Intrusion Detection Techniques For Web Application Attacks

[1]Fasnamol A A, [2]Ajith S
[1]PG Scholar, [2]Assistant Professor
[1]Cmputer Science And Engineering,
[1]Rajagiri School Of Engineering and Technology, Kakkanad, India

Abstract: Quick expansion in web development provides a way to deliver complex business solution. The rise in web dependency led to the increase in web application attacks. Hackers use different techniques or methods for exploiting the vulnerabilities of a web application. Based on these methods or techniques the attacks are categorized. This paper discusses some of the most prevailing attacks on web applications and various methods for detection of such attacks. Different anomaly detection techniques are used to resolve attacks as well as to attain decent security. A Intrusion detection system has been used to attain security. It is a process which is used to identify unusual activities in a computer system. When a user opens a website his or her clicks are stored in a log file. This paper highlights the survey done on different methods in web log intrusion detection. In few works certain additional features were used for live detection of anomaly which has been included in this survey.

IndexTerms - Web application attacks, Intrusion detection system, Web security.

Introduction

Nowadays the use of web applications for banking, social interactions, business purposes etc. are drastically increasing. People are choosing these web applications in this busy life to get things done easily. Because of this mentality of user developers try to make the web application more fast and dependable. As the developers are trying to achieve more reliability they may compromises in standard coding. These drawbacks are utilized by the attackers to get private information's or to completely destroy particular web application[1].

Vulnerabilities in a web application are increasing. Web application attackers utilizes these vulnerability for adding malicious code to the application which helps in destroying the particular application. Hackers are mostly exploiting the vulnerability of a client and web server interaction phase in a web application architecture. In client and web server interaction phase the client sends http request to the server, using this request hacker will be able to make changes to the web application. It is essential to detect and prevent the attacks in a web application. There are intrusion detection systems available to detect these attacks.

Intrusion detection system is software that observes web application for malicious activity. Intrusion Detection System is classified into two misuse detection and anomaly detection. Misuse detection helps in finding known attacks and it is also called signature based attacks. It helps in detecting the known attacks by matching input with the set of signature that represents the attacks. Anomaly detection system is for detecting unknown attacks. The detection of anomaly is based on finding abnormality by analyzing collection of normal behavioral pattern.

In most of the web applications intrusion detection techniques the web log is taken as the input. Web log can be stored in different formats. Choosing the web log format may vary based on each intrusion detection system. Most commonly used one is W3C format , which can be customized . Fields contained in the web log of W3C format are:

- Request date and time.
- Users IP address.
- Number of Bytes returned
- HTTP status code
- URI
- Http Version

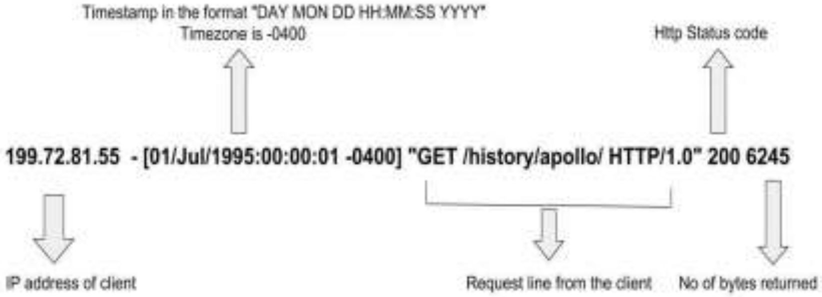

Timestamp in the format "DAY MON DD HH:MM:SS YYYY"
Timezone is -0400
Http Status code
199.72.81.55 - [01/Jul/1995:00:00:01 -0400] "GET /history/apollo/ HTTP/1.0" 200 6245
IP address of client
Request line from the client
No of bytes returned

Fig1: Example of web log with field description

In this paper we are discussing some of the most prevailing attacks as well as some of the intrusion detection systems which helps in protecting web application from attacks. The comparison of these detection systems are carried out in this work to find the efficient one.

I. BACKGROUND KNOWLEDGE

Security in web application becomes a important issue now a days. Research on various web application attacks and their detection techniques are described here.

