|| Volume 5 || Issue 12 || January 2021 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH & ET
Multidisciplinary Journal
Double-Blind Peer Reviewed Refereed Open Access International Journal

DETECTION OF PHISHING WEBSITES USING MACHINE LEARNING

Prof Meera Ranadive[1] , Sachin Kumar[2], Rohit Shinde[3], MD Wais[4], Kothari Utkarsh Chandrakant[5]
Dept of Computer Engineering, DYPIET, Ambi Savitribai Phule Pune University, Pune, India.[1,2,3,4,5]

Abstract:- Phishing attack is a simplest way to obtain sensitive information from innocent users. Aim of the phishers is to acquire critical information like username, password and bank account details. Cyber security persons are now looking for trustworthy and steady detection techniques for phishing websites detection. This paper deals with machine learning technology for detection of phishing URLs by extracting and analyzing various features of legitimate and phishing URLs. Decision Tree, random forest and Support vector machine algorithms are used to detect phishing websites. Aim of the paper is to detect phishing URLs as well as narrow down to best machine learning algorithm by comparing accuracy rate, false positive and false negative rate of each algorithm. Keywords: Phishing, Feature Classification, Random Forest classifier, etc

---

I INTRODUCTION

The Technology is growing rapidly day-by-day and with this rapid growing technol- ogy internet has become an essential part of humans daily activities. Use of internet has grown due to the rapid growth of technology and intensive use of digital systems and thus data security has gained great importance. The primary objective of maintaining security in information technologies is to ensure that necessary precau- tions are taken against threats and dangers likely to be faced by users during the use of these technologies. Phishing is the fraudulent attempt to obtain sensitive in- formation such as usernames, passwords and credit card details by disguising as a trustworthy entity in an electronic communication. Typically carried out by email spoofing or instant messaging, it often directs users to enter personal information at a fake website, the look and feel of which is identical to the legitimate site. Information security threats have been seen and developed through time along development in the internet and information systems. The impact is the intrusion of information security through the compromise of private data, and the victim may lose money or other kinds of assets at the end. Internet users can be affected from different types of cyber threats such as private information loss, identity theft, and financial dam- ages. Hence, using of the internet may suspect for home and official environments. Identify and defend against privacy leakage efficient analytical tools are required for users to reduce security threats. Effective systems that can improve self-interventionmust be formed using artificial intelligence-based information security management system at the time of an attack.

II LITERATURE SURVEY:

1. Detecting Phishing Websites Using Machine Learning Amani Alswailem Bashayr Alabdullah Norah Alrumayh Dr.Aram Alsedrani 2019 IEEE The system is based on a machine learning method, particularly supervised learning. Here is selected the Random Forest technique due to its good performance in classification. The focus is to pursue a higher performance classifier by studying the features of phishing websites and choose the better combination of them to train the classifier. As a result, the conclusion is the paper is with accuracy of 98.8

2. A Machine-Learning Framework for Supporting Intelligent Web-Phishing Detection and Analysis Alfredo Cuzzocrea, Fabio Martinelli, Francesco Mercaldo 2019 ACM In particular the system makes use of state-of-the-art decision tree algorithms for detecting whether a Web site is able to perform phishing activities. If this is the case, the Web site is classified as a Web-phishing site. Experimental evaluation confirms the benefits of applying machine learning methods to the well-known web-phishing detection problem.

3. Phishing Web Sites Features Classification Based on Extreme Learning Machine. Yasin S¨onmez1 T¨urker Tuncer2 H¨useyin G¨okal 3 Engin Avc4 2018 IEEE The purpose of this study is to perform Extreme Learning Machine (ELM) based classification for 30 features including Phishing Websites Data in UC Irvine Machine Learning

Repository database. For results assessment, ELM was compared with other machine learning methods such as Support Vector Machine (SVM), Na¨ıve Bayes (NB) and detected to have the highest accuracy of 95.34 percentage

4. Intelligent Phishing Website Detection using Random Forest Classifier Abdulhamit Subasi, Esraa Molah, Fatin Almkallawi, Touseef J. Chaudhery IEEE 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA) In this paper, an intelligent system to detect phishing attacks is presented. We used different data mining techniques to decide categories of websites: legitimate or phishing. Different classifiers were used in order to construct accurate intelligent system for phishing website detection. Classification accuracy, area under receiver operating characteristic (ROC) curves (AUC) and F-measure is used to evaluate the performance of the data mining techniques. Results showed that Random Forest has outperformed best among the classification methods by achieving the highest accuracy 97.36 percentage. Random forest runtimes are quite fast, and it can deal with different websites for phishing detection.

5. Phishing Website Detection Framework Through Web Scraping and Data Mining Andrew J. Park Ruhi Naaz Quadari Herbert H. Tsang 2017 IEEE The focus of this research is to establish a strong relationship between those identified heuristics(content-based) and the legitimacy of a website by analyzing training sets of websites (both phishing and legitimate websites) and in the process analyze new patterns and report findings. Many existing phishing detection tools are often not very accurate as they depend mostly on the old database of previously identified phishing websites. However, there are thousands of new phishing websites appearing every year targeting financial institutions, cloud storage/file hosting sites, government websites, and others. This paper presents a framework called Phishing-Detective that detects phishing websites based on existing and newly found heuristics. For this framework, a web crawler was developed to scrape the contents of phishing and legitimate websites. These contents were analyzed to rate the heuristics and their contribution scale factor towards the illegitimacy of a website. The data set collected from Web Scraper was then analyzed using a data mining tool to find patterns and report findings. A case study shows how this framework can be used to detect a phishing website. This research is still in progress but shows a new way of finding and using heuristics and the sum of their contributing weights to effectively and accurately detect phishing websites.

