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 2, November 2024

Illegitimate Websites Detection Using Deep Learning Framework

Dr. P. C. Latane[1] Vikas Ramdas Takale[2], Prathamesh Kailas Shirke[3], Mugdha Govardhan Khobare[4]

Department of Information Technology[1-4]

Sinhgad Institute of Technology, Lonavala, Maharashtra, India

Abstract: *Phishing is a crime involving robbery of confidential user data. The phishing websites are aimed at individuals, businesses, and cloud storage and government websites. Hardware- based anti-phishing methods are generally used, but software- based approaches are favored because of costs and operational factors. There is no solution to the problem such as zero-day phishing attacks from current phishing detection approaches. A three-phase attack detection called the Phishing Attack Detector based on Web Crawler was proposed to resolve these problems and precisely detect phishing incidences using recurrent neural network. It includes the input features Web traffic, web content and Uniform Resource Locator (URL) based on the classification of phishing and non-phishing pages.*

Keywords: Recurrent Neural Network, Deep Learning, illegitimate URLs, cyberattacks

I. INTRODUCTION

Phishing is a cyber-crime where a person who poses as a legitimate agency contacts a victim or target via email, phone or text message to attract the person to supply information, information about personal identity, banking and credit card information and passwords. Phishing is a crime. The new term ’fishing’ refers to the attacker’s invitation to visit a counterfeit site by creating a website look, and to get personal information from users such as username, password, financial information, account details, national security identifier, etc.. Phishing is a new term that was developed using ’fishing.’The information collected is used for potential target ads or even identity robberies, attacks (for example, money transfer from one’s account). The attack method that is widely used is to send e-mails, messages that can lead to data theft or personal information. Social networking account Passwords, credit cards or attackers provide upgrades to their websites, encourage you to comply with your personal information and change it via fake website. If you are entering your personal data, the attackers will collect it successfully on your server side, and will be able to carry out the next move with your information and to use it for their malicious purposes.

Phishing is described as a reverberation of a website of a remarkable business that snaps private data of consumers, for example usernames, passwords and structured savings numbers. Mail spammers can be categorized with their target in mind. Some telemarketers are spammers who send a few hundred/a large number of e-mail customer’s spontaneous messages. Spammers have the following classification, which continues to randomly send messages, but are near zero enthusiastic. Often they spam or promote materials with irrelevant topics. Some of the cases are sees, knowledgeable news, or statements about meetings. Phishing is itself a new idea, but the criminals, i.e. the phishers, have more and more used it in recent years to steal your personal data and carry out business and social crimes. The number of phishing attacks has significantly risen in four to five years. Phishing is widely used and is easy to carry out on your destination. Phishing usually uses social engineering to attract a victim by submitting a spoofed link to a fake website. The spoofed connection can be found on common web pages or sent to the victim via email. Similar to the legitimate website the fake website is made. So it is directed to the attacker site instead of guiding the victim request to the true web server.

II. LITERATURE SURVEY

This paper provides a broad and comprehensive review of the state of the art in this field by discussing the main challenges and findings. More specifically, the discussion is centered around three important categories of detection approaches, namely, list-based, similarity-based and machine learning-based[1].

DOI: 10.48175/IJARSCT-22156

ISSN 2581-9429 IJARSCT

In this paper, we present CrawlPhish, a framework for automatically detecting and categorizing client-side cloaking used by known phishing websites. We deploy CrawlPhish over 14 months between 2018 and 2019 to collect and thoroughly analyze a dataset of 112,005 phishing websites in the wild[2].

Phishing is the technique by which the attacker tries to obtain confidential information from the user, with the purpose of using it fraudulently. These days, three ways to mitigate such attacks stand out: Focus based on awareness, based on blacklists, and based on machine learning (ML)[3].