2.1 Web Application Attacks

Web application vulnerabilities are been exploited by hackers for attacking purposes. Hacker uses different ways for attacking, out of which most prevailing ways are classified as various prevailing attacks. According to the Web Application attack statistics by positive research group the Sql injection attack becomes the common attack because out of all attacks 25.5% is sql injection attack. The web application statistics of all major attack is depicted in Fig 2.

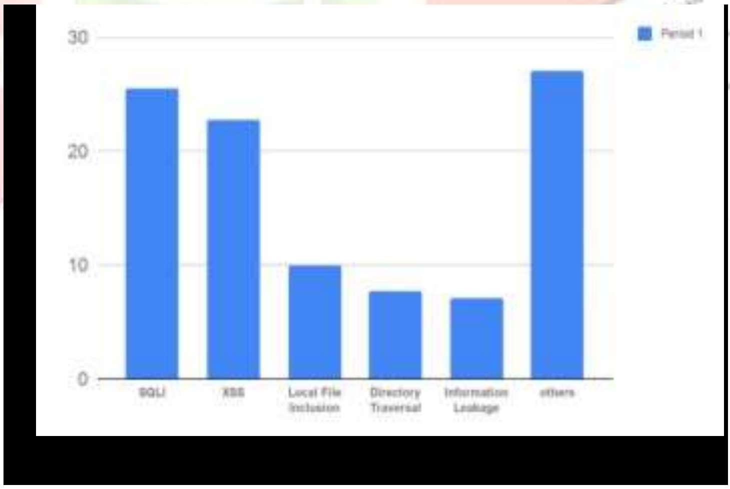


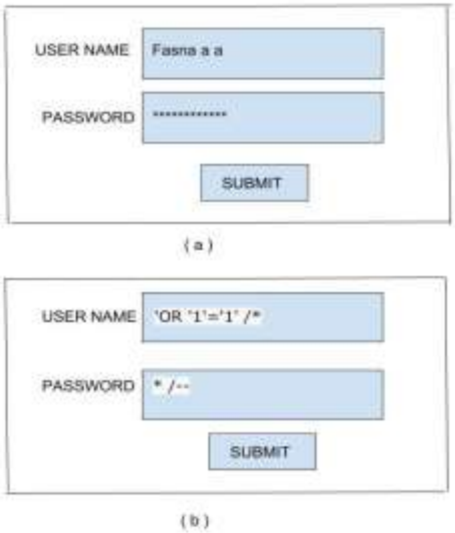
Fig 2. Web application attacks: Statistics

In this section we are focusing on some of the most prevailing attacks for web application such as:

2.1.1 Sql Injection Attack

Sql injection attack is one of the most prevalent attack in which hackers use malicious sql query for manipulating the database. Through this attack hackers can be able to display , delete or manipulate most relevant information of a web application. A malicious user can generate a web request from which sql query can be invoked and that can be able to manipulate the database.

Hackers can manipulate sensitive information through this sql injection attack. By using this type of attack hackers can use any sql operations such as UPDATE, DELETE, UNION etc. to manipulate information of a web application.


USER NAME Fasna a a

PASSWORD *************

(a)

USER NAME 'OR '1'='1' /*

PASSWORD * /--

(b)

Fig 3. Example of web log with field description

Sql injection attack example is depicted in the above figure Fig 3. In the above example in order to get an access to the web site particular user name and password should be valid. It is possible to access the user name and password of all the users through sql injection attack. Fig 3a depicts a valid request but Fig 3b depicts invalid injection techniques to forcibly display the information of all the users. So that hackers can access an account and manipulate entire databases[2].

2.1.2 XSS Attack

Cross Site Scripting is a type of attack where hackers inject malicious script to a web application. It is actually a front end attack. By injecting these malicious content attackers will be able to attain sanctioned access over sensitive content,cookies etc. For the successful execution of the XSS, attackers should first be clear about the vulnerability of the web applications and after that can inject the malicious script in to the server. In case of XSS attacks hackers can add malicious links to the web application by utilizing its vulnerability. For example; in most of the social web applications it is possible to add comments, So hackers utilize the vulnerability and add following comments for example : " Great Discount For This Item " Click the below Link "$<$script src=http://xyz.com/unknown$>$<$/script$>$. The malicious script get executed whenever we access the particular web page of an application where the script is written. In XSS attacks links looks normal at the beginning and it turns to malicious one[3].

