ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2, 2024

DETECTION OF PHISHING WEBSITES USING URLS

[1]Minhaj Begum, [2]K Krithika Reddy, [3]Tipirishetty Lahari, P Madhuri Sri

[1]Assistant professor in Department of Information Technology Bhoj Reddy Engineering College for Women

[2,3,4] UG Scholars in Department of Information Technology Bhoj Reddy Engineering College for Women

Abstract

Phishing remains a persistent and impactful security threat, posing serious risks to individuals and targeted brands. Despite its longstanding presence, this malicious activity continues to thrive, adapting and refining tactics over time to enhance its effectiveness. This article aims to underscore the critical importance of identifying and thwarting phishing websites, which serve as a significant conduit for these attacks. The landscape of phishing attacks is dynamic, with threat actors constantly evolving their strategies to create more convincing and potent schemes. Recognizing the gravity of this ongoing threat, we delve into an overview of the state of the art in phishing detection, emphasizing three primary categories of detection approaches: list-based, similarity-based, and machine learning-based methods. List-based approaches rely on predefined lists of known phishing websites, allowing for the identification of malicious sites based on historical data and reported instances. Simultaneously, similarity-based techniques leverage patterns and similarities with established legitimate websites, aiming to flag deviations indicative of potential phishing attempts. Machine learning-based approaches, on the other hand, harness the power of advanced algorithms to analyze vast datasets, learning to distinguish between legitimate and malicious sites based on various features and patterns. This comprehensive review encompasses an examination of the diverse detection methods proposed in existing literature, highlighting the strengths and limitations of each approach. Additionally, the article delves into the datasets commonly employed for assessing the efficacy of these detection methods, shedding light on the empirical foundations of current research in the field. Despite significant advancements, there are notable research gaps that warrant further exploration. These gaps may include he need for more robust and diverse datasets, improved algorithmic sophistication, and a deeper understanding of emerging phishing tactics.

Keywords: Phishing, phishing detection, cyber security

I INTRODUCTION

Phishing is a dangerous security threat that exploits sophisticated psychological and social engineering techniques to trick individuals into clicking links of malicious websites and submit highly valuable sensitive information, such as personal or corporate information and account credentials. Phishing attacks are far from being technologically complex and their deployment requires little effort. Nevertheless, they are

generally very effective. Attackers create well crafted phishing websites with a look and feel of the legitimate sites they are trying to impersonate, thus making it very challenging for individuals to identifyphishing sites. In addition, to avoid being detected, attackers have refined over the years their tactics and evasion techniques. Phishing attacks have several direct and indirect impacts. They affect the individuals being phished, whose identity and accounts might be compromised, thus leading to money being stolen as well as to a potential crisis of trust towards online services. These attacks also affect the companies and organizations being impersonated, whose brands might be abused, thus leading to potential data breaches, financial losses and reputation damages.A study by Enisa reveals that phishing attacks are among the most common cyber incidents European small medium enterprises are likely to be exposed to. In the Cybersecurity threat rends report Cisco suggests that in 2020 phishing accounts for around 90% of data breaches. Moreover, 86% of organizations had at least one user try to connect to a phishing site. In fact, individuals tend to fall prey of phishing attacks especially because of the insufficient attention paid in assessing the legitimacy of a website and the lack of appropriate education

II LITERATURE SURVEY

CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing:

Phishing is a critical threat to Internet users. Although an extensive ecosystem serves to protect users, phishing websites are growing in sophistication, and they can slip past the ecosystem's detection systems—and subsequently cause real-world damage—with the help of evasion techniques. Sophisticated client-side evasion techniques, known as cloaking, leverage JavaScript to enable complex interactions between potential victims and the phishing website, and can thus be particularly effective in slowing or entirely preventing automated mitigations. Yet, neither the prevalence nor the impact of client-side cloaking has been studied. 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. By adapting state- of-the-art static and dynamic code analysis, we find that 35,067 of these websites have 1,128 distinct implementations of client-side cloaking techniques. Moreover, we find that attackers' use of cloaking grew from 23.32% initially to 33.70% by the end of our data collection period. Detection of cloaking by our framework exhibited low false-positive and false-negative rates of 1.45% and 1.75%, respectively. We analyze the semantics of the techniques we detected and propose a taxonomy of eight types of evasion across three high-level

categories: User Interaction, Fingerprinting, and Bot Behavior. Using 150 artificial phishing websites, we empirically show that each category of evasion technique is effective in avoiding browser-based phishing detection (a key ecosystem defense). Additionally, through a user study, we verify that the techniques generally do not discourage victim visits. Therefore, we propose ways in which our methodology can be used to not only improve the ecosystem's ability to mitigate phishing websites with client-side cloaking, but also continuously identify emerging cloaking techniques as they are launched by attackers.

Why phishing still works: User strategies for combating phishing attacks:

We have conducted a user study to assess whether improved browser security indicators and increased awareness of phishing have led to users′ improved ability to protect themselves against such attacks. Participants were shown a series of websites and asked to identify the phishing websites. We use eye tracking to obtain objective quantitative data on which visual cues draw users′ attention as they determine the legitimacy of websites. Our results show that users successfully detected only 53% of phishing websites even when primed to identify them and that they generally spend very little time gazing at security indicators compared to website content when making assessments. However, we found that gaze time on browser chrome elements does correlate to increased ability to detect phishing. Interestingly, users′ general technical proficiency does not correlate with improved detection scores

A survey of phishing attacks: Their types, vectors and technical approaches:

