SHIV SHAKTI

International Journal of in Multidisciplinary and Academic Research (SSIJMAR)

Vol. 4, No. 3, June 2015 (ISSN 2278 – 5973)

AN APPROACH TO REVEAL WEBSITE DEFACEMENT

Garima Singh

M.Tech Student

Dept. of Computer Sci.& Engineering

Somany institute of Tech. & Management, Rewari

Haryana, India

garima1989singh@gmail.com

Dr. Pushpender Sarao

Professor

Dept. of Computer Sci. & Engineering

Somany inst. Of Tech. & Mgmt, Rewari

Haryana, India

pushpendersarao@gmail.com

Impact Factor = 3.133 (Scientific Journal Impact Factor Value for 2012 by Inno Space Scientific Journal Impact Factor)

Global Impact Factor (2013)= 0.326 (By GIF)

Indexing:

ABSTRACT

Due to adhoc nature of web application development and design complexity of web application, it is difficult to attain fool proof web security. In recent years invaders defaced several web sites by projecting techniques such as phishing, code injection etc. In the web defacement attack the invader changes the visual appearance of the webpage. The business competitor, Insurgent and extremist groups defame the reputation of the organizations and mislead public through these types of attacks. Manual monitoring and scrutinizing these attacks on web sites is a time consuming and tedious task for law enforcement agencies. Hence there is a need to develop a system which effectively monitors the content of web sites and automatically generate alarm for any suspicious or threatening activity.

In this work a prototype system is developed to scrutinize and detects the defacement activities automatically. At first phase web contents are preprocessed and stored in the web domain dictionary. In second phase integrity of web contents through CRC32, MD5, SHA 512, PSNR and SSIM techniques is checked. The developed system successfully scrutinizes the web defacement attacks and it would be helpful for web administrator to capture the web defacement cases automatically.

1. INTORDUCTION

In today's world internet has become an indispensable part of one's everyday life. Most of the routine transactions are available online, either it be information regarding a subject or other service like reservation, online shopping also known as e-shopping. Websites serve as the source of information. They also contain proprietary data which can be misused. It is the need of the hour to make websites and its services secure

"Web site defacement is the process of introducing unauthorized modifications to a Website."

A brief description of information security, cyber security, web security, internet security, Attacks on Websites & its Countermeasures and Website Defacement is explored in the following sections:

1.1 Information Security :- Information security (InfoSec), as defined by the standards published by the Committee on National Security Systems (CNSS), formerly the National Security Telecommunications and Information Systems Security Committee (NSTISSC), "the protection of information and its critical elements, including the systems and hardware that use, store, and transmit that information".

1.2 Cyber Security: - It is defines as to protect a computer or computer system against unauthorized access or attack on the internet. It is a collection of tools, policies, security concepts and safeguards, guidelines, risk management approaches, assurance and technologies that can be used to protect the organization and user's assets which includes connected computing devices, personnel, infrastructure,

applications, services, telecommunications systems, and the totality of transmitted and/or stored information in the cyber environment. The general security objectives comprise availability, integrity, confidentiality, authenticity and non-repudiation.

1.3 **Web Security: -** Web security involves protection of the confidential data on the web by preventing, detecting and responding to the attacks on websites. The prevention is done by authenticating the users who access the web. This can be done in two ways – basic access and digest access. In basic access authentication the user requires a username and password to make a request. This method is implemented on HTTPS protocol. This does not provide any confidentiality. This is the simplest method to provide access control as it does not require any cookies rather it uses static HTTP headers. HTTP headers are components of the message header of requests and responses in the Hypertext Transfer Protocol (HTTP). They define the operating parameters of an HTTP transaction.

1.4 **Internet Security: -** Internet security is to provide security to internet from hacker or intruder. The internet security often involves browser security but also network security on a more general level as it applies to other applications or operating systems on a whole. Its objective is to set up the rules against attacks over the Internet. Internet security problems can be divided roughly into four closely intertwined areas: secrecy, authentication, non-repudiation, and integrity control. Secrecy, also called confidentiality which means keeping information out of the hands of unauthorized users.

1.5 **Attacks on Websites and its Counter Measures:** - Rapid growth of internet has created numerous services, which have become integral part of one's day to day life. Web applications are used for making reservations, paying bills, and shopping on-line. However wide spread usage of this applications came with increase in number and type of attacks affecting confidentiality, integrity and availability of information. Almost every day, new security vulnerabilities are discovered which are exploited by hackers in accessing confidential information.

1.6 Website Defacement: - Web site defacement is the process of introducing unauthorized modifications to a Web site which changes the visual appearance of the site or a webpage [8]. It is an attack on a website that changes the visual appearance of the site or a webpage. In this type of attack, the attacker defaces the reputation of an organization by modifying the content of home page. Targets are mainly government, official agencies and trade groups' websites. Affects of website defacement are *disturbed images, political messages, forms of signature of the attacker*.

1.7 Motivation: - Internet is the basic requirement in the present era. As it is an open source, it is more vulnerable to serious defacement. Even it is a secure website; hackers have developed advanced ways to attack on the contents of web pages.

The law enforcement agencies are trying their best to catch such hackers. Hence there is a need to develop a system which effectively monitors the content of web sites and automatically generate an alarm to reveal even a minor suspicious or threatening activity.

This served as a motivation towards designing of a framework, which effectively and efficiently provides a method to prevent the website defacement. A framework is proposed in this thesis which provides a solution to the defacement of text as well as images of a webpage.