2.1.3 Cross Site Request Forgery

Cross Site Request Forgery attack is also known as session Riding attack. CS RF can cause huge damage to the authenticated web applications. Cross Site Request Forgery attack is implemented using links or email that tricks users browser for sending request to a web application server. When the user tries to open the link or email generated by the hacker, it automatically results in sending HTTP request to some authenticated web application server. While sending this http request browser of the user will include cookie header. And this cookie header contained a confidential data such as session identifier of the user. Because of having this confidential information in the request the web server may think that the request is a legitimate one and the web server will serve the request. This attack is mainly effected by Banking sites to hack money from account [4].

2.1.4 Directory traversal attack

Directory traversal attack is also called as Path Traversal Attack. In this type of attack attacker tries to perform unauthorized access to directories outside the web servers root directory. Here the unauthorized directory is accessed and manipulated by making malicious http request to the server. For implementing this method ,First the attacker try to find the possibility of accessing a file inside web servers root directory. If it is possible to access it then attacker will note the particular url that helps in accessing file inside the servers web root directory. Then attacker tries to manipulate that url logically to access unauthorized data. Attacker may add some " ../" to the url, because that permits system to access one directory up.

2.1.5 CRLF injection attack

In the CRLF injection attack Cr represents carriage return and LF represent line feed. Cr and Lf are used to indicate the end of a line. In HTTP protocol CR-LF sequence is used to indicate the end of line. In a normal web application the http header is identified based on the CR, LF sequence in the request. In this CRLF injection attack, Hackers add CR-LF sequence in the http header in order to get the control of http response message. Control over the http response message can be utilized to modify the response message header and can add malicious activity.

2.1.6 Denial Of Service Attack

It is a type of attack that can make web services unavailable. Web services can be made unavailable by loading target server with lot of request. This type of attack is possible for crashing a particular server or the attack can make the server down. With this kind of attack can make web server possible to reject valid user request. DDOS attack is possible to get accomplished by a single host as well as multiple host[5].

2.2 Intrusion Detection Techniques

There are several techniques for detecting the intrusion in the web application by analyzing the web log. Some of these techniques are described below.

2.2.1 Entropy Based Detection Technique

T. Threepak and A. Watcharapupong[6] proposed a model using entropy based technique. In this model web server's log file is taken as the input ans Shannon textual entropy[7] calculation method is used for web application attack detection. In entropy based method, First prepossessing of web log data is done by removing erroneous log and log containing negligible attribute. Then will extract request string from each web log, this request string is further split-ted into word tokens. After splitting will calculate Shannon textual entropy of each word token and based on entropy value of word tokens in each request string its relative entropy is calculated. Average and Standard deviation of the relative entropy value in a specific time duration.

$$E_t(p_1\ldots\ldots..p_n)=\frac{1}{\lambda}\ \sum_{i=0}^{n} p_i\ [\log_{10}(\lambda)\ -\ \log_{10}(P_i)]\ldots.(1)$$

$$E_{Rel}\ =\frac{E_t}{E_{max}}\times 100\ldots\ldots\ldots.(2)$$

$$E_t(p_1\ldots\ldots..p_n)=\frac{1}{\lambda}\ \sum_{i=0}^{n} p_i\ [\log_{10}(\lambda)\ -\ \log_{10}(1)] = \log_{10}(\lambda)..(3)$$

In this proposed model Shannon textual entropy and relative entropy is calculated using the above equations, Where E_t represent textual entropy and E_{Rel} represent relative entropy. Lambda represents number of words in a text with n different words

2.2.2 Hybrid Intrusion Detection Technique

Jing Yo, Dan Tao and Zhawen Lin proposed a hybrid model for detecting intrusion in a web log generated by a web application[site].In [8] author combines misuse detection technique and anomaly detection technique using clustering to improve the rate of attack detection. Here web log data with attribute file is taken as the input. At first preprocessing of web log data is done to remove both the log without having a standard format and the logs that containing error. Then the preprocessed log is given to misuse detection model, In this set of rules were defined to detect the known attacks. Regular expressions are used in this model to define the rules. Using these rules matching process is performed, In order to check the similarity of each web log with these defined rules. If matching is found that web log was an anomaly else that web log is added to anomaly detection part.

