Phoenix: International Multidisciplinary Research Journal
Vol 1, No. 3, July-Sep, 2023
ISSN:2583-6897

AN EXPLAINABLE MULTI-MODAL HIERARCHICAL ATTENTION MODEL FOR DEVELOPING PHISHING THREAT INTELLIGENCE

First Author[1], Second Author[2]

[1]Nikita, Master of Computer Application BKIT-Bhalki
[2]Prof . Gayatri Mugli, Master of Computer Application BKIT-Bhalki

Abstract - Today, phishing is one of the most dangerous online risks since it allows malicious websites to steal users' login information. Sites that use phishing to steal users' personal information. sensitive information when they browse a phony website. Website that seems like the real thing is another Internet crime is on the rise, and it's one of the most in particular worries about several other sectors, including electronic account management and retail. Phishing is, in general, a large-scale fraud that occurs when a rogue website behave like a genuine server. The identification of phishing websites is a real and a complex and ambiguous matter with many factors and unreliable standards of evaluation. This article describes a method which can identify and stop both preexisting and freshly created threats URLs used in phishing attacks that have absolutely no history of any kind evaluate the use of Data Mining. An online sorting system model will be developed for the same, with many taken from parameters obtained from the URL's properties. The model will be taught to recognize patterns in a large dataset to maximize precision and precision. Random Forest was used for this purpose. (RF) is a subset of machine-learning-based Phishing website detection algorithms. Now, at long last, we Delete the website from our network.

***Key Words*:** Data Mining, Phishing, URL, RF

INTRODUCTION

Phishing is a method used by hackers and attackers to deceive people into divulging important information, such as login credentials and credit card information, to a fake website. In this kind of assault, the bad guys pretend to be the good guys by using a phony logo or website. In this way, the users are fooled by the phony website's design, which is almost similar to the real one. Attackers often target users of e-commerce platforms, social networking platforms, and online financial services. The frequency of dangerous mass E-mails that include links to phishing sites increased and spam flows varied widely in 2016 [6]. PhishTank has confirmed 2,259,845 sites as phishing sites [10] up till quite recently. As a result, phishing has superseded all other methods as the primary vector for spreading malware [6]. A highly efficient anti-phishing solution is thus in high demand.

Phishing is a kind of widespread fraud in which a fake website is used to trick users into divulging sensitive information such as passwords, account details, or credit card numbers. The goal of phishing is to get sensitive information, such as passwords and credit card details, by making the target believe they are communicating with a legitimate organization or person via the Internet.

Misleading websites used in phishing attacks are designed to seem like those of legitimate businesses, such as financial institutions and online marketplaces.[11].We propose a method that use Machine Learning to identify malicious URLs, such as those used for phishing, spamming, etc. Online fraud that might compromise sensitive user information could be avoided with the use of this solution.

A technique called "Detection and Prevention of Phishing Websites using Machine Learning Approach" is presented to achieve this high level of safety[7]. In this system, we deal with URLs and use a machine learning method to determine whether or not they lead to a phishing website. In this article, we will develop a web browser. When we visit a site, its URL will be validated using a machine learning method. If it is expected that the outcome will be bad, access to that website will be denied. If so, we won't be able to use that computer

to go online. The primary benefit of our approach is that it prevents access to the banned site across all browsers, not just the one we built. In this video, we create a basic browser application in Java and have it relay the user-entered URL to Python.

Problem statement:

In 2018, 80% of cyber assaults were reported to be directed towards emails, according to a study. When a person clicks on a phishing attack's URL, their computer is infected. Phishing emails allow an attacker to get a foothold from which to launch more sophisticated assaults. Phishing attacks are directed at the human target since they are the most vulnerable part of any security system. The credit card industry and others like it play a significant role in an assault. The assaults have a significant financial impact.

Modern email security is more vulnerable to phishing assaults. The assaults are constantly evolving to circumvent safeguards that have been put in place. However, research on how to spot suspicious emails is expanding rapidly. The creation of a smart method for identifying and stopping phishing attempts is urgently required. To improve email security and make it less reliant on user knowledge, we need effective solutions that use machine learning methods.

