I.J. Wireless and Microwave Technologies, **2015, 6, 53-62**
Published Online November 2015 in MECS(http://www.mecs-press.net)
DOI: 10.5815/ijwmt.2015.06.07

Available online at http://www.mecs-press.net/ijwmt

A New Technique for Detection and Prevention of Passive Attacks in Web Usage Mining

Roop Kamal Kaur [a,*1], Kamaljit Kaur [b,*2]

[a] *Computer Science and Engineering Department, Sri Guru Granth Sahib World University, Fatehgarh sahib,140406, India*
[b] *Computer Science and Engineering Department, Sri Guru Granth Sahib World University, Fatehgarh sahib,140406, India*

Abstract

Data theft and data attack is very common with internet users and web log files. Web log file is the proof of authentication of any user over any website. Many unauthorized users change them according to their requirement. Attacks like sql injection, brute force attack, IP spoofing might cause severe damage to the database system In this encryption techniques and optimization technique have been applied in order to check the unexpected arrival of messages or queries and the system parameter.

Index Terms: Attacks, Genetic algorithm, Encryption techniques.

1. Introduction

The internet has become the important part of our life. The World Wide Web has been affecting both its user and web sites owners. The websites and visitors have increased remarkably. Data mining techniques are used to mine interesting data. The data mining techniques are not applied directly to the web data because the web data is unstructured or semi-structured that's why we use web mining which can be applied to web data. Web mining uses the techniques of Data mining for discovering and extracting the information from Web related documents and services. There are three types of information that can be discovered by using web mining that are Web activity, server logs and activity of Web browser tracking. Web usage mining is one of the types of Web mining which is used to extract meaningful and interesting patterns of usage in web site which can be uses in a various ways like improvement of web sites, detecting of wrong information, and behavior of user etc [1]. Web Security plays an important role in web application. Destructive causes are because of lack of security in the organization. There are two types of web security namely web browser security and web application

* Corresponding author
E-mail address:

security [2].Web browser security is important when the website is attacked by the attacker.

Nomenclature
n used as the modulus for both the public and private keys.
e termed as the public key exponent.

1.1. Web Mining

Web Mining is the application of data mining . Data mining is the process of extracting data from information. Web mining is used to search the various types of patterns from the Web. Web mining is of three different types:

1.1.1. Web Usage Mining

Web usage mining is used for extracting important information from server logs .Web usage mining is the process of finding out the behavior of users that what the users are looking for on the Internet. Some users search for text based data, whereas some other users may be interested in multimedia data. Web Usage Mining fulfills the requirement of web based applications by discovering the interesting usage patterns from Web. Web usage mining classification is based on the kind of usage data considered [3]:

- *Web Server Data*: Web server contains the user logs. It consists of IP address, page reference and access time.
- *Application Server Data*: Commercial application servers have important features to enable e-commerce applications to be built on top of them with less effort. An important feature is the ability to track different kinds of business events and log them in to the application server logs.
- *Application Level Data:* various types of events can be defined in an application, and logging can be turned on for them thus generating histories of these specially defined events.

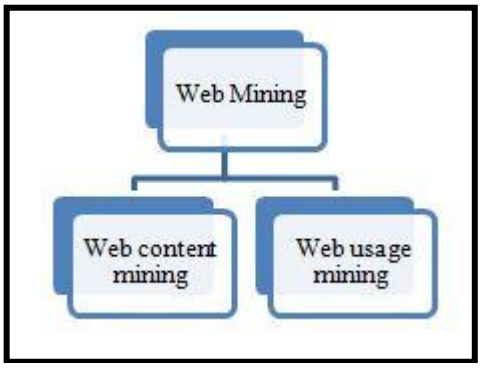

Fig.1. Types of Web Mining

1.1.2. Web Structure Mining

Web structure mining uses the graph theory to analyze the node and connection structure of a web site. Based on web structural data, web structure mining is divided into two kinds:

- Hyperlinks connects the web page to different locations.

- Structure of pages look like tree structures analysis is used to describe HTML or XML tag usage.

1.1.3. Web Content Mining

Web content mining consists of extraction and integration of useful data, information and knowledge from Web page content.

1.2. Attack

An attack is an attempt to destroy, expose, alter, disable, steal information or gain unauthorized access to and make unauthorized use of an asset. A security attack is defined as a threat, intrusion, and denial of service attack on a network that will analyze your network and gain information to cause your network to crash or corrupted. The types of attack have been shown in fig.2.

