ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

Vol. 2, Issue 5, May 2014

# A Survey on Attacks in Web Usage Mining

R.Natarajan, Dr.R.Sugumar

Research Scholar, Dept. of CSE, St. Peter's University, Chennai, India

Associate Professor, Dept. of CSE, Veltech Multitech Dr.RR Dr.SR Engineering College, Chennai, India

**ABSTRACT:** As the use of internet is high nowadays, threats are taking places on the use of internet. There are several kinds of online attacks which affect the system badly and infect the system in such a manner that the server finds itself to recover. In the same proceedings attacks like brute force or IP spoofing has made its place in the crowd of online theft and attacks. Brute force attack is an attack in which the combination of password is thrown to the login system to crack the password. It has been seen often that the combination of password cracks the database. To prevent the system for such attacks optimization algorithms are designed to ensure that the system remains safe. In addition to the brute force attack there is another attack named IP spoofing attack which leads to send request to the server in a random manner, before the server deals with one ip another request hits the system and makes the server response slow. Such a system can be often seen in the university sites where results day become a hectic task for the server to respond to all the request at the same time and it leads to slowing down the server. This research deals with three kinds of attacks simultaneously namely brute force attack, ip spoofing and sql Injection attack with the prevention mechanisms like ant colony optimization and genetic algorithm simultaneously. The research would also compare the results of genetic algorithm and ant colony optimization algorithm.

**KEYWORDS:** KDD, Securiy, Passive attack

## I. INTRODUCTION

Web Mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents, us-age logs of web sites, etc Internet has became an indispensable part of our lives now a days so the techniques which are helpful in extracting data present on the web is an interesting area of research. These techniques help to extract knowledge from Web data, in which at least one of structure or usage (Web log) data is used in the mining process (with or without other types of Web). According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining.

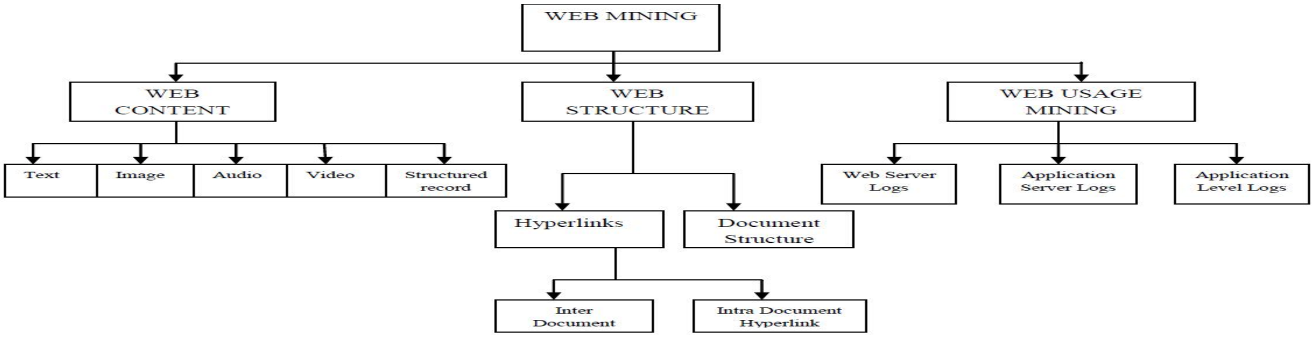

**Fig 1.2. Web Mining Taxonomy**

**a. Web content mining:** Web content mining is the mining, extraction and integration of useful data, information and knowledge from Web page content. It examines content of the web pages as well and web searching. Content data corresponds to the collection of facts a Web page was designed to convey to the users. Web content may be unstructured (plaintext), semi- structured (HTML documents), or structured (extracted from databases into dynamic Web pages).

**b. Web structure mining**: Web structure mining is the process of using graph theory to analyze the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:
1. Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location.
2. Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.