In anomaly detection part first feature extraction is done, In feature extraction phase request path length and other seven implicit features such as request parameter number etc are extracted and these extracted features of each web log are used for clustering purpose. Here K-mean clustering is done. After clustering average mew and standard deviation of each cluster is computed . It is based on calculating Euclidean distance of the centroid of a cluster to its data points. After these Chebyshev inequality helps to detect anomaly of web log.

2.2.3 N-gram model based Intrusion Detection Technique

Eric Asselin,Carlos and Gentian[9] proposed a model to detect intrusion in a web log using N-gram model. Web log without any anomaly used for training purpose. In this system first phase is applying crawling method. In crawling method, the web crawler helps to construct the training data for a web application whose intrusion needed to be detected. Next is N-gram model phase, In N-gram model used for creating connecting sequence of n-item from a word. Here in this system sequence of 2-items are selected from each web log which is named as bigram model. Then the probability of each bigram in the URL is calculated. Then harmonic mean of the probabilities of each bigram in the URL calculated. A threshold value is computed in order to separate normal and anomalous URL'S. The value assigned to the threshold is the lowest of all the harmonic mean calculated for each URL. Then in order to detect the intrusion in a URL data either one of the criteria need to satisfy. First harmonic mean of the probabilities of bigram should be greater than given threshold and the second is the Http status code of the given URL should not exceed 400.

2.2.4 Frequent Closed Episode Rule Based Intrusion Detection Technique

Lei Wang1, Shoufeng Cao2, Lin Wan3, and Fengyu Wang1[10] all proposed a model for detecting attacks in a web application by analyzing the web log. In this work detecting the abnormality in the web log is based on generating Frequent Closed Episode Rule by using FCER mining algorithm. Frequent Closed Episode Rules are used for analyzing the pattern of training data set. The architecture of the model is depicted in the Fig 4. Training and testing phase are there in the system. In training phase, At first log data is preprocessed which is done by extracting necessary information from each log line such as Ip address, date and requested resource and removed the log requested for the image. Last in preprocessing will generate symbols corresponding to each log line. Frequent Closed Episodes are generated from the preprocessed data. From these the rules are mined using FCER mining algorithm.

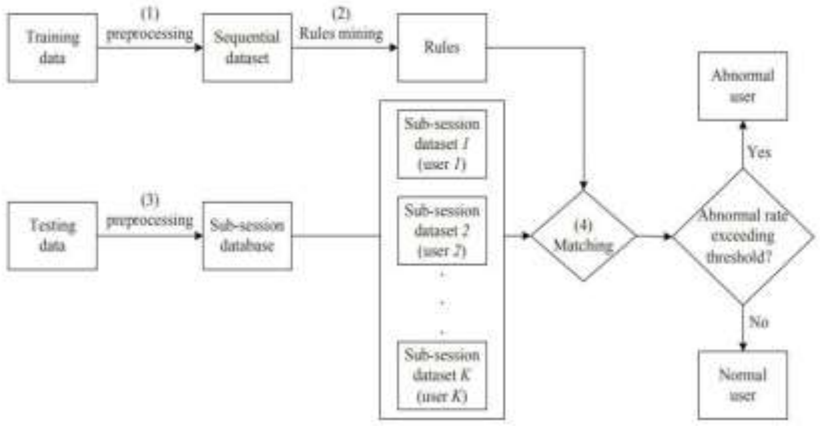


Fig. 4. Low Level Architecture

In the testing phase, Preprocessing of testing data is the first step. Here procedure done in preprocessing is same as that done in training phase. Then preprocessed data needed to be converted into sub-session database by setting the window size. Then sub-session data is further divided based on each user. Then calculated the abnormal rate for all the training data using below equation where and set the threshold as the greatest of all the abnormal rate in the normal data which is used for training purpose. In matching process abnormal rate of the testing data compared with the threshold and if abnormal rate of the testing data greater than threshold then it is anomalous one.

