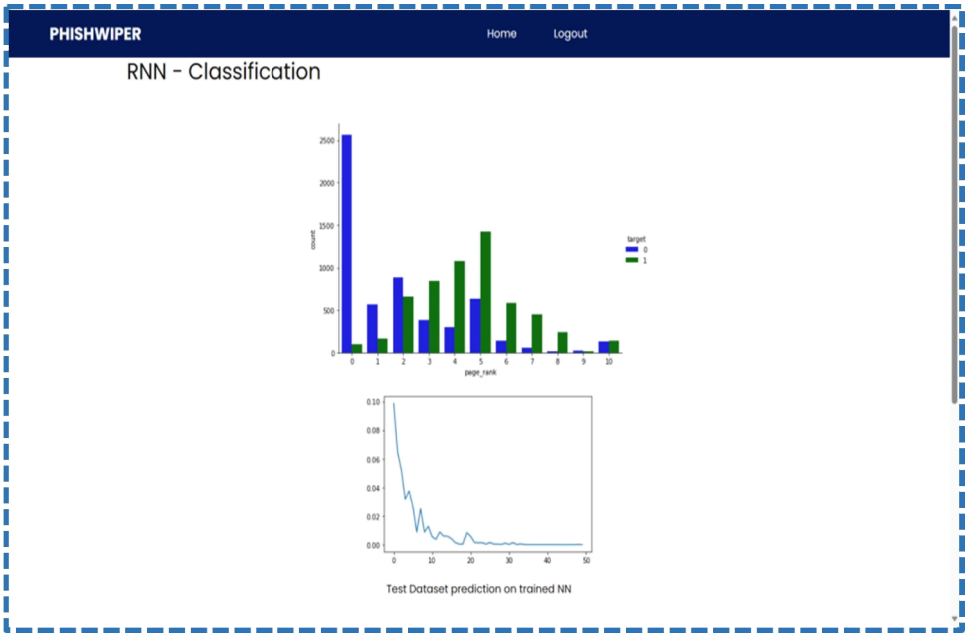


IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

User registration:



RNN Classification:



V. CONCLUSION

In conclusion, this project 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, the proposed system of the project provides a reliable and effective solution to protect against phishing attacks, which remain a significant threat to internet users.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

REFERENCES

[1]. Alazab, M., Watters, P., & Alazab, M. (2017). Machine learning-based phishing detection: An empirical study. Journal of Information Security and Applications, 32, 102-115.

[2]. Alhazmi, O., & Almuhaideb, S. (2019). A comparative study of machine learning algorithms for detecting phishing websites. International Journal of Advanced Computer Science and Applications, 10(8), 283-289.

[3]. Gandotra, E., & Singh, S. (2018). An analysis of machine learning algorithms for phishing detection. In 2018 3rd International Conference on Computing and Communications Technologies (ICCCT) (pp. 46-50). IEEE.

[4]. Hidayanto, A. N., Bayuaji, R., & Kurniawan, F. (2019). Phishing websites detection using machine learning and entropy feature selection. Journal of Physics: Conference Series, 1317(1), 012016.

[5]. Hu, Y., & Chau, M. (2018). Machine learning-based detection of phishing websites using content and link features. Decision Support Systems, 114, 26-38.

[6]. Jaiswal, A. K., Mishra, S. K., & Tyagi, S. (2020). A study of machine learning-based approaches for phishing website detection. International Journal of Computer Science and Information Security, 18(9), 51-57.

[7]. Kalita, J. K., & Sarma, M. (2018). Machine learning based phishing website detection. In 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU) (pp. 1-6). IEEE.

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

[9]. 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.

[10]. 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.

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