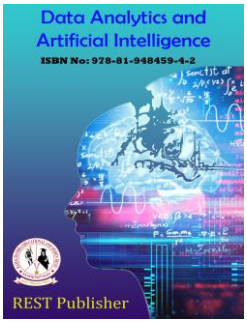
Data Analytics and Artificial Intelligence

Vol: 3(3), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/



Website Change and Detection Monitoring: "Get Aware of Phishes and Viruses"

Moratanch N, *Adnan Faseeh M S, Aafaq Ahmed Namazi, Jacob Raj R

Adhiyamaan College of Engineering, Hosur, Tamil Nadu, India.

*Corresponding Author Email: adnanfaseeh118@gmail.com

Abstract: *Web attacks and web defacement attacks are issues in the web security world. Recently, website defacement attacks have become the main security threats for many organizations and governments that provide web-based services. Website defacement attacks can cause huge financial and data losses that badly affect the users and website owners and can lead to political and economic problems. Several detection techniques and tools are used to detect and monitor website defacement attacks. However, some of the techniques can work on static web pages, dynamic web pages, or both, but need to focus on false alarms. Many techniques can detect web defacement. Some are based on available online tools and some on comparing and classification techniques; the evaluation criteria are based on detection accuracies with 100% standards and false alarms that cannot reach 1.5% (and never 2%); this paper presents a literature review of the previous works related to website defacement, comparing the works based on the accuracy results, the techniques used, as well as the most efficient techniques.*

Keywords: *website defacement, machine learning, web security, phishing, virus, malware.*

1. INTRODUCTION

A website defacement attack exploits a vulnerable website or a web server to launch malicious code to deface, modify, or delete the web page content (e.g., for personal or political reasons, or just for fun) via a text, picture, or both, which describe what the attacker wants to show (or simply block the web page). This can lead to financial and reputation consequences; see Figure 1. Some criminals use brute-force techniques; they might vulnerable points inside websites and perform techniques such as SQL injection or cross-site scripting (XSS), and send malware to website administrators. This requires defensive measures to be taken, such as frequent updating of the system, using monitoring and detection tools, and creating employee awareness and education about this type of attack. Detection and monitoring systems are important because they may allocate and prevent the attack from occurring again. Then the weakness or flaw in the system or website is out and can be fixed. As a consequence of the lack of detection or prevention systems, the website is vulnerable to defacement attacks. In this paper, we present several techniques to monitor a website and measurements to detect website defacement, regardless of whether the web page is static, dynamic, or both. Website defacement may be associated with website users or owners. Figure 2 shows the possible reasons for website defacement. Web security involves protecting the website or web application from unauthorized users, especially those who want to sabotage devices and networks via vulnerabilities, to damage, steal, and disrupt data. Web security is important because government, economic, commercial, and educational websites are uploaded on the web. Defacing websites is a threat to web security and a nuisance. Attackers may do this for fun or political reasons. Web security confidentiality, integrity, and availability can be considered protection measures for static and dynamic web pages. However, dynamic sites are less secure and more exposed to the internet than static sites. Because of the presence of interactive communication between the site and the database, a database can be exploited. Figure 3 presents an overview of what a static and a dynamic website look like. A static web page is a prepared HTML file fully formed and located on the server storage waiting to be requested by a client and displayed. In this case, the client is a web browser. The browser generates a URL to address a particular HTML file. The URL is sent to the server to request the page from the server's database.

