IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 3, November 2022

Detecting Phishing Website using Machine Learning

Prof. K. S. Mulani[1], Vaibhav Shewale[2], Mansi Salve[3], Niranjan Kale[4], Komal Shinde[5]

Professor, Department of Information Technology[1]

Students, Department of Information Technology[2,3,4,y]

Sinhgad Institute of Technology, Lonavala, Maharashtra, India

Abstract: *Phishing attacks continue to pose a major threat for computer system defenders, often forming the first step in a multi-stage attack. There have been great strides made in phishing detection; however, some phishing emails appear to pass through filters by making simple structural and semantic changes to the messages. We tackle this problem through the use of a machine learning classifier operating on a large corpus of phishing and legitimate emails. We design SAFEPC (Semi-Automated Feature generation for Phish Classification), a system to extract features, elevating some to higher level features, that are meant to defeat common phishing email detection strategies. To evaluate SAFE-PC, we collect a large corpus of phishing emails from the central IT organization at a tier-1 university. The execution of SAFE-PC on the dataset exposes hitherto unknown insights on phishing campaigns directed at university users. SAFEPC detects more than 70a state-of-the-art email filtering tool. It also outperforms Spam Assassin, a commonly used email filtering tool. We also developed an online version of SAFE-PC, that can be incrementally retrained with new samples. Its detection performance improves with time as new samples are collected, while the time to retrain the classifier stays constant.*

Keywords: Detecting Phishing Website, Website management, Safety tips, Safety Requirement.

I. INTRODUCTION

Phishing can be defined as impersonating a valid site to trick users by stealing their personal data comprising usernames, passwords, accounts numbers, national insurance numbers, etc.

Phishing frauds might be the most widespread cybercrime used today. There are countless domains where phishing attack can occur like online payment sector, webmail, and financial institution, file hosting or cloud storage and many others.

The webmail and online payment sector was embattled by phishing more than in any other industry sector. Phishing can be done through email phishing scams and spear phishing hence user should be aware of the consequences and should not give their 100 percent trust on common security application. Machine Learning is one of the efficient techniques to detect phishing as it removes drawback of existing approach

II. LITERATURE REVIEW

Paper Name: Real Time Detection of Phishing Websites

Author: Abdulghani Ali Ahmed, Nurul Amirah Abdullah

Abstract ::- Web Spoofing lures the user to interact with the fake websites rather than the real ones. The main objective of this attack is to steal the sensitive information from the users. The attacker creates a 'shadow' website that looks similar to the legitimate website. This fraudulent act allows the attacker to observe and modify any information from the user. This paper proposes a detection technique of phishing websites based on checking Uniform Resources Locators (URLs) of web pages. The proposed solution is able to distinguish between the legitimate web page and fake web page by checking the Uniform Resources Locators (URLs) of suspected web pages. URLs are inspected based on particular characteristics to check the phishing web pages. The detected attacks are reported for prevention. The performance of the proposed solution is evaluated using Phistank and Yahoo directory datasets. The obtained results show that the detection mechanism is deployable and capable to detect various types of phishing attacks maintaining a low rate of false alarm

DOI: 10.48175/IJARSCT-7596

III. PROPOSED WORK

The purpose of this project is the link Guard algorithm is the idea for finding the phishing electronic messages sent by the phisher to get hold on the data of the end user.

Link Guard depends on investigation of the attributes of phishing hyperlinks. Each end user is implemented with Link Guard algorithm.

Since Link Guard is qualities based, it can identify and prevent not only known phishing but also obscure ones Subsequent to doing as such, the end user perceives the phishing emails and can abstain from responding to such mails.

IV. FEATURE SCOPE

- The following category of features are selected:
- Address Bar based Features
- Domain based Features
- HTML &JavaScript based Feature
- Address Bar based Features considered are:
- Domain of URL • Redirection '//' in URL
- IP Address in URL
- 'http/https' in Domain name
- '@' Symbol in URL
- Using URL Shortening Service
- Length of URL

