Global Scientific JOURNALS

GSJ: Volume 13, Issue 6, June 2025, Online: ISSN 2320-9186
www.globalscientificjournal.com

DETECTION APPROACHES FOR WEB DEFACEMENT ATTACKS

Islam Uddin

Department of Computer Science, Univerity of Engineering & Techology Peshawar, Pakistan
Email: islamuop@gmail.com

KeyWords
Detection; Web Defacement; Web Applications; Websites

ABSTRACT

Defacement of web attacks significantly threatens web security and integrity, often resulting in loss of finances, reputation, and spread of misinformation. Websites serve as critical online platforms for commerce, communication, and information dissemination. For many organizations and government agencies that provide web-based services, defacement of web attacks has become the main security threat. The attacks are usually politically motivated, ideologically motivated, or intended to discredit an institution. As web applications are becoming dynamic and content-rich, defacement has become more challenging to detect, especially subtle or partial. Traditional detection methods, such as file integrity monitoring, checksums, and signature-based intrusion detection systems, are susceptible to missing small or hidden changes, especially those introduced by injecting content or by taking advantage of application-level vulnerabilities.

This paper offers a comprehensive overview of the existing web defacement detection methods and classifies them as static, dynamic, and hybrid methods. Through critical examination, we highlight the limitations of existing methods, including high false positives, delayed detection, ineffectiveness in responding to new threats, and low-visibility content spoofing detection. We then present an innovative detection method that combines real-time content matching, DOM structure matching, and machine learning-based anomaly detection for improved responsiveness and accuracy. Our method is capable of detecting not only gross defacements, i.e., homepage hijacking and banner replacement, but subtle modifications, i.e., unauthorized sentence-level changes or hidden scripts.

We developed a prototype system to test and utilize the suggested method. Experimental results show that there is dramatic improvement in detection accuracy and recall over current approaches. The system can effectively detect minor defacements that are not detected by other tools, but with minimal false positives. The results of this work are added to the development of stronger, adaptive, and intelligent web defacement detection systems to protect digital assets in real-time.

INTRODUCTION

In this era of digitalization, the evolution of human society, economy, politics, and social environments is promoted by continuous development. The amount of information can easily be accessed via the internet (a network of networks). The main entity of the internet is a website consisting of various interconnected web pages. A web server is a system for hosting websites, whereas a web browser is a gateway for displaying websites on the internet. As a result, websites hosted by web servers and accessed through web browsers spread a tremendous amount of information that is accessible from anywhere in the world. As these websites are accessed by billions of users globally, the chances of cybercrime attacks have increased manifold. Thus, the activity of solving this diverse problem, involving software and hardware, and inhibiting the prevalence of these cybercrimes is called cybersecurity [1].

In this paper, we describe the existing techniques used in literature for the detection of web defacement attacks.

LITERATURE REVIEW

Web defacement scientific literature is scarcer and the availability of dealing literature with hacktivism and website defacements are even poor [2]. This section focuses on the existing methodologies and tools used in the detection of defacement attacks on the web. Different techniques, approaches, and methodologies, such as signature-based, hybrid, machine learning, and anomaly-based approaches, are presented. This section also focuses on the types of web defacement attacks and the impact of those attacks on the surrounding environment. Furthermore, the study analyzes the key strengths and limitations of each approach. This chapter compares the proposed study who the existing literature in terms of adaptability, scalability, reliability, false positive detection, and real-time detection. By thoroughly evaluating the current work, this section focuses on the foundation for the research study and justification for the need for a more efficient and robust system for the detection of defacement attacks on the web [2].

When a cyber attacker gains access by unauthorized means to a website and modify its content, functionality and visual appearance, it is known as web defacement attack. Vulnerabilities are exploited such as Remote File Inclusion (RFI), Cross- Site Scripting (XSS) and SQL Injection to execute attacks by attackers. The main motivation behind these types of attacks ranges from financial gain to political statements or just by showcasing their hacking skills

Signature-based Detection

The signature-based detection technique compares the current state of a website to a previously recognized signature or baseline It plays a vital role in the arsenal of security components for most of the organizations. Signature-based detection methods are amongst the oldest and most popular web defacement attack detection methods. These methods continuously scan files, scripts, or content of a site being monitored against a database of known attack behavior signatures, file changes, or malicious content. The instant a match is found, an alarm is triggered, and a suspected defacement is reported. Signature-based tools will generally search for exact wording, altered scripts, changes to primary files like index.html or home page banners, or unauthorized changes to stylesheets and embedded code. Signature-based systems excel at finding well-documented and well-known attacks as they are easy and fast. They are suitable where threats are known beforehand and are deterministic [3-4].