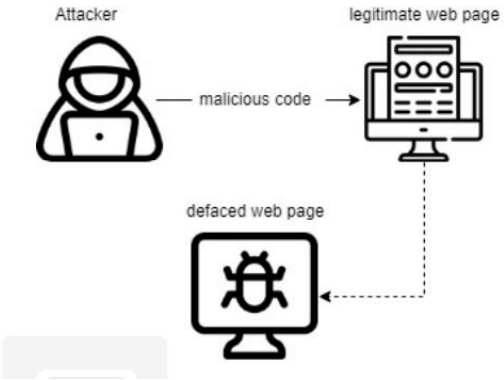


FIGURE 1. Overview of website defacement

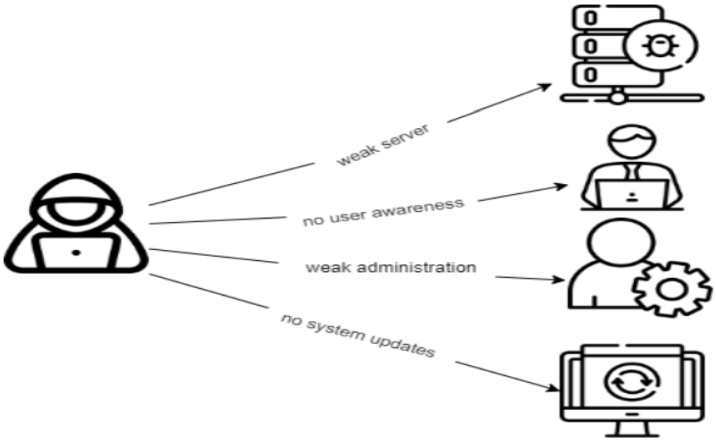


FIGURE 2. Defacement causes

2. RELATED WORK

Webpage change detection based on image processing and image comparison change detection and notification systems are poor choice for less frequently changing websites. [1] We propose alternative approach based on Image Processing and Image Comparison. The proposed system includes functionalities for fetching webpage from the Internet. using Image Comparison Algorithm, two different versions of webpage are compared. This paper presents a literature review of the techniques and methods used to identify website defacements. However, some of the techniques can work on static web pages, dynamic web pages, or both, but need to focus on false alarms.[2] Many techniques can detect web defacement. Some are based on available online tools and some on comparing and classification techniques; the evaluation criteria are based on detection accuracies with 100% standards. Change Detector, an enforced prototype, addresses this challenge by incorporating a number of machine language ways. [3] The top backend factors of Change Detector are machine learning intelligent crawling, runner bracket and reality- grounded change detection Finally, the frontal end presents a flexible way to interact with the database of detected changes to pinpoint those changes. Analysis of phishing sites using Web technology. [4] A phishing attack is one of the biggest security concerns in URL and domain identifier dot, abnormal attributes and domain attributes were used as semantic features to deploy phishing sites. Machine readable modelling algorithms are used for detection. Versioning and change detection on the web paper looks at versioning and change detection on the web, proposes a system for efficiently monitoring changes to Web documents.[5] researchers say the algorithm can be used to detect changes in Web pages. web-based system for detecting change in ontology on the Web. Vandalism in web - grounded services website vandalization attacks are the main security pitfalls for web-grounded services. [6] Websites can be attacked with false admonitions and fiscal losses. There are several ways to descry and cover website. vandals. To eradicate this detection tool is been used. Pre-Phish algorithm is an automated machine learning approach to analyze phishing and non-phishing URL to produce reliable result. [7] Phishing URLs typically have connections between registered domain level and query level URL. Pre-Phish is a real-time phishing

website detection method. This thesis aims to assess the felicity of web technologies for the development of administrative control and data accession systems. The Industrial Internet is used as the Industrial Internet platform for OPC UA standard for artificial information modeling and data transfer. [8] A central theme is the integration of software factors in the Java and JavaScript languages. Vandalization of web spots has come a wide problem. A paper assesses the performance of several approaches to descry web vandalism automatically. [9] The approaches construct a profile of the covered runner and use machine literacy to descry vandalism. The authors say the approaches perform well in terms of false cons and negatives. Webvigil: a change monitoring system for the web Web Vigil is a change monitoring system to detect changes to the data on the web and notify runners in a timely manner. [10] We present the semantics of the system and its use of event-condition-action rules and a literacy algorithm for costing runners. The change detection on the web paper looks at versioning and change detection on the web, proposes a system for efficiently monitoring changes to Web documents. researchers say the algorithm can be used to detect changes in Web pages. [11] web-based system for detecting change in ontology on the Web. Change Detector are machine learning intelligent crawling, runner bracket and reality- grounded change detection. [12] Finally, the frontal end presents a flexible way to interact with the database of detected changes to pinpoint those changes. Web-based analysis of phishing websites. [13] One of the largest security problems in URL and domain identifier dot is a phishing attack. Phishing sites were deployed using anomalous characteristics and domain attributes as semantic elements. Algorithms for machine-readable modelling are utilized for detection. A system for effectively tracking changes to Web publications is suggested in the paper on change detection on the web. It examines versioning and change detection on the web. The technique, according to experts, can be used to find changes in Web pages. [14] Web-based method for monitoring ontology changes on the Internet. The purpose of this thesis is to evaluate the suitability of web technologies for the creation of systems for administrative control and data access. [15] The OPC UA standard for artificial information modelling and data transfer is implemented on the Industrial Internet as the Industrial Internet platform. The integration of software components into the Java and JavaScript languages is a major theme.