SYSTEM ANALYSIS:

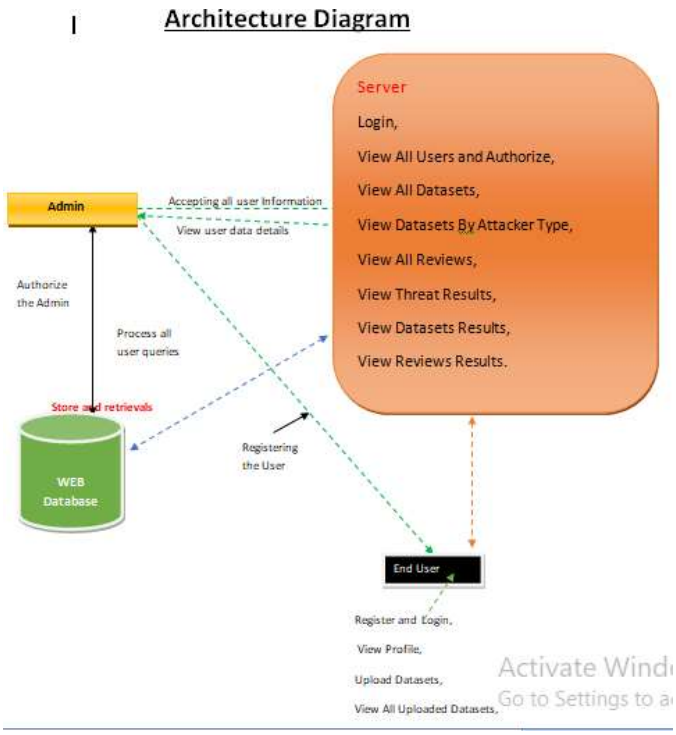
Existing System:

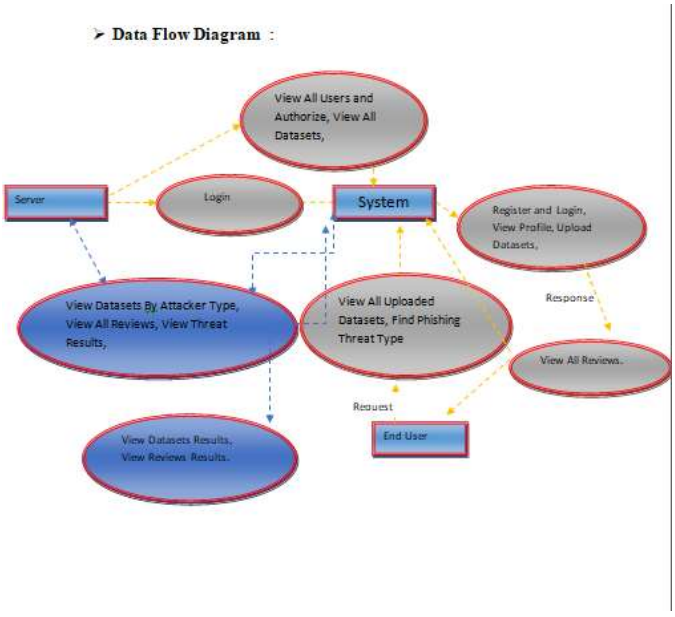
Ankit Kumar Jain et al. [3] presented a comprehensive review of all well-known Phishing attacks and their apparent outcomes. In addition, a comparison of the different machine learning based approaches for phishing detection is provided, which is of great use. This opens up new possibilities for the future development of efficient solutions via the use of machine learning. Abdulghani Ali Ahmed et al. [1] propose a method for detecting phishing websites by comparing the Uniform Resource Locators (URLs) of questionable web pages with a set of five extracted attributes. The suggested solution's efficacy is evaluated using the Phishtank and Yahoo directory datasets. The completed report proves, therefore, that the detection system is capable of unearthing a wide variety of phishing assaults. There is still a possibility of believing a false alert. To determine whether or not a particular website is a phishing attempt, Shree Jaswal et al. suggested a method in their paper "dynamically heuristic antifraudulence system" [12]. If not, we construct the performance metric by extracting four heuristics from the URL. More efficiency is guaranteed by distinguishing pictures or trademarks under certain situations. Varsharani Ramdas Hawanna et al. suggest using the Alexa Rank and the number of URL-based attributes to identify phishing URLs.