However, signature-based detection has several limitations that restrict its effectiveness against new or sophisticated defacement attacks. Since these systems can only detect threats whose signatures are known, they are inherently blind to zero-day attacks or minor variations that lie outside the known paradigm. Attackers can simply bypass such systems by making minor variations to their approaches or payloads to remain undetected. Furthermore, updating the signature database and maintaining it current requires constant human effort and fails to scale in ever-changing web environments. Signature-based detection can thus be a simple defense mechanism, but in most instances, it is not sufficient by itself, and even less so against increasingly subtle and dynamic web defacement techniques.

Anomaly Detection

Anomaly detection comprises strictly monitoring the behavior of the website over time and identifying deviations from usual actions. This methodology is more flexible and efficient as compared to signature-based detection and recognizes new attacking techniques that do not fit established patterns.

Anomaly-based detection methods aim to identify web defacement attacks by recognizing deviations from the normal behavior

or appearance of a website, rather than relying on predefined signatures. These techniques typically involve creating a baseline model of legitimate website content, structure, or traffic patterns during a training phase. Once the normal state is established, the system continuously monitors for anomalies—such as unexpected changes in text content, HTML tags, layout structure, or user interaction patterns. For example, a sudden insertion of unfamiliar JavaScript code, a modified title tag, or an altered DOM element hierarchy could trigger an anomaly alert. Anomaly-based detection is particularly useful for identifying zero-day attacks and subtle content modifications that signature-based systems often miss.

Despite their advantages, anomaly-based detection methods also come with several challenges. One major issue is the high rate of false positives, where legitimate website updates—such as routine content changes or layout redesigns—are mistakenly flagged as defacement attempts. This requires fine-tuning the sensitivity of the detection system and, in some cases, incorporating human validation to avoid unnecessary alerts. Additionally, anomaly detection systems can be resource-intensive, especially when complex machine learning or behavioral models are used. Maintaining an up-to-date baseline in dynamic websites is also a challenge, as frequent content changes may require continuous retraining of the detection model. Nevertheless, with proper calibration and adaptive learning mechanisms, anomaly-based detection remains a powerful approach for identifying both overt and covert web defacement attacks in real time [5-6].

Machine Learning and AI-based Detection

Machine learning (ML) and artificial intelligence (AI) techniques have garnered attention in recent years as methods for enhancing web defacement detection. Machine learning-based detection methods for web defacement represent a modern and intelligent approach to identifying unauthorized changes to website content. Unlike traditional signature or anomaly-based techniques that rely on static rules or handcrafted baselines, machine learning models are trained on large datasets of normal and defaced web content to learn patterns and features that distinguish legitimate behavior from malicious alterations. These models can analyze a variety of indicators—such as text features, HTML tag sequences, visual layout characteristics, and metadata—to detect both obvious and subtle forms of defacement. Supervised learning techniques, such as support vector machines (SVM), decision trees, or neural networks, are commonly employed to classify web pages as normal or defaced based on labeled training data. Some advanced systems also use unsupervised learning to identify novel or previously unseen defacement attempts by clustering similar patterns and detecting outliers [7-8].

Hybrid Approaches

To provide a more robust solution in the detection of web defacements, hybrid detection techniques combine manifold methods like anomaly-based detection, signature-based detection, and machine learning techniques the combination of anomaly based and signature-based detection. Hybrid approaches for web defacement detection combine the strengths of multiple detection techniques—such as signature-based, anomaly-based, and machine learning-based methods—to achieve more accurate, robust, and adaptive detection capabilities. The rationale behind hybrid systems is that no single technique can effectively detect all types of defacement attacks, especially when considering the spectrum from overt homepage takeovers to minor, stealthy content manipulations. In a typical hybrid model, signature-based detection is used for identifying known attacks with high precision, while anomaly detection or machine learning models monitor for subtle or unknown changes. This layered architecture ensures both low false-positive rates for known threats and improved sensitivity to novel or evolving attacks. Some hybrid systems also incorporate DOM structure analysis, visual similarity comparison, and behavioral analysis to enhance detection accuracy [9-10].