III SYSTEM ARCHITECTURE:

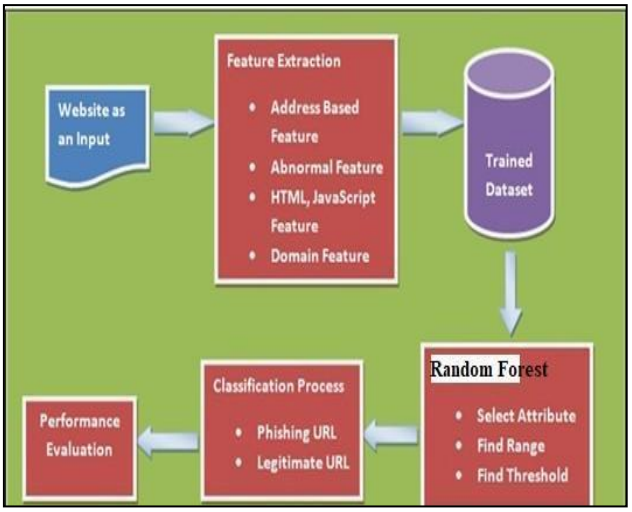

Fig 1: System Architecture

The proposed methodology which imports dataset of phishing and legitimate URLs from dataset and imported data preprocessed. Detecting Phishing Website is performed based on four categories of URL features: domain based, address based, abnormal based and HTML, JavaScript features. These URL features are extracted with processed data and values for each URL attribute are generated. The analysis of URL is performed by machine learning technique which computes range value and the threshold value for URL attributes. Then it is classified into phishing and legitimate URL. The attribute values are computed using feature extraction of phishing websites and it is used to identify the range value and threshold value. The value for each phishing attribute is ranging from -1, 0, 1 these values are defined as low, medium and high according to phishing website feature. The classification of phishing and legitimate website is based on the values of attributes extracted using four types of phishing categories and a machine learning approach.

Algorithm:

Random Forest:

Working of Random Forest Algorithm

We can understand the working of Random Forest algorithm with the help Figure . System Architecture.

Step 1 First, start with the selection of random samples from a given dataset.

Step2 Next, this algorithm will construct a decision tree for every sample.

Then it will get the prediction result from every decision tree.

Step 3 In this step, voting will be performed for every predicted result.

Step 4 At last, select the most voted prediction result as the final prediction

result.

IV CONCLUSION:

Thus we are going to implement a prototype model for phishing website detection using Ml. We are going to develop a system which will efficiently identify the phishing sites. The programming language used will be python.

Future Scope:

There are billions of social media users around the world. However, little is known about social media usages and its determinants that conclusively affect user's vulnerability phishing attacks on social media platforms. Nevertheless, social networking platforms are a popular way used by cyber criminals to swindle their targets. Billions of people logging onto their favorite social media accounts, that is to say, it is a rich source for cybercriminals to gain profit. So, in future we will try to extend our work further for social media platforms like Facebook, Instagram, etc. after successful completion of this proposed work for bachelor of engineering.

REFERENCES:

[1] Matthew Dunlop, Stephen Groat, David Shelly (2010) "

GoldPhish: Using Images for Content-Based Phishing Analysis"

[2] Rishikesh Mahajan (2018) "Phishing Website Detection using Machine Learning Algorithms"

[3] Purvi Pujara, M. B.Chaudhari (2018) "Phishing Website

Detection using Machine Learning : A Review"

[4] David G. Dobolyi, Ahmed Abbasi (2016) "PhishMonger: A Free and Open Source Public Archive of Real-World Phishing Websites"

[5] Satish.S, Suresh Babu.K (2013) "Phishing Websites Detection Based On Web Source Code And Url In The Webpage"

[6] Purvi Pujara, M. B.Chaudhari (2018) "Phishing Website

Detection using Machine Learning : A Review"

[7] Satish.S, Suresh Babu.K (2013) "Phishing Websites Detection Based On Web Source Code And Url In The Webpage"

[8] Tenzin Dakpa, Peter Augustine (2017) "Study of Phishing Attacks and Preventions"

[9] Ping Yi (2018) "Web Phishing Detection Using a Deep Learning Framework"

[10] Jalil Nourmohammadi Khiarak (2017) "What is Machine Learning"

[11] Sadia Afroz, Rachel Greenstadt (2018) "PhishZoo: An

Automated Web Phishing Detection Approach Based on Profiling and Fuzzy Matching"

[12] Arun Kulkarni, Leonard L. Brown (2019) "Phishing Websites Detection using Machine Learning"

[13] Rohan Saraf , Mayur Khatri , Mona Mulchandani (2014) "Phish Tank-A Phishing Detection Tool"

[14] Sadia Afroz, Rachel Greenstadt (2017) "PhishZoo: Detecting Phishing Websites By Looking at Them"

[15] Matthew Dunlop, Stephen Groat, David Shelly (2010) "

GoldPhish: Using Images for Content-Based Phishing Analysis"