Phishing attacks are one of the most common and least defended security threats today. We present an approach which uses natural language processing techniques to analyze text and detect inappropriate statements which are indicative of phishing attacks. Our approach is novel compared to previous work because it focuses on the natural language text contained in the attack, performing semantic analysis of the text to detect malicious intent. To demonstrate the effectiveness of our approach, we have evaluated it using a large benchmark set of phishing emails[4].

III. FLOWCHART

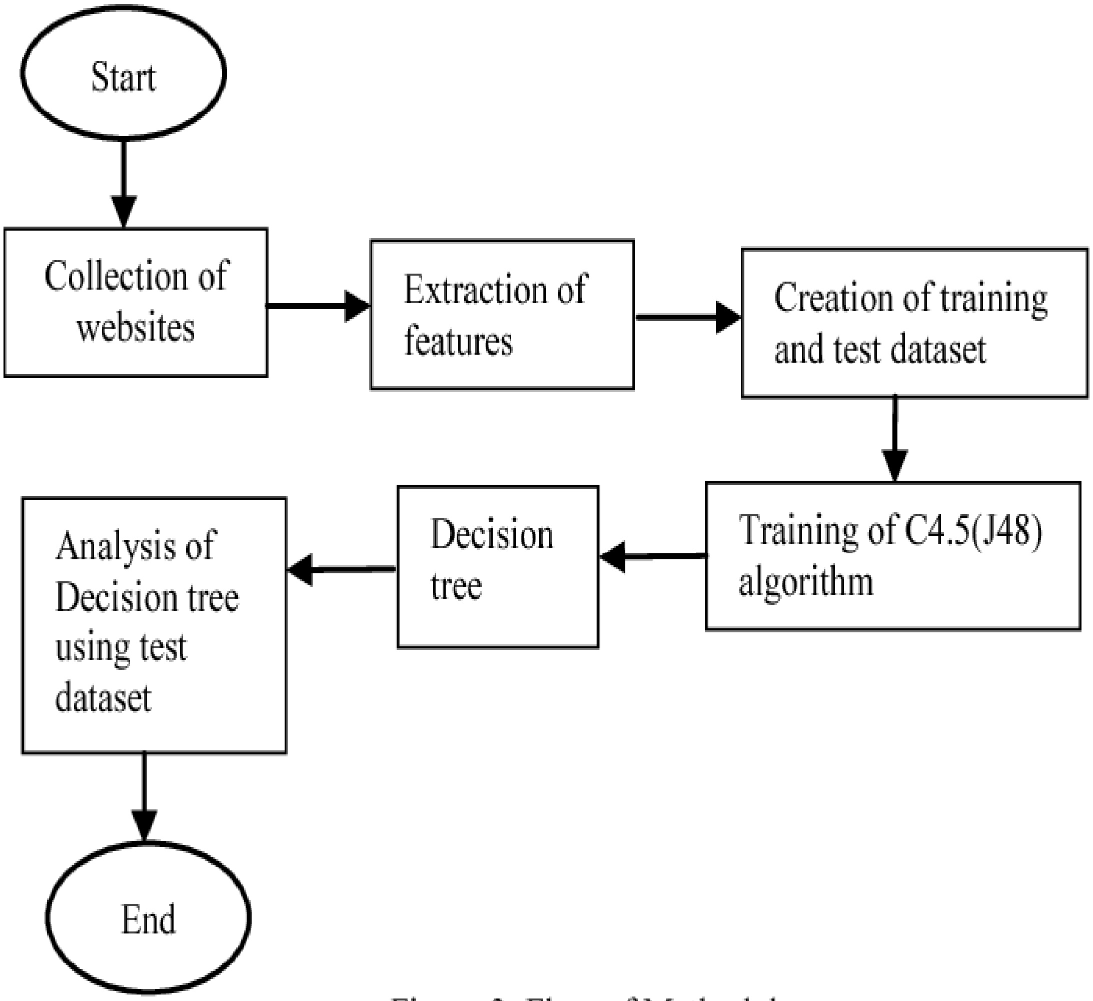
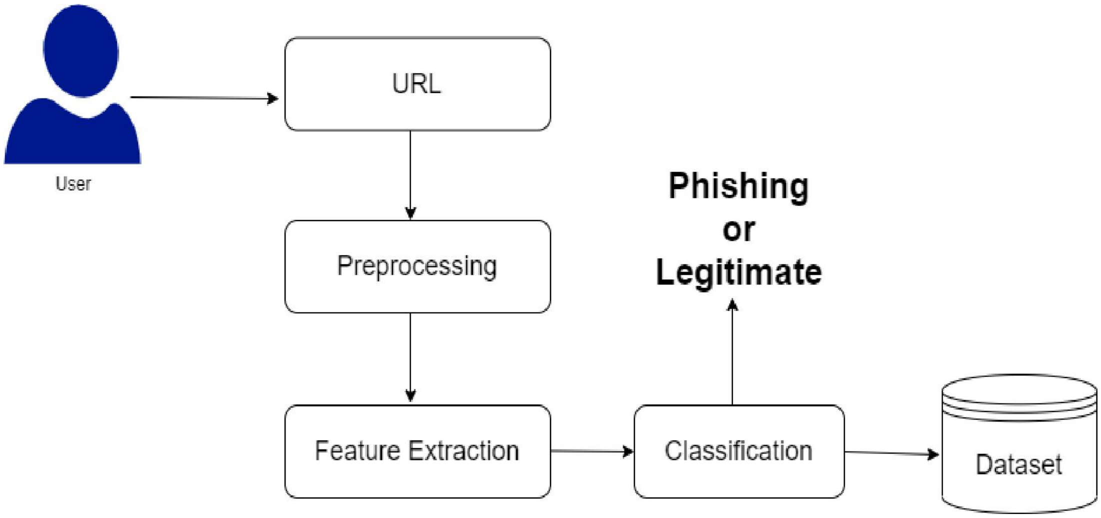