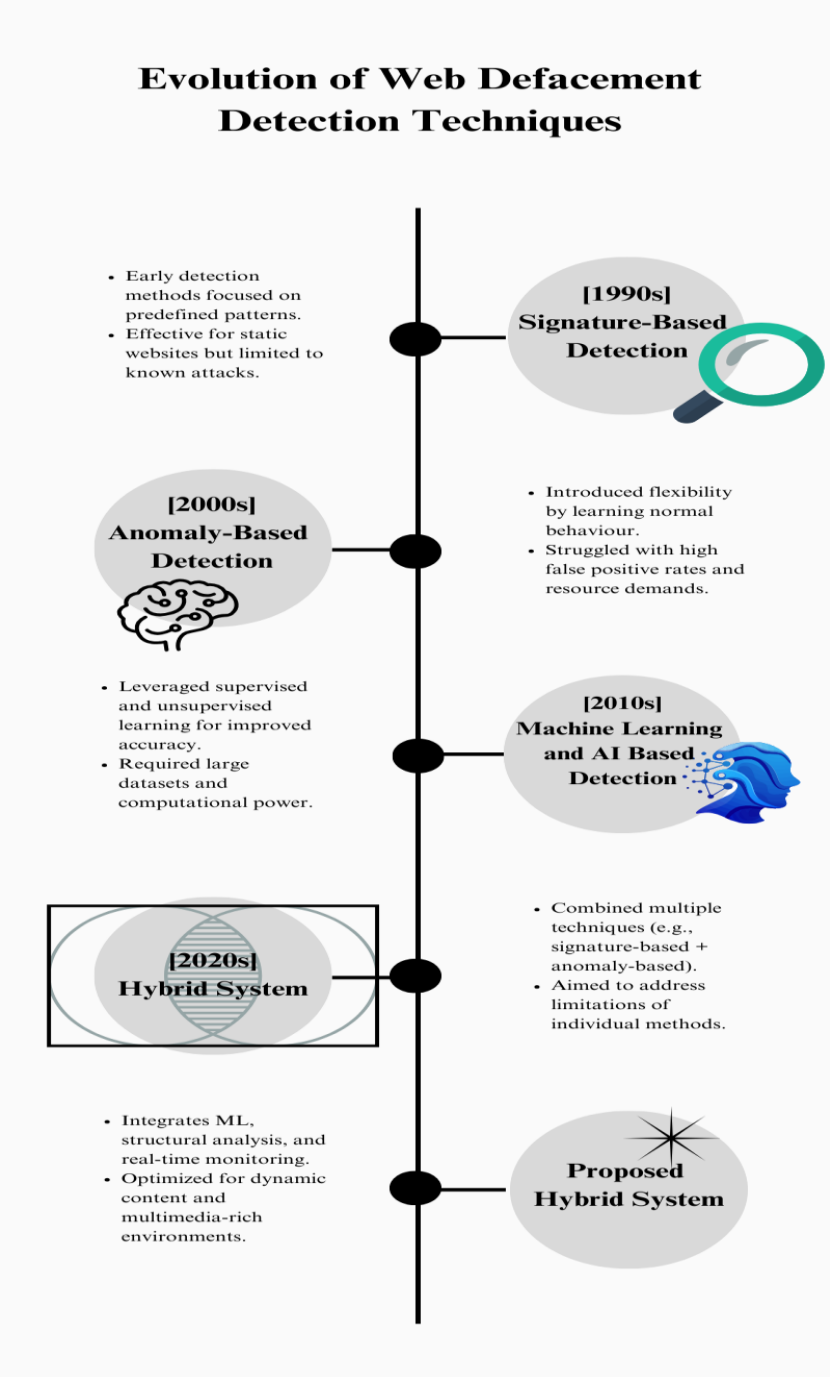

Figure 1: Detection Approaches for Web Defacement Attacks

RESEARCH METHODOLOGY

Figure 2 below shows the detailed and comprehensive workflow of the proposed system, which shows major steps from developing the proposed system to deploying it.

- Data Collection: Through web pages, the data is collected and then preprocessed.
- Feature Extraction: Extraction of relevant features for the analysis.
- Algorithm Design: Detection of developed criteria and comparison techniques.

- Implementation: Integrate and code the algorithm into a security framework.
- Evaluation: To test and validate the system using real-world data.
- Deployment: For real-time monitoring system is deployed.