2. Background and literature Survey

2.1 History of Website

Website came in contemplation following the key idea of hypertext. First it is explained in Jorge Luis Borges short story 'The Garden of Forking Paths', published in 1941. The origin of the idea of HyperText is considered in 1945 when the Atlantic Monthly publishes the seminal 'As We May Think' by Vannevar Bush, an American engineer. In 1963 Ted Nelson, father of the Xanadu hypermedia system, coins the term HyperText.

The term 'Markup Language' is coined in 1970 by Charles Goldfarb, co-inventor of the first markup language GML, and designer of SGML. He invented this Generalised Markup Language (GML) at IBM with his colleagues Ed Mosher and Ray Lorie.

The first website using HTML tags is designed by Tim Berner-Lee in 1991. HTML tags are a list of tags used in the HTML language. Each tag starts with a tag opener (a less than sign) and ends with a tag closer (a greater than sign).

2.2 Web Security

Simson Garfinkel et al. [11] gave a definition of web security, using that definition, web security is a set of procedures, practices, and technologies for protecting web servers, web users, and their surrounding organizations. Security protects against unexpected behavior. Website users and owners have a set of postulations regarding expected security problems of web.

A Web Security Solution based WALSG (Web Application Level Security Gateway) on XML Technology has been given in ref. [14]. Their solution WALSG provides administrators with means to describe specific security settings for specific web pages and as attacker always search for web pages with weaker security defacement could be happened and our system work for all pages of website.

2.3 Webpage Pre-Processing

The researcher from China Weitong Huang et al. [55] introduced the details of a Chinese Web-page Classification system. Then they proposed the Preprocessing and Feature Preparation in Chinese Web Page Classification [56]. In that proposed system there are six procedures in web-page preprocessing: HTML parsing, English lexical analysis, Chinese word segmentation, stopword removal, stemming, and vocabulary selection.

The web page extraction is a subpart of pre-processing of web page. Alberto H. F. Laender et al. surveyed a brief of Web Data Extraction Tools [57]. The taxonomy is based on the technique which is used by each tool. Here, technique was to generate a wrapper.

To generate a wrapper following existing tools are surveyed: Languages for Wrapper Development, Wrapper Introduction Tools, HTML-aware Tools, Natural Language Processing (NLP) based Tool, Modeling-based Tools, and Ontology-based Tools. Then Quality Analysis was explored in that paper. The Quality Analysis involves Degree of Automation, Support for Objects with Complex Structure, Page Contents, Availability of GUI, XML output, Support for Non-HTML Sources, Resilience and Addictiveness.

2.4 Text Integrity

It is the process to detect the alteration of text. It is done by comparing text character-by-character. It requires at least one file contains unaltered or original text. Yasushi Yamazaki et al. [59] proposed a model named as Hidden Markov Models for Text-indicated Writer Verification. In his proposed model a different text including ordinary characters is used as a dataset which is used for verification of text.

Researcher Tatjana Rusko focuses on aspects of scientific text organization that adequately account for its integrity [60]. He examined Cohesion text-forming role, its functioning, actualization, formal-logical and semantic means. The main function of text cohesion was to provide a connection between the text elements and introducing a hierarchy of its constituent parts for the integration of a text. In that cohesion actualization was to transmit certain information from the addresser to the addressee and cohesion realization in the text were lexical, syntactic, graphic and logical-semantic ones depending on the author's intentions.

2.5 Image Integrity

It is the process to detect alteration in pixels of image. In this images are compared pixel by pixel. Two images are called same when each and every pixel has same value.

Jozsef Lenti et al. proposed a feature vector generation algorithm where the Feature vector generation for image integrity verification [64]. The proposed algorithm distinguishes JPEG compression from other malicious manipulations. They conclude that the feature vector based authentication can detect the image modifications and it is possible to trace image blocks which are modified compared to the original blocks.

2.6 Website Defacement

Major work done in the field of website defacement is done by Alberto Bartoli. In Ref. [49] A Framework for Large-Scale Detection of Web Site Defacements is proposed which is based on Anamoli detection technique. During a preliminary learning phase Goldrake automatically builds a profile of the monitored resource. Then, while monitoring, Goldrake will retrieve the remote resource periodically and generate an alert whenever something "unusual" is shown. Implementing this simple idea is hard because Web

content is highly dynamic; that is, different readings of the same Web resource may be very dissimilar from each other.

The system proposed by them is illustrated as follows: The resources R which are analyzed by a number of sensors S. A sensor Scan operates in either of two modes named as learning mode and monitoring mode. In learning mode, S takes the tuning sequence and executes its own tuning procedure. As a result of this procedure, whose details are sensor-specific, S builds its own portion of the profile of R and verifies whether it is indeed able to operate with R. If not, S will state that it is unsuitable for R, in which case M will not apply S to R.

REFERENCES

[1] Arapaho Nsuok, "Introduction to Information Security", Available: http://arapaho.nsuok.edu/~hutchisd/IS_4853/ C6572_01.pdf.

[2] "44 USC § 3542 - Definitions", Available: http://www.law.cornell.edu/uscode/text/44/3542

[3] Gralla, and Preston, How the Internet Works, Indianapolis: Que Pub, 2007.

[4] Rhee, M. Y., Internet Security: Cryptographic Principles, Algorithms and Protocols, Chichester: Wiley, 2003.

[5] Andrew S Tanenbaum, Computer Networks, 4th ed., New Delhi: Prentice-Hall, 2002.

[6] Cyber Security, Available: http://www.itu.int/en/ITU-T/studygroups/com17/Pages/ cybersecurity.aspx.