Phishing was a threat in the cyber world a couple of decades ago and still is today. It has

grown and evolved over the years as phishers are getting creative in planning and executing the attacks. Thus, there is a need for a review of the past and current phishing approaches. A systematic, comprehensive and easy-to-follow review of these approaches is presented here. The relevant mediums and vectors of these approaches are identified for each approach. The medium is the platform which the approaches reside and the vector is the means of propagation utilised by the phisher to deploy the attack. The paper focuses primarily on the detailed discussion of these approaches. The combination of these approaches that the phishers utilised in conducting their phishing attacks is also discussed.This review will give a better understanding of the characteristics of the existing phishing techniques which then acts as a stepping stone to the development of a holistic anti-phishing system. This review creates awareness of these phishing techniques and encourages the practice of phishing prevention among the readers. Furthermore, this reviewwill

gear the research direction through the types of phishing, while also allowing the identification of areas where the anti-phishing effort is lacking. This review will benefit

A comprehensive survey of AI-enabled phishing attacks detection techniques

In recent times, a phishing attack has become one of the most prominent attacks faced by internet users, governments, and service-providing organizations. In a phishing attack, the

attacker(s) collects the client's sensitive data (i.e., user account login details, credit/debit card numbers, etc.) by using spoofed emails or fake websites. Phishing websites are common entry points of online social engineering attacks, including numerous frauds on the websites. In such types of attacks, the attacker(s) create website pages by copying the behavior of legitimate websites and sends URL(s) to the targeted victims through spam messages, texts, or social networking. To provide a thorough understanding of phishing attack(s), this paper provides a literature review of Artificial Intelligence (AI) techniques:Machine Learning, Deep Learning, Hybrid Learning, and Scenario-based techniques for phishing attack detection. This paper also presents the comparison of different studies detecting the phishing attack for each AI technique and examines the qualities and shortcomings of these methodologies. Furthermore, this paper provides a comprehensive set of current challenges of phishing attacks and future research direction in this domain

III EXISTING SYSTEM

Phishing is a popular topic that has been researched over the years under different perspectives. Several surveys have summarized the state of the art in this field. Some of them analyze general aspects, such as attack strategies, training and education approaches, whereas some others specifically focus on detection and prevention approaches. In literature they introduced phishing mitigation techniques. A high-level overview of various categories of phishing mitigation techniques is also presented, such as: detection, offensive defense, correction, and prevention, which we belief is critical to present where the phishing detection techniques fit in the overall mitigation process another researcher gives a detailed review of the strategies offered in the literature for the detection of phishing websites. These strategies are subdivided in six categories according to the techniques they are based upon, namely, search-based, heuristics and machine learning, black and whitelists, DNS based, visual similarity, and proactive phishing URL based techniques. The advantages and disadvantages of the various strategies are discussed in detail.

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2, 2024

Problems in existing system:

1. The short URLs affect machine learning based approaches since most URL features suggested in the literature become meaningless in this context, thus making the detection mechanisms fail.

2. Insufficient attention paid in assessing the legitimacy of a website and the lack of appropriate education.

3. Other open issues associated with features are related to their relationships with the evasion techniques implemented by attackers.

4. In addition, the existing surveys mainly focuses the detection approaches based on heuristics, while machine learning approaches are covered to a rather limited extent

. 5. In general, it is not sufficient to retrain a machine learning model whenever new data becomes available, instead there is the compelling need to quickly identify the tactics used by these ever-evolving attacks and automatically extract appropriate features.

6. Hence, further research efforts should be dedicated to investigate these issues.

IV PROPOSED SYSTEM

By addressing the key issues and discoveries, this paper offers an extensive and thorough evaluation of the state of the art in this area. More particular, three significant categories of detection approaches—list-based, similarity-based, and machine learning-based—are the focus of the discussion. We present the detection strategies suggested in the literature for each category, along with the datasets used for their evaluation, and we talk about certain unmet research needs

V IMPLEMENTATION

Module is a part of a program. Programs are composed of one or more independently developed modules. A module description provides detailed information about a module and its supported components. The included description is available directly by making environment check for supported components. The modules are

- Data exploration: Using this module we will load data into system.
- Processing: Using this module we will read data for processing.
- Splitting data into train & test: Using this module data will be divided into train & test

. • Model generation: Model building - SVM - Random Forest - KNN - Decision Tree – Logistic Regression - CNN (Convolutional Neural Network). Algorithms accuracy calculated.

- User signup & login: Using this module we will be able to register and login into the website.
- User input: Using this module we will give input for predicting.
- Prediction: Using this module final prediction will be displayed.

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2, 2024

VI CONCLUSION

The project offers a thorough examination of current methods for detecting phishing websites, addressing key challenges, discoveries, and areas for further research. It delves into diverse detection strategies, datasets utilized, and existing gaps in knowledge. Emphasis is placed on the critical selection of features, considering their strengths, weaknesses, discriminating capabilities, and alignment with attacker tactics, while adhering to the principle of simplicity.

REFERENCES

[1] N.A. Azeez, S. Misra, I.A. Margaret, L. Fernandez-Sanz, and S.M. Abdulhamid, ‘‘Adopting automated whitelist approach for detecting phishing attacks,’’ Comput. Secur., vol. 108, Sep. 2021, Art. no. 102328.

[2] W. Khan, A. Ahmad, A. Qamar, M. Kamran, and M. Altaf, ‘‘SpoofCatch: A clientside protection tool against phishing attacks,’’ IT Prof., vol. 23, no. 2, pp. 65–74, Mar. 2021.

[3] Y. Lin, R. Liu, D.M. Divakaran, J. Ng, Q. Chan, Y. Lu, Y. Si, F. Zhang, and J. Dong, ‘‘Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages,’’ in Proc. 30th USENIX Secur. Symp., 2021, pp. 3793–3810.

[4]A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, ‘‘A comprehensive survey of AI-enabled phishing attacks detection techniques,’’ Telecommand. Syst., vol. 76, no. 1, pp. 139–154, Jan. 2021.