3. PROPOSED SYSTEM

According to our study, only a limited number of surveys have been carried out regarding webpage CDN techniques. Additionally, it is challenging to find comprehensive evaluations of existing CDN systems which discuss different aspects of such systems. Oita et al. [84] have reviewed major approaches used for the change detection process in webpages. They have reviewed temporal aspects of webpages, types of webpage changes, models used to represent webpages to facilitate change detection and various similarity metrics that are used for detecting changes. Shobhna and Chaudhary [104] discuss about a selected set of CDN systems with different types of change detection algorithms in a summarized manner.

3.1 System Architecture: There has been a wide variety of techniques for detecting and preventing defaced websites; machine-learning detection techniques are great, but sometimes they are resource-consuming and slow because they need to be Therefore, some tools have been found to detect and prevent defaced websites, fed with set of data and apply more than one in FIG.1 to give accurate results.

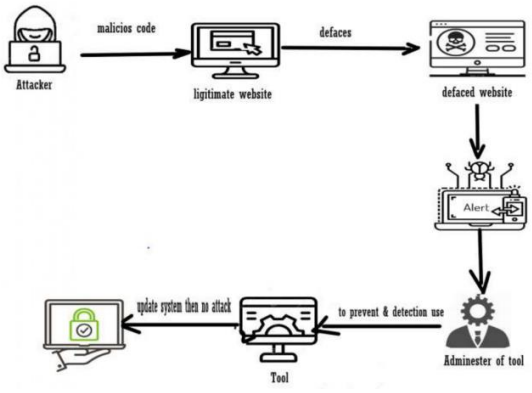


FIGURE 3.

3.2 Modules: In the module 1, changes detected will be notified within 24 hours of time as depicted in the figure 1, so that we can easily identify the changes that are been done.

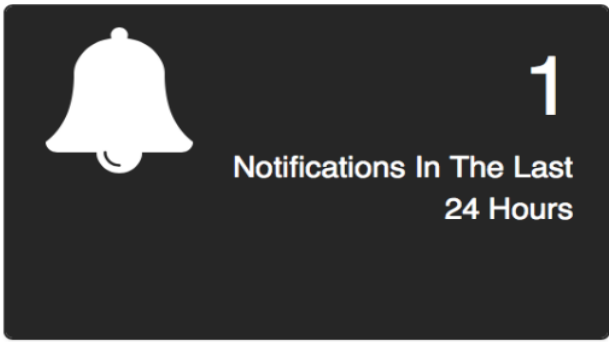


FIGURE 4. Module 1

In the module 2, the monitoring will be active for every single minute so that it could identify is trying to crash our website as depicted in the figure 2.