V. UML DIAGRAMS

5.1 Class Diagram

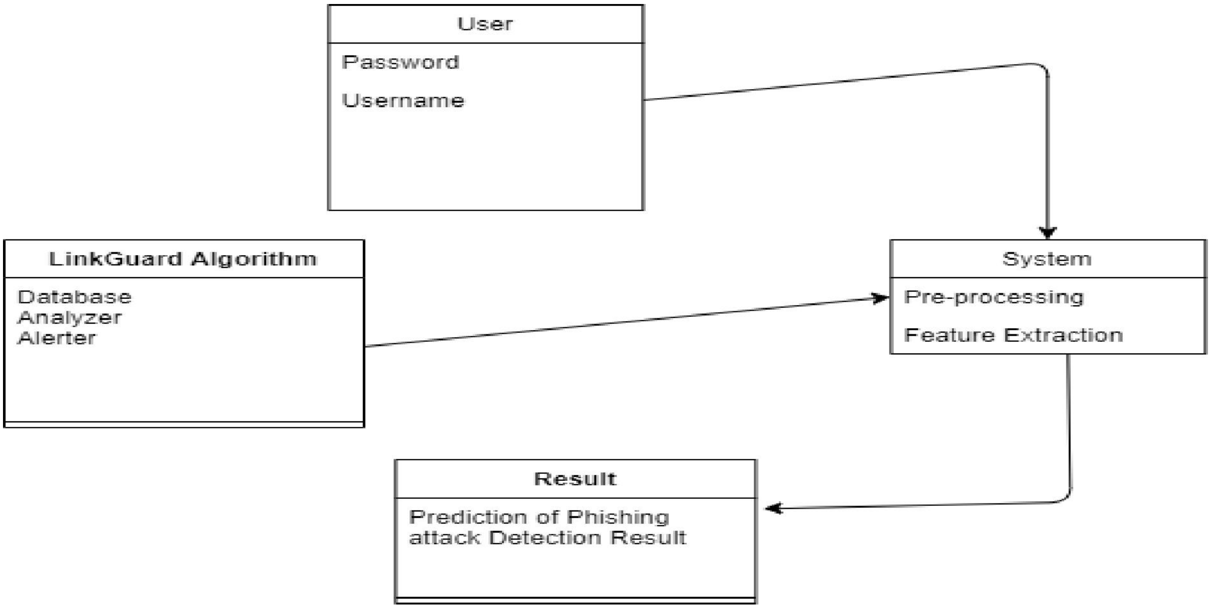

Figure 1: Class Diagram for Detecting Phishing Website Using Machine Learning

DOI: 10.48175/IJARSCT-7596

5.2 Use Case Diagram

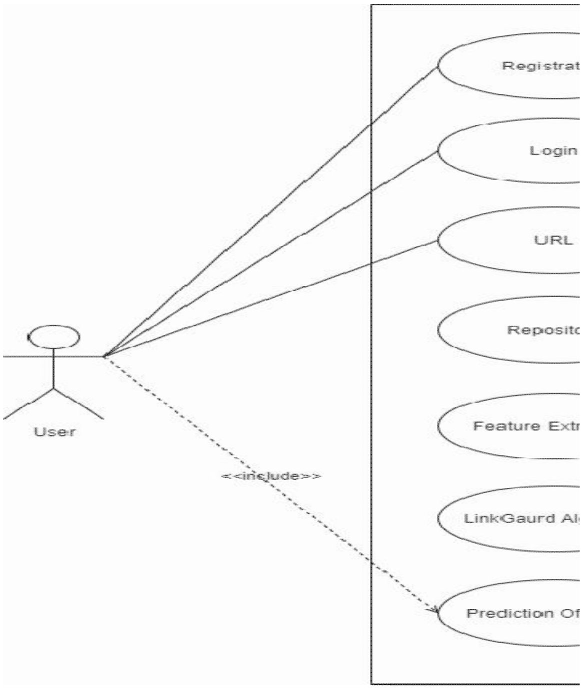

Figure 2: Use Case Diagram Detecting Phishing Website Using Machine Learning

VI. HARDWARE AND SOFTWARE REQUIREMENTS

6.1 Hardware Requirements

- Processor: Intel i5
- Hard Disk: 500 GB.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram: 8 GB

6.2 Software Requirements

- Operating system : Windows 10,11
- Coding Language: Python
- IDE: Spyder
- Database:Xampp

VII. APPLICATIONS

This study is useful for the users to prevent themselves from phishing activities which lure individuals into providing sensitive data such as personally identifiable information, banking and credit card details and passwords.It is very interactive. Easy to use.

Effective system for preventing URL attacks

VIII. METHODOLOGY

- The Phishing is a malicious website that impersonates as a legitimate one to get sensitive data like credit card number or bank account password.
- A phisher uses social engineering and technical deception to fetch private information from the web user.
- The phishing web pages generally have alike page layouts, blocks and fonts to mimic legitimate web pages in an endeavour to influence web users to obtain personal details such as username and password

DOI: 10.48175/IJARSCT-7596

IX. CONCLUSION

The Phishing has become a serious network security problem, causing financial loss to both consumers and e-commerce companies. In this paper we've discussed about implemented system, using link guard algorithm

ACKNOWLEDGMENT

The authors would like to thank Assistant Professor of the Department of Information Technology in Sinhgad institute of technology, Lonavala Prof. K. S. Mulani for their time and efforts he provided throughout the project and his advice and suggestions were really helpful to us.

REFERENCES

[1]. “WC-PAD: Web Crawling based Phishing Attack Detection” Nathezhtha.T, Sangeetha.D,Vaidehi.V

[2]. “Detection of Phishing Attacks with Machine Learning Techniques in Cognitive Security Architecture” Ivan Ortiz-Garces, Roberto O. Andrade, and Maria Cazares

[3]. “A Methodical Overview on Phishing Detection along with an OrganizedWay to Construct an Anti-Phishing Framework” Srushti Patil, Sudhir Dhage

[4]. “A survey of the QR code phishing: the current attacks and countermeasures” Kelvin S. C. Yong, Kang Leng Chiew and Choon Lin Tan

[5]. “Fuzzy Rough Set Feature Selection to Enhance Phishing Attack Detection” Mahdieh Zabihimayvan and Derek Doran

BIOGRAPHY

- VaibhavShewale- an Undergraduate Scholar pursuing Bachelors of Engineering in Information Technology from Sinhgad Institute of Technology. He is working under the guidance of Prof. K. S. Mulani
- Mansi Salve- an Undergraduate Scholar pursuing Bachelors of Engineering in Information Technology from Sinhgad Institute of Technology. She is working under the guidance of Prof. K. S. Mulani
- Niranjan Kale - An Undergraduate Scholar pursuing Bachelors of Engineering in Information Technology from Sinhgad Institute of Technology. He is working under the guidance of Prof. K. S. Mulani
- Komal Shinde - An Undergraduate Scholar pursuing Bachelors of Engineering in Information Technology from Sinhgad Institute of Technology. She is working under the guidance of Prof. K. S. Mulani

DOI: 10.48175/IJARSCT-7596