**c. Web usage mining**: Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Discovery of meaningful patterns from data generated by client-server transactions on one or more Web servers. Usage mining is valuable not only to businesses using online marketing, but also to e-businesses whose business is based solely on the traffic provided through search engines. The use of this type of web mining helps to gather the important information from customers visiting the site. This enables an in-depth log to complete analysis of a company's productivity flow. E-businesses depend on this information to direct the company to the most effective Web server for promotion of their product or service.

## II. LITERATURE SURVEY

The important task of web mining is web usage mining, which mines web log records to discover user access patterns of web pages. Since web log data provide information about what kind of user will access what kind of web pages, web log information can be integrated with web content mining and web linkage structure mining.

**Abdelhakim Herrouz, Chabane Khentout, Mahieddine Djoudi** describes web mining is a class of data mining. In order to relieve a "Data Rich but Information Poor" dilemma, Data Mining emerged. Web Mining is a variation of this field that distils untapped source of abundantly available free textual information. The importance of web mining is growing along with the massive volumes of data generated in web day-to-day life. In general, web data always arrives in a multiple, continuous, rapid and time varying flow. Most of the existing conventional algorithms fail while handling such dynamic data. Web data extraction algorithms are important in extracting useful documents from streaming on-line sources. We propose a new method for web data extraction. It has three phases. In the first phase list of web documents are selected, second phase documents are preprocessed, in the final phase results are presented to users.

**Daxin Jiang Jian Pei Hang Li, "Mining Search and Browse Logs for Web Search:** Huge amounts of search log data have been accumulated at web search engines. Currently, a popular web search engine may every day receive billions of queries and collect tera-bytes of records about user search behavior. Beside search log data, huge amounts of browse log data have also been collected through client-side browser plug-ins. Such massive amounts of search and browse log data provide great opportunities for mining the wisdom of crowds and improving web search. At the same time, designing effective and efficient methods to clean, process, and model log data also presents great challenges.

**Jianli Duan, Shuxia Liu:** depicts that with the rapid development of Internet, web data mining, especially weblog mining, plays an important role in many fields, including personalized information service, improving designs, services of websites and so on. This paper introduces web data mining firstly, and then discusses the process of weblog mining.

**Sivakumar J, Ravichandran K.S, :** This paper focuses on the efficient application of the Web Mining Algorithm for web log analysis which is applied to identify the context associated with the web design of an e commerce web portal that demands security. As priority is given to efficiency, the comparative study made with other similar algorithm like E-web Miner Algorithm and Apriori algorithm, it has been proved that this proposed Web Page Collection web mining algorithm as the best [or say the most suited] performer to manage time and space complexity .thus this algorithm, better known as Efficient web Miner possesses valid by computational comparative performance analysis. The number of data base scanning drastically gets reduced in Web Page Collection algorithm. Here it may be noted that E -Web Miner can be applied successfully in any weblog analysis which includes information centric network design.

.**Maryam Jafari, Shahram Jamali, Farzad Soleymani Sabzchi:** World Wide Web plays a significant role in human life. It requires a technological improvement to satisfy the user needs. Web log data is essential for improving the performance of the web. It contains large, heterogeneous and diverse data. Analyzing g the web log data is a tedious process for Web developers, Web designers, technologists and end users. In this work, a new weighted association mining algorithm is developed to identify the best association rules that are useful for web site restructuring and recommendation that reduces false visit and improve users' navigation behavior. The algorithm finds the frequent item set from a large uncertain database. Frequent scanning of database in each time is the problem with the existing algorithms which leads to complex output set and time consuming process. The proposed algorithm scans the database only once at the beginning of the process and the generated frequent item sets, which are stored into the database. The evaluation parameters such as support, confidence, lift and number of rules are considered to analyze the performance of proposed algorithm and traditional association mining algorithm. The new algorithm produced best result that helps the developer to restructure their website in a way to meet the requirements of the end user within short time span

**Roop Ranjan, Sameena Naaz, Neeraj Kaushik,** in current trend, most of the businesses are running through online web applications such as banking, shopping, and several other e-commerce applications. Hence, securing the web sites is becomes must do task in order to secure sensitive information of end users as well as organizations. Web log files are generated for each user whenever he/she navigates through such e-commerce websites, users every click is recorded into such web log files. The analysis of such web log files now a day's done using concepts of data mining. Further results of this data mining techniques are used in many applications. Most important use of such mining of web logs is in web intrusion detection. To improve the efficiency of intrusion detection on web, we must have efficient web mining technique which will process web log files. In this project, our first aim is to present the efficient web mining technique, in which we will present how various web log files in different format will combined together in one XML format to further mine and detect web attacks. And because log files usually contain noisy and ambiguous data this project will show how data will be preprocessed before applying mining process in order to detect attacks.