Start
Collection of websites
Extraction of features
Creation of training and test dataset
Training of C4.5(J48) algorithm
Decision tree
Analysis of Decision tree using test dataset
End

Figure 3: Flow of Methodology



IV. METHODOLOGY

A lot of work has been done in this field thanks to its extensive use and applications. This section mentions some of the approaches that have been implemented to achieve the same purpose. These works are mainly differentiated from the techniques for Phishing systems.

The fundamental principle behind the development of such a system is to ensure that financial information for a customer is safe, and so banks and other financial institutions provide various security measures to minimize the risk of unauthorized access to their online account. Online banking has been completely relayed on online transactions through various applications nowadays, so it is most important that this online banking activity is secured.

V. CONCLUSION

Phishing is one of the most damaging web security threats. We have created a prediction model for the detection of Phishing websites by analyzing the attributes of the attack according to our study. The deep-seated learning model of the deep recurrent neural Network overcomes other machine learning models via prediction and achieves the highest precision.

REFERENCES

[1]. RASHA ZIENI , LUISA MASSARI , AND MARIA CARLA CALZAROSSA” Phishing or Not Phishing? A Survey on the Detection of Phishing Websites”.IEEE 2023

[2]. Penghui Zhang; Adam Oest; Haehyun Cho; Zhibo Sun; RC Johnson; Brad Wardman “CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing” 2022 IEEE Symposium on Security and Privacy

[3]. Eduardo Benavides, Walter Fuertes, Sandra Sanchez & Manuel Sanchez “Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review https://link.springer.com/chapter/10.1007/978-981-13-9155-2_5

[4]. T. Peng, I. Harris, and Y. Sawa, “Detecting phishing attacks using natural language processing and machine learning,” in 2019IEEE 12th International Conference on Semantic Computing (ICSC), Jan 2018, pp. 300–301.

[5]. O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, “Machine learning based phishing detection from urls,” Expert Systems with Applications, vol. 117, pp. 345–357, 2020.

DOI: 10.48175/IJARSCT-22156

ISSN
2581-9429
IJARSCT