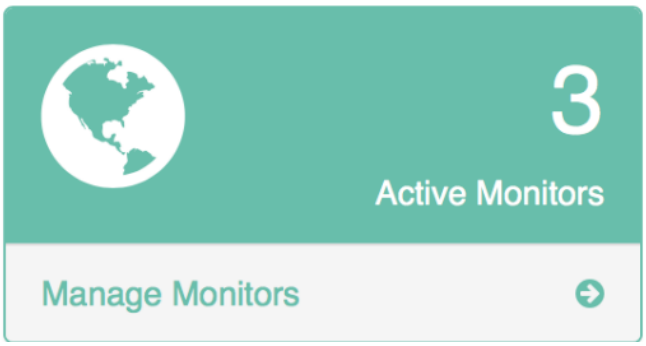


FIGURE 5. Module 2

4. RESULT

We intend to look into certain real-world applications in a variety of tools to examine the web pages, and then we'll compare these tools to see which one is best (and fastest) at looking for vulnerabilities in websites. In order to increase detection precision and decrease false positive alarms, a hybrid approach combining two distinct approaches, such as machine learning and another fundamental algorithm for detecting defacement, will be taken into consideration.

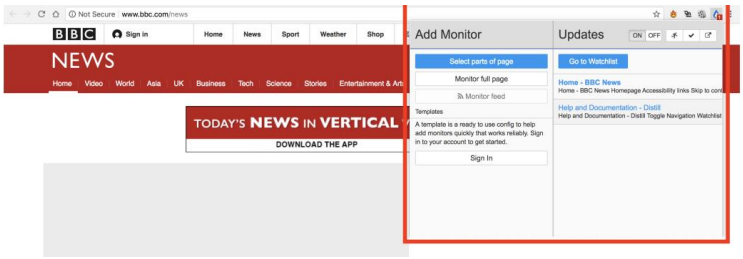


FIGURE 6.

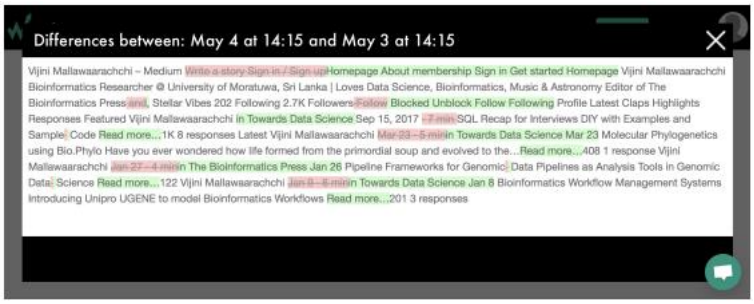


FIGURE 7.

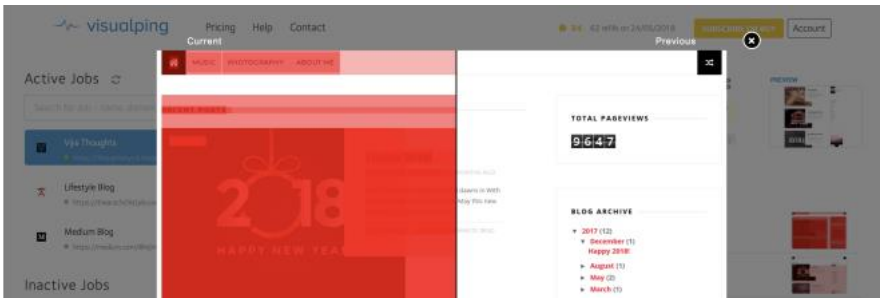


FIGURE 8.

5. CONCLUSION

There is a need for a beneficial approach to secure websites on the internet. The web will become more understandable as we study how to construct security mechanisms. The growth of web-based services will force us to become more careful and build highly secure web infrastructure, depending on a secure web server. One of the most critical attacks involve the internet is the defacement of websites. We reviewed website defacement detection techniques and defacement detection tools and specified each study's implementation, machine-based learning techniques, and other tools used in this field. Defacement detection techniques may be categorized into three categories: anomaly-based detection, signature-based detection, and machine-learning techniques. The study may be repeated using a new method of website defacement detection and monitoring. In the future, we will apply some of these techniques against web defacement attacks. We plan to investigate certain practical implementations in multiple tools to examine the web pages, and then compare these tools to determine which one is the best (and the fastest) to examine website vulnerabilities. Moreover, a hybrid between two different methods, such as machine learning and another basic algorithm for detecting defacement, to improve detection accuracy and reduce false positive alarms, will be considered.