Table 1: Comparison of different intrusion detection techniques

PAPER TITLE	TECHNIQUES USED		DATA SET	ACCURACY/RESULT	COMMENTS
Web Attack Detection Using Entropy Based Analysis	Shannon Textual Entropy	Sql injection XSS Directory Traversal	Normal web log from a university web server and Some attacks are added from a Nessus.	Out of 93550, 11337 are classified High risk,Medium Risk and Low risk attacks	Its feasibility is less compared to all types of input log. High false alarm rate compared to others
Hybrid Web Log Based Intrusion Detection Model	K-means Clustering Chebyshev Inequality	XSS CSRF SQL Injection	Official web site of BUPT	96.70% attacks are detected	It is an efficient technique and suitable for any new kind of attack.

Anomaly Detection For Web Log: A Simple Yet Crawling Based Approch.	N-gram model Harmonic Mean	Brute Force Attack Denial Of service Missing common resource	Web log collected from a 30 live production of web sites hosted on a single server. Access log consist of 10 million queries.	Performance of log reduction scheme varies from13% to 95% of the web application.	Inefficient for detecting attack that uses cookies or post data as a vector.
Web log anomaly Detection Based On Frequent Closed Episode Rule	FCER mining Algorithm	Log collected from DWVA application 30308 log line	Brute Force Attack CSRF XSS Sql injection	96% of attacks are detected	Implemented in Hadoop environment so possible to handle large data effectively. Less feasible to all inputs

References

[1] R. **V**. Bhor and H. K. Khanuja, *”Analysis of web application security mechanism and Attack Detection using Vulnerability injection tech- nique”*,2016 International Conference on Computing Communication Control and automation (ICCUBEA), Pune, 2016, pp. 1-6.

[2] O. B. Al-Khurafi and M. A. Al-Ahmad, *”Survey of Web Application Vulnerability Attacks”*,2015 4th International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Kuala Lumpur, 2015, pp. 154-158.

[3] P. A. Sonewar and S. D. Thosar, *”Detection of SQL injection and XSS attacks in three tier web applications”*,2016 International Conference on Computing Communication Control and automation (ICCUBEA), Pune, 2016, pp. 1-4.

[4] G. P. Bherde and M. A. Pund, *”Recent attack prevention techniques in web service applications”*,2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp. 1174-1180.

[5] D. S. N. Mary and A. T. Begum, *An Algorithm for Moderating DoS Attack in Web Based Application”*,2017 International Conference on Technical Advancements in Computers and Communications (ICTACC), Melmaurvathur, 2017, pp. 26-31.

[6] T. Threepak and A. Watcharapupong, *Web attack detection using entropy- based analysis”*,The International Conference on Information Network- ing 2014 (ICOIN2014), Phuket, 2014, pp. 244-247.

[7] C. Shannon, *Prediction and entropy of printed English”*, Bell Systems Tech Jour 30, pp. 50-64, January 1951.

[8] J. Yu, D. Tao and Z. Lin, *A hybrid web log based intrusion detection model”*,2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), Beijing, 2016, pp. 356-360.

[9] E. Asselin, C. Aguilar-Melchor and G. Jakllari, *Anomaly detection for web server log reduction: A simple yet efficient crawling based approach”*,2016 IEEE Conference on Communications and Network Security (CNS), Philadelphia, PA, 2016, pp. 586-590.

[10] L. Wang, S. Cao, L. Wan and F. Wang, *”Web Anomaly Detec- tion Based on Frequent Closed Episode Rules”*,2017 IEEE Trust- com/BigDataSE/ICESS, Sydney, NSW, 2017, pp. 967-972.