WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Comprehensive Review of Website Defacement Technique

Malavika N

Dept. Computer Science and Engineering WYD23CSNS05
malavikan878@gmail.com
DOI : https://doi.org/10.55248/gengpi.6.0125.0627

ABSTRACT—

Web attacks, particularly website defacement at- tacks, pose significant challenges in the realm of web security. In recent times, these attacks have emerged as primary security threats for numerous organizations and governmental entities providing web-based services. The repercussions of website defacement attacks extend beyond financial and data losses, affecting users, website owners, and even contributing to political and economic problems. To address this issue, various detection techniques and tools have been employed to identify and monitor website defacement attacks.

While some detection techniques focus on static web pages, dynamic web pages, or a combination of both, they often grapple with the challenge of minimizing false alarms. Numerous approaches exist for detecting web defacement, utilizing a range of methods such as online tools, comparison, and classification techniques. The evaluation criteria for these techniques typically revolve around achieving high detection accuracies, often target- ing 100

This paper conducts a comprehensive literature review of prior works related to website defacement. The review compares these works based on their accuracy results, the employed techniques, and highlights the most efficient approaches in the context of website defacement detection.

Keywords: website defacement; machine learning; web security

I. Introduction

A website defacement attack involves exploiting vulnerabil- ities in a website or web server to deploy malicious code with the intent of altering, modifying, or deleting web page content. The alterations, often conveyed through text, images, or both, serve various purposes such as personal or political motives or simply for amusement, and may even involve blocking the entire web page. The repercussions of such attacks extend to both financial losses and damage to the reputation of the targeted entity, as illustrated in Figure 1.

Perpetrators of these attacks may employ brute-force tech- niques, identifying vulnerable points within websites and executing methods like SQL injection or cross-site scripting (XSS). They may also disseminate malware to website ad- ministrators. Defensive measures are crucial to mitigate such risks, including regular system updates, the implementation of monitoring and detection tools, and educating employees about potential threats.

Detection and monitoring systems play a pivotal role in preventing recurrent attacks by identifying and thwarting them in real-time. Once an attack is detected, it sheds light on weak- nesses or flaws in the system or website, allowing for prompt remediation. In the absence of robust detection or prevention systems, websites remain susceptible to defacement attacks. This paper introduces various techniques for monitoring web- sites and outlines measurements to detect website defacement, irrespective of whether the web page is static, dynamic, or a combination of both. Website defacement incidents may have implications for both website users and owners.

II. RELATED WORKS

A. N-Gram Method

The N-gram method, or N meta-model, serves as a text or language processing tool employed for the classification and categorization of text. It demonstrates effectiveness in extracting features from text in various languages and finds application in filtering, monitoring, and sorting emails, news, scientific articles, and other high-value information. This method involves constructing N-grams, where N represents the gram value, such as N = 3, to break down the text into tokens. For instance, the word ”beautiful” with N = 3 results in tokens like ”bea,” ”eau,” ”aut,” ”uti,” ”tif,” ”ifu,” and ”ful.” The word length (W) is considered as w-2 to ensure reliable classification.

In assessing the outcomes of an experiment, four criteria play a crucial role in the final results, facilitating comparisons and enhancements for more accurate findings: - True Positive (TP): Correctly classifying a defaced website. - True Negative (TN): Accurate classification of a legitimate website. -

False Positive (FP): Incorrectly classifying a legitimate website as defaced. - False Negative (FN): Incorrectly classifying a defaced website as legitimate.

B. . Machine-Learning-Based Techniques

Wu et al. [3] conducted a study introducing a novel classification model utilizing three machine-learning algo- rithms—support vector machine (SVM), random forest (RF), and gradient-boosted decision trees (GBDT)—to discern web- site defacement. Their approach involved building a classifier by extracting web pages and Trojan features for use in the classification step. The model was evaluated on a dataset con- taining 4620 websites from various sources, employing cross- validation. The evaluation criteria included true positive (TP), true negative (TN), false positive (FP), and false negative (FN) measurements. While significant differences were observed in the performance of the three algorithms, SVM demonstrated the highest accuracy with a false positive rate below 1

Dau Hoang [6] proposed a website defacement detection method based on machine-learning techniques, inspired by Woonyon Kim et al. [1]. Hoang incorporated n-gram methods and occurrence frequency for dynamic web page defacement detection, reducing false alarms through threshold adjustment. The experiment showed good results, but limitations were identified, particularly regarding highly dynamic web pages. Hoang achieved high detection accuracy of over 93

Building on Hoang's previous work [5], Hoang et al. [6] proposed a hybrid defacement detection method combining machine-learning techniques and attack signatures. They used supervised machine-learning algorithms, naive Bayes and ran- dom forest, to classify web page HTML code on a large dataset containing English and Vietnamese web pages. The method exhibited improved performance on both English and Vietnamese web pages, achieving accuracy rates exceeding 99.26

Hoang et al. [7] introduced a multi-layer model for website defacement detection, focusing on integrity checks as the final step. The model demonstrated high detection accuracy of over 98.80

Dau Hoang et al. [10] proposed a CNN-based model for website defacement detection, deviating from traditional machine-learning algorithms. They achieved high detection accuracy of 98.86

Nguyen et al. [9] proposed a combination model for website defacement detection, incorporating text and image features with deep-learning techniques. The model achieved high ac- curacy of 97.49

In summary, these studies employed various machine- learning and deep-learning techniques, including SVM, ran- dom forest, gradient-boosted decision trees, naive Bayes, J48 decision tree, and CNN, to develop effective website deface- ment detection methods. Each approach had its strengths and limitations, and the choice of technique depended on factors such as dataset characteristics and computational resources. The integration of attack signatures, n-gram methods, and integrity checks showcased the diversity of strategies in en- hancing detection accuracy and reducing false alarms.