REFERENCES

[1]. E. Adar, J. Teevan, S. T. Dumais, and J. L. Elsas. 2009. The Web Changes Everything: Understanding the Dynamics of Web Content. In Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM '09). ACM, New York, NY, USA, 282–291. https://doi.org/10.1145/1498759.1498837

[2]. F. Ahmadi-Abkenari and A. Selamat. 2012. An architecture for a focused trend parallel Web crawler with the application of clickstream analysis. Information Sciences 184, 1 (2012), 266–281. https://doi.org/10.1016/j.ins.2011.08.022

[3]. A. Anjum and A. Anjum. 2012. Aiding web crawlers; projecting web page last modification. In 2012 15th International Multitopic Conference (INMIC). 245–252. https://doi.org/10.1109/INMIC.2012.6511443

[4]. R. Baeza-Yates, C. Castillo, and F. Saint-Jean. 2004. Web Dynamics, Structure, and Page Quality. Springer Berlin Heidelberg, Berlin, Heidelberg, 93–109. https://doi.org/10.1007/978-3-662-10874-1_5

[5]. K. Benjamin, G. von Bochmann, M. E. Dincturk, G.-V. Jourdan, and I. V. Onut. 2011. A Strategy for Efficient Crawling of Rich Internet Applications. In Web Engineering, S. Auer, O. Díaz, and G. A. Papadopoulos (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 74–89.

[6]. D. Bhatt, D. A. Vyas, and S. Pandya. 2015. Focused Web Crawler. Advances in Computer Science and Information Technology (ACSIT) 2, 11 (April 2015), 1–6.

[7]. B. H. Bloom. 1970. Space/Time Trade-offs in Hash Coding with Allowable Errors. Commun. ACM 13, 7 (July 1970), 422–426. https://doi.org/10.1145/362686.362692

[8]. P. Boldi, B. Codenotti, M. Santini, and S. Vigna. 2004. UbiCrawler: a scalable fully distributed Web crawler. Software: Practice and Experience 34, 8 (2004), 711–726. https://doi.org/10.1002/spe.587 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/spe.587

[9]. P. Boldi, A. Marino, M. Santini, and S. Vigna. 2018. BUbiNG: Massive Crawling for the Masses. ACM Trans. Web 12, 2, Article 12 (June 2018), 26 pages. https://doi.org/10.1145/3160017

[10]. K. Borgolte, C. Kruegel, and G. Vigna. 2014. Relevant Change Detection: A Framework for the Precise Extraction of Modified and Novel Web-based Content As a Filtering Technique for Analysis Engines. In Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion). ACM, New York, NY, USA, 595–598. https://doi.org/10.1145/2567948.2578039

[11]. O. Brandman, J. Cho, H. Garcia-Molina, and N. Shivakumar. 2000. Crawler-Friendly Web Servers. SIGMETRICS Perform. Eval. Rev. 28, 2 (Sept. 2000), 9–14. https://doi.org/10.1145/362883.362894

[12]. B. E. Brewington and G. Cybenko. 2000. How dynamic is the Web?1This research was partially supported by AFOSR grant F49620-97-1-0382, DARPA grant F30602-98-2-0107 and NSF grant CCR-9813744. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the above agencies.1. Computer Networks 33, 1 (2000), 257–276. https://doi.org/10.1016/S1389-1286(00)00045-1

[13]. S. Brin and L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 30, 1 (1998), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X Proceedings of the Seventh International World Wide Web Conference.

[14]. D. Buytaert. 2000. Drupal - Open Source CMS | Drupal.org. Drupal community. Retrieved November 8, 2019 from https://www.drupal.org/

[15]. G. C. Canavos. 1972. A Bayesian Approach to Parameter and Reliability Estimation in the Poisson Distribution. IEEE Transactions on Reliability R-21, 1 (Feb 1972), 52–56. https://doi.org/10.1109/TR.1972.5216172