**Dilpreet kaur, Sukhpreet Kaur** Web usage mining is an important type of web mining which deals with log files for extracting the information about users how to use website. It is the process of finding out what users are looking for on internet. Some users are looking at only textual data, where others might be interested multimedia data. Web log file is a log file automatically created and manipulated by the web server. The lots of research has done in this field but this paper deals with user future request prediction using web log record or user information. The main aim of this paper is to provide an overview of past and current evaluation in user future request prediction using web usage mining

**Rimmy Chuchra, Bharti Mehta, Sumandeep Kaur:** This research paper is merging the concept of web mining with the network security so that we can easily detect the online attacks occur on the network by using web agents (i.e. - web agents are basically web robots) rather than using man power effort. The major objective is to reduce cost as well as time while identifying online attack. Here we use “rule induction data mining technique” to achieve maximum accuracy of results. The special focus is to detect online active attack by the web agents after that they will provide security by using various mechanisms and techniques. In this way, Paper can also say that these web agents help to protect us from attacker during online data transfer which follows the concept of network security. The first task of web agents is to identify the type of active attack after that provide several ways to prevent security. In this way we can use a “Hybrid approach” (i.e. - web mining with network security).The major benefit to use this hybrid approach is to save time and cost which are the major objectives of data mining.

**S.Mirdula, D.Manivannan:** Security is the essential and important topic in web applications. The choice of communication made the web technology a essential one in the environment. The importance of web application and its security increasing day by day, but traditional networks fails to provide security for web application. This paper discuss about some of the vulnerable online attacks commonly occurs in web applications and providing solution for preventing such attacks by using penetration tool backtrack. The testing aspect of vulnerabilities is carried out for SQL injection.

**Diallo Abdoulaye Kindy and Al-Sakib Khan Pathan:** Web applications play a very important role in individual life as well as in any country's development. Web applications have gone through a very rapid growth in the recent years and their adoption is moving faster than that was expected few years ago. Now-a-days, billions of transactions are done online with the aid of different Web applications. Though these applications are used by hundreds of people, in many cases the security level is weak, which makes them vulnerable to get compromised. In most of the scenarios, a user has to be identified before any communication is established with the backend database. An arbitrary user should not be allowed access to the system without proof of valid credentials. However, a crafted injection gives access to unauthorized users. This is mostly accomplished via SQL Injection input. In spite of the development of different approaches to prevent SQL injection, it still remains an alarming threat to Web applications.

**Tajpour, A :** Database driven web application are threaten by SQL Injection Attacks (SQLIAs) because this type of attack can compromise confidentiality and integrity of information in databases. Actually, an attacker intrudes to the web application database and consequently, access to data. For stopping this type of attack different approaches have been proposed by researchers but they are not enough because usually they have limitations. Indeed, some of these approaches have not implemented yet and also most of implemented approaches cannot stop all type of attacks. In this paper all type of SQL injection attack and also different approaches which can detect or prevent them are presented. Finally we evaluate these approaches against all types of SQL injection attacks and deployment requirements.

**Govind Murari Upadhyay, Kanika Dhingra:** The focus of this paper is to bring in light the value of Web Content Mining. The paper gives an insight into its techniques, processes and its applications in the current cut-throat business environment as well in research and extracting contents for educational purposes. It further explains how using web content mining plays an integral role by getting rich set of contents and uses those contents in the decision making in the corporate environment, education and research.