Proposed System

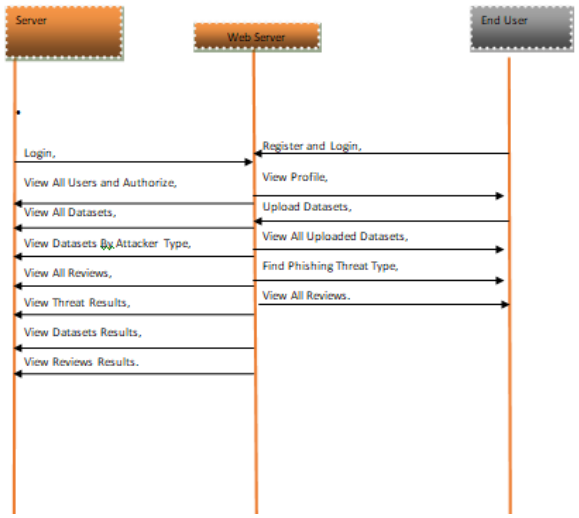
A system for the detection and prevention of phishing websites using a machine learning approach is suggested. A cloud-based approach is utilized to identify potentially malicious websites in the proposed system.
We'll use a training dataset to teach the model, which will be based on a classification technique. The cloud-based architecture will have two-way communication with the browser add-on. The phishing website will be identified based on the URL and characteristics of the website itself. All current client- and server-side processes will be merged into this system. A classifier model trained using the random forest technique will be stored on the server, while a Chrome extension will be developed and installed on users' computers. Here, we anticipate if a website is malicious due to our use of a machine learning approach applied to URLs. Fig -1 shows a schematic of the system's layout.In this article, we will develop a web browser. When we visit a site, a machine learning method will be used to verify the URL. Based on the forecasted outcome, a voice notification will be broadcast. The site will be banned if the projected outcome is unfavorable. Then we can't use that computer to go to the site. Our system's primary benefit is that it prevents us from accessing the banned website using any browser, not only the one we developed.

ARCHITECTURE





➢ **Sequence Diagram**



Results and Analysis:

DATA-BASE

The proposed model will be trained using data made available via the UCI repository [9]. There are a total of 11055 entries; 4,898 of them are for phishing sites, while the remaining 6,157 are for real ones.

ALGORITHM

To classify data, we're use a random forest. By combining numerous decision trees with Bootstrap Aggregation, often known as bagging, a Random Forest[8] is an ensemble approach capable of solving both regression and classification problems. The objective here is to combine the results of many different decision trees to arrive at a single conclusion.

Conclusion:

The system's intended use is to use data mining to discover phishing websites. To do this, we'll parse the user's URL and pull out information about the site's features. The characteristics obtained will be used as validation data for the model. We suggested a system based on a classification Data mining algorithm that is smart, adaptable, and successful in detecting and predicting phishing websites. We used a classification algorithm and several strategies to derive criteria for evaluating the veracity of the phishing data sets. The suggested model may be trained using the Random Forest Algorithm. To prevent users from having their credentials stolen, the system can identify the phishing website and provide a warning in advance.

REFERENCES

[1] Listed as Reference: Ahmed, Abdulghani Ali, and Nurul Amirah Abdullah.
There is now "real-time detection of phishing websites." IEEE's 7th Annual Conference on Information Technology, Electronics, and Mobile Communication (IEMCON), 2016.
[2] Hawanna, Varsharani Ramdas; Kulkarni, V. Y.; and Rane, R. A. "A new algorithm for spotting malicious URLs," it reads.
IEEE, 2016, pp. 548-552, International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT).
[3] Jain, Ankit Kumar, and B. B. Gupta. Phishing detection method comparison using machine learning and feature comparison. 2016 Third International Conference on Computing for Sustainable Global Development (INDIACom). IEEE. pp. 2125-2130.
According to
[4] Tahir, M. Amaad Ul Haq, et al. 2016 IEEE International Conference on Computational Science and Computational Intelligence (CSCI), "A Hybrid Model to Detect Phishing Sites Using Supervised Learning Algorithms." pp. 1126-1133.
[5] Singh, Priyanka; Maravi, Yogendra P.; Sharma, Sanjeev. IEEE, 2015. "Phishing websites detection using supervised learning networks." Proceedings of the 2015 International Conference on Computing and Communications Technologies (ICCCT), pp. 61-65.
2016 Phishing Statistics That You Need to Know, by J. Crowe
[6]. [Online]. Available: https://blog.barkly.com/phishingstatistics-2016.