C. Based on Other Tools

Over the years, a diverse range of techniques has emerged for detecting and preventing website defacement. While machine-learning detection techniques are powerful, they can be resource-intensive and slow due to the need for extensive data and the application of multiple techniques for accurate results. As an alternative, various tools have been developed to swiftly detect and prevent defacement, offering faster solutions compared to machine-learning approaches. Figure 9 illustrates the functioning of these tools, and Tables 2 and 3 provide an overview of studies utilizing tools for website defacement detection or prevention.

Mfundo et al. [2] addressed the common threat of website defacement and introduced the Web Defacement and Intrusion.

Monitoring Tool (WDIMT). This tool swiftly detects deface- ment, issues rapid alerts, and facilitates the restoration of a web page's original content. The WDIMT's architecture comprises three layers: the presentation layer for user interaction, the business layer for database interactions, and the data access layer for storing user data and web page hashes. The WDIMT, executable in a Linux environment, offers a visual representa- tion of web page status through a terminal. Figure 10 outlines the flowchart of the WDIMT process.

In [2], Tran Dac Tot et al. emphasized the importance of promptly detecting changes in a website's interface and con- tent. They proposed a method combining local area network (LAN) and remote monitoring, utilizing hash functions (specif- ically MD5) for server and database monitoring. Employing the Boyer–Moore algorithm for content change detection in HTML documents, their method integrates the C4.5 algorithm for enhanced security alerts accuracy. The proposed approach collects and compares information to identify potential alter- ations in website content.

III. CONCLUSION

Securing websites on the internet is imperative, and there is a growing need for effective approaches to enhance web infrastructure security. As web-based services continue to proliferate, it becomes crucial to construct highly secure web servers to mitigate potential threats. Among the significant threats facing the internet, website defacement stands out as a critical concern. Our comprehensive review focused on website defacement detection techniques and tools, providing insights into the implementation details, machine-based learn- ing techniques, and other tools employed in this domain.

Defacement detection techniques can be broadly categorized into three groups: anomaly-based detection, signature-based detection, and machine-learning techniques. The specific fo- cus of our review was on defacement detection leveraging machine-learning algorithms, delving into the nuances

of each research study. Notably,the paper emphasized the effectiveness of a defacement detection model that combines machine- learning techniques with attack signatures, showcasing an impressive accuracy rate of 99.26

Various techniques were explored in the reviewed studies, including WDIMT, random monitoring, dork search engine, JavaScript, TPM, OWASP Zed Attack Proxy (ZAP), Acunetix, Burp Suite, and self-protection mechanisms. A comparative analysis of these techniques revealed variations in accuracy and false positive rates. Looking ahead, the future endeavors aim to identify practical outcomes against website defacement attacks. The envision repeating the study with novel website defacement detection and monitoring methods.

References

[1] Rajiv Kumar Gurjwar, Divya Rishi Sahu, Deepak Singh Tomar ”An Approach to Reveal Website Defacement” 2014

[2] Mfundo Masango, Francois Mouton, Palesa Antony and Bokang Man- goale ”Web Defacement and Intrusion Monitoring Tool: WDIMT” 2018

[3] Siyan Wu, Xiaojun Tong, Wei Wang, Guodong Xin, Bailing Wang*, Qi Zhou ”Website Defacements Detection Based on Support Vector Machine Classification Method” 2018

[4] Xuan Dau Hoang ”A Website Defacement Detection Method Based on Machine Learning” 2019

[5] Barerem-Melgueba Mao, Kanlanfei Damnam Bagolibe ”A contribution to detect and prevent a website defacement” 2019

[6] Xuan Dau Hoang , Ngoc Tuong Nguyen ”Detecting Website Deface- ments Based on Machine Learning Techniques and Attack Signatures” 2019

[7] Xuan Dau Hoang, Ngoc Tuong Nguyen ”A Multi-layer Model for Website Defacement Detection” 2019

[8] Kevin Borgolte, Christopher Kruegel, Giovanni Vigna ”Meerkat: De- tecting Website Defacements through Image-based Object Recognition” 2020

[9] Trong Hung Nguyen, Xuan Dau Hoang, Hanoi ”Detecting Website Defacement Attacks using Web-page Text and Image Features” 2021

[10] Hoang Xuan Dau, Nguyen Trong Hung ”A CNN-BASED MODEL FOR DETECTING WEBSITE DEFACEMENTS” 2021