**Sandra Sarasan :** Today’s Web applications can contain dangerous security flaws. The global distribution of these applications makes them prone to attacks that uncover and maliciously exploit a variety of security vulnerabilities. Research reports indicate that more than 80 percent of the web applications are vulnerable to security threats. User friendly web applications are developed to increase the customer base and hackers utilize the features provided by the web applications to inject their malicious code. Web applications might contain security vulnerabilities that are not seen to the owner of the application. This paper presents multiple solutions to prevent web applications from the major security attacks such as SQL Injection and Cross Site Scripting. Each of the solutions have their own strengths and weaknesses, and the developers must choose the solutions according to their software development requirements.

**Sharmin Rashid, Subhra Prosun Paul:** This paper is on "Proposed methods of IP Spoofing Detection & Prevention". This paper contains an overview of IP address and IP Spoofing and its background. It also shortly discusses various types of IP Spoofing, how they attack on communication system. This paper also describes some methods to detection and prevention methods of IP spoofing and also describes impacts on communication system by IP Spoofing. We think that our proposed methods will be very helpful to detect and stop IP spoofing and give a secured communication system

**Pallavi Asrodia, Hemlata Patel:** In the past five decades computer networks have kept up growing in size, complexity and, overall, in the number of its users as well as being in a permanent evolution. Hence the amount of network traffic flowing over their nodes has increased drastically. With the development and popularization of network Technology, the management, maintenance and monitoring of network is Important to keep the network smooth and improve Economic efficiency. For this purpose packet sniffer is used. Packet sniffing is important in network monitoring to troubleshoot and to log network. Packet sniffers are useful for analyzing network traffic over wired or wireless networks. This paper focuses on the basics of packet sniffer; it's working Principle which used for analysis network traffic.

## III. COMPARISON

| Paper | Features | Advantages | Limitations |
|---|---|---|---|
| Web Content Mining: Its Techniques and Uses [12] | Web content mining tools are used to extract the useful information from web pages. | By these tools we can make our search of contents over the web faster and exact. | The web continues to increase in size and complexity with time hence making it difficult to extract relevant information. |
| Security vulnerabilities in Web Application-An Attack Perspective.[9] | Preventive measures of vulnerable online attacks of web servers are use. | All kind of vulnerable attacks can be prevented using this penetration tool. | The importance of web application and its security increasing day by day, but traditional networks fails to provide security for web application. |

## VI. CONCLUSION

The increasing use of the web paradigm for the development of pervasive applications is opening new security threats against the infrastructures behind such applications. Web applications developers must consider the use of support tools to guarantee a deployment free of vulnerabilities, such as secure coding practices, etc. However, attackers continue managing new strategies to exploit web applications. The significance of such attacks can be seen by the pervasive presence of web applications for health care, banking, government administration, and so on. Here we discuss some of the defensive measures which can be deployed in web applications to prevent attacks. But each has their own strengths and weaknesses to deal with. Thus an efficient solution to prevent attacks should be enforced to provide security in all environments of web applications. The current problem lies in the integration of these attack prevention techniques in a practical environment.

## REFERENCES

[1] Govind Murari Upadhyay, Kanika Dhingra., " Web Content Mining: Its Techniques and Uses,"
International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2013.

[2] D Sharmin Rashid, Subhra Prosun Paul, "Proposed Methods of IP Spoofing Detection and Prevention ", International Journal of Science and Research (IJSR), India Volume 2 , August 2013.

[3] Pallavi Asrodia, Hemlata Patel, "Network Traffic Analysis Using Packet Sniffer" International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, May-Jun 2012, pp.854-856; https://www.owasp.org/index.php/XSS Cross Site Scripting) Prevention Cheat Sheet.

[4] Rimmy Chuchra, Bharti Mehta, Sumandeep Kaur, "Use of web Mining in Network Security", International Journal of Emerging Technology and Advanced Engineering , Volume 3, Issue 4, April 2013.

[5] Roop Ranjan, Sameena Naaz, Neeraj Kaushik, "Web Miner: A Tool for Discovery of Usage Patterns from Web Data,", Volume 5 ,Issue 5,May 2013.