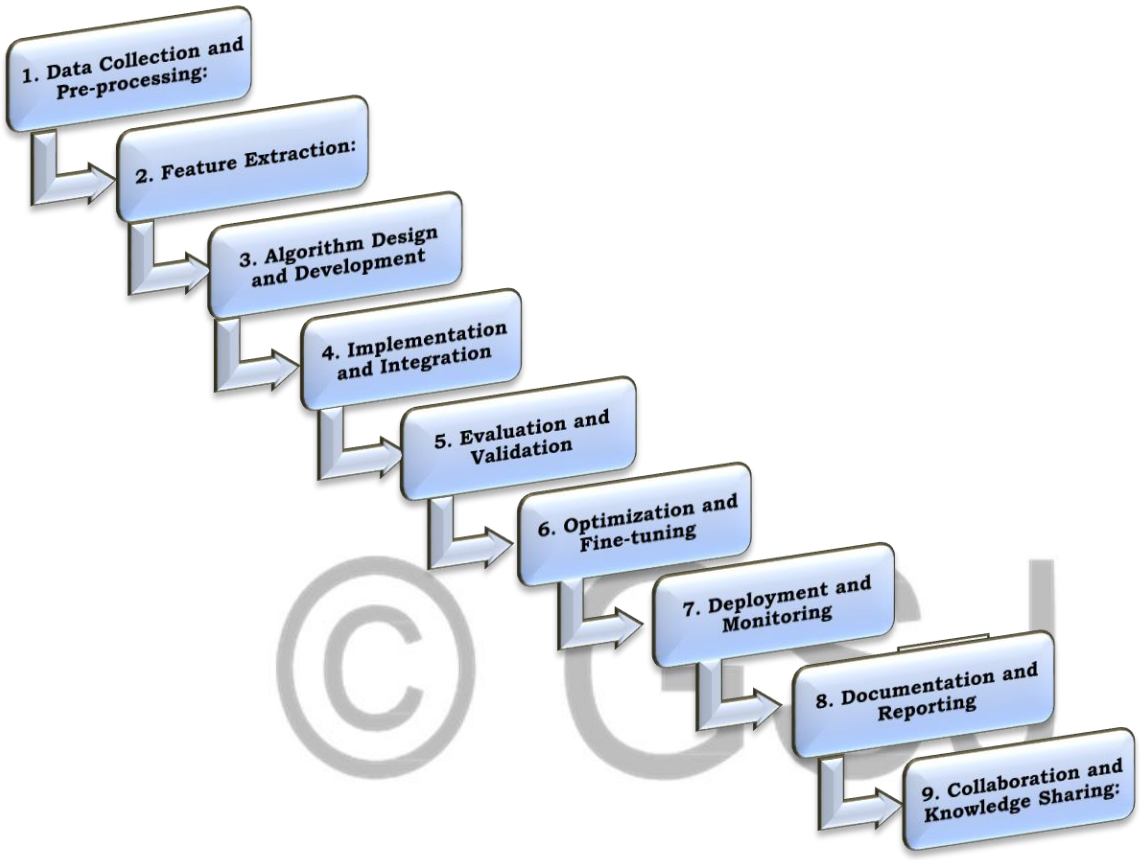

Figure 2 Proposed Workflow of the Study

Results

Figure 3 shows the result of our proposed tool compared to the existing tools.

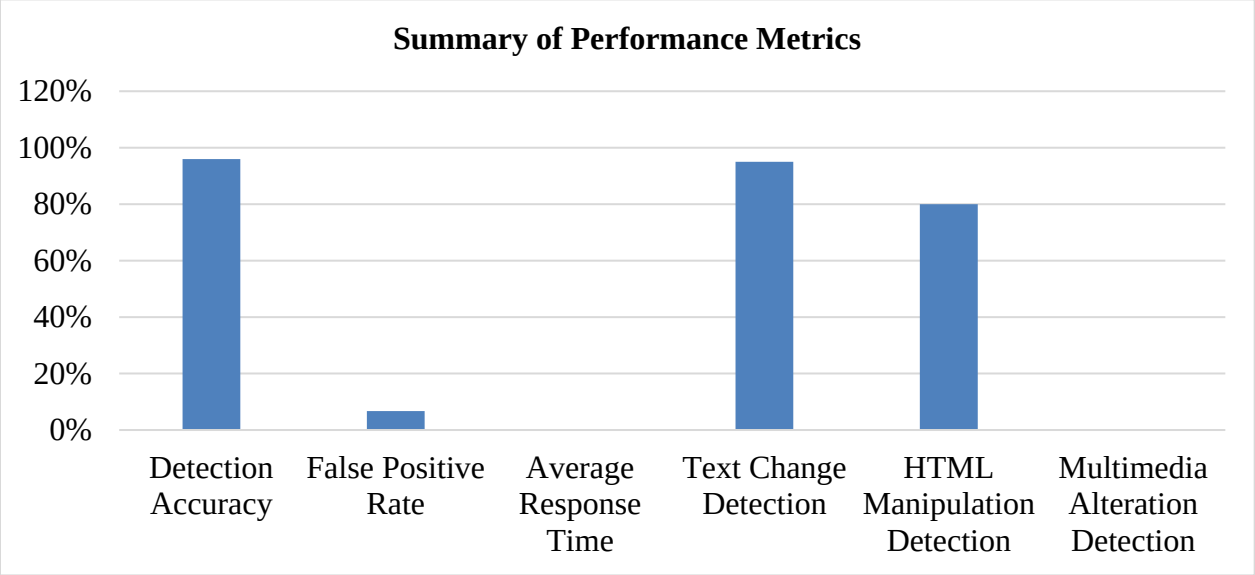

Figure 3: Result of proposed tool compared to existing approaches

Conclusion

The growth in web technologies can be observed by bringing exceptional opportunities for organizations to connect with their customers. However, with these opportunities and progress, a significant number of security challenges have also arisen, and one of the most common is defacement attacks on websites. These defacement attacks on websites not only undermine the trust of the visitors but also compromise the integrity of the website, which further disrupts the business operations and causes reputational damage.

The proposed tools for web defacement attacks detection demonstrated promising results as can be seen in the previous section. The existing techniques lacks several features and generates false positives while our tool performed much better in these aspects.

References

[1] Albalawi, M., Aloufi, R., Alamrani, N., Albalawi, N., Aljaedi, A., & Alharbi, A. R. (2022). Website defacement detection and monitoring methods: A review. *Electronics, 11*(21), 3573. https://doi.org/10.3390/electronics11213573

[2] Okereafor, K. (2008). Impacts of cyber attacks on corporate business continuity: Fostering cyber security consciousness in the citizenry. In *The 1st National Conference on Cybercrime and Cybersecurity*.

[3] Masango, M., Mouton, F., Antony, P., & Mangoale, B. (2017, September). Web defacement and intrusion monitoring tool: Wdimt. In *2017 International Conference on Cyberworlds (CW)* (pp. 72–79). IEEE.

[4] Hoang, X. D., & Nguyen, N. T. (2019). Detecting website defacements based on machine learning techniques and attack signatures. *Computers, 8*(2), 35. https://doi.org/10.3390/computers8020035

[5] Odeh, A., Keshta, I., & Abdelfattah, E. (2021, January). Machine learning techniques for detection of website phishing: A review for promises and challenges. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0813–0818). IEEE.

[6] Romagna, M., & Van den Hout, N. J. (2017, October). Hacktivism and website defacement: Motivations, capabilities and potential threats. In *27th Virus Bulletin International Conference* (Vol. 1, pp. 1–10).

[7] Bansal, S. (n.d.). The anatomy of web attacks: Understanding XSS, SQLi, and other threats. Insights2Techinfo. https://insights2techinfo.com/the-anatomy-of-web-attacks-understanding-xss-sqli-and-other-threats/, Accessed on: 2024-02-04

[8] Almomani, O., Alsaaidah, A., Abu-Shareha, A. A., Alzaqebah, A., Almaiah, M. A., & Shambour, Q. (2025). Enhance URL defacement attack detection using particle swarm optimization and machine learning. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE52024668ojs.bonviewpress.com

[9] Kurniawan, C., & Triayudi, A. (2023). File integrity monitoring as a method for detecting and preventing web defacement attacks. *Jurnal Online Informatika, 9*(2). https://doi.org/10.15575/join.v9i2.1326Jurnal Online Informatika

[10] Hoang, X. D. (2019). A website defacement detection method based on machine learning. In H. Fujita, D. C. Nguyen, N. P. Vu, T. L. Banh, & H. H. Puta (Eds.), *Advances in Engineering Research and Application* (pp. 116–124). Springer. https://doi.org/10.1007/978-3-030-04792-4_17SpringerLink