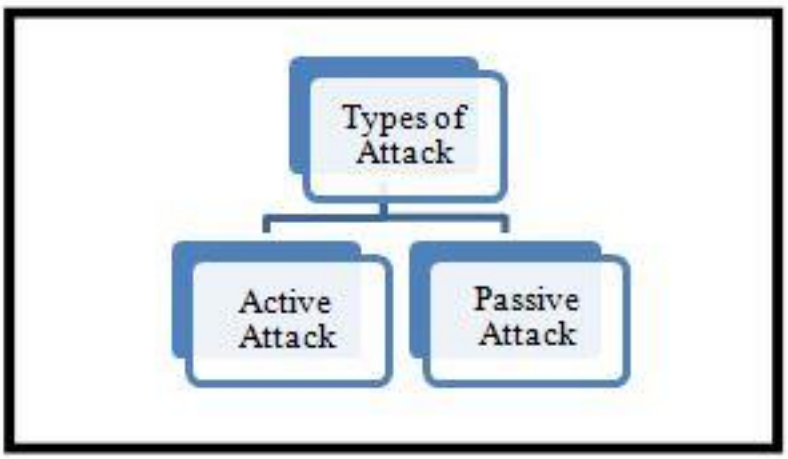

Fig.2. Types of Attack

1.2.1. Active attack

Active attack**,** the attacker tries to bypass into systems through viruses, worms and Trojan horses. Active attacks include attempts to destroy protection features introduction of malicious code, to steal or modify information.

1.2.2. Passive Attack

A passive attack monitors unencrypted data and looks for clear-text passwords and important information that can be used in other types of attacks. Passive attacks are traffic analysis, observing of data which is not protected and capturing authentication information such as passwords. Results of Passive attacks are exposure of information to an attacker without the knowledge of the user. The types of passive attacks are shown in fig 3.

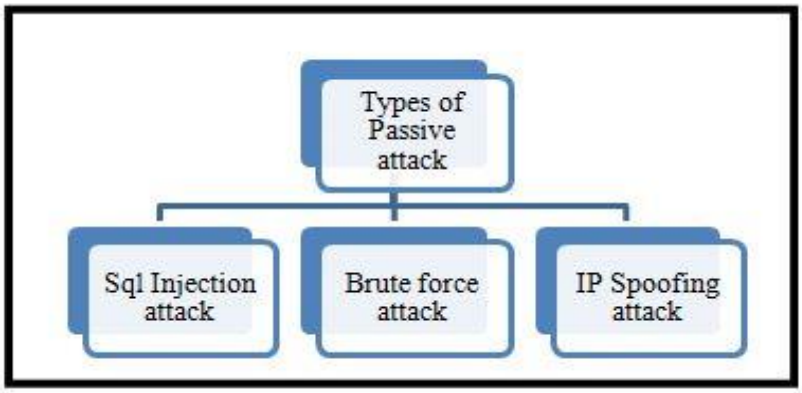

Fig.3. Types of Passive Attack

1.2.2.1. SQL injection

SQL injection is a technique of code injection, used to attack data-driven applications, so malicious SQL statements are inserted into an entry field for execution. SQL injection must destroy a security vulnerability in an application's software when user input is either incorrectly filtered for string literal various escape characters are embedded in SQL statements or user input is not strongly typed and unexpectedly executed. SQL injection is known as an attack vector for websites but can be used to attack any type of SQL database [6].

1.2.2.2. Brute force attack

Brute-force attack is an attack that can be used against any encrypted data . It consists of guessing all possible keys or passwords again and again until the correct one is found. The encryption types used are i.e. 64-bit, 128-bit or 256-bit encryption. The amount of time is dependent on the complexity of the password, encryption strength and the strength of the computer used to conduct the attack.

1.2.2.3. IP spoofing attack

IP address spoofing or IP spoofing is the creation of Internet Protocol (IP) packets with a forged source IP address, to fulfill the purpose of concealing the identity of the sender. IP spoofing attack is an attack in which one person or program successfully masquerades as another by falsifying data and gains an illegitimate advantage [8].

2. Literature Review

2.1. Encryption

Encryption is the process of encoding information in such a way that unauthorized parties cannot read it. The information is referred to as plaintext, which is then encrypted by an algorithm, cipher text is generated that can only be read if decrypted [9].

2.1.1. Types of Encryption

- *Symmetric key encryption*

In symmetric-key schemes, the encryption and decryption keys are the same. Both sender and receiver have the same key before they can achieve secret communication.

- *Public key encryption*

In public-key encryption schemes, the encrypted key is used to encrypt messages. However, only the receiver has access to the decryption key that enables messages to be read.

2.2. Types of Encryption techniques

Various types of algorithms for encryption are like

2.2.1. DES (Data Encryption Standard):

DES is a symmetric block cipher. 56-bit key is used to encipher and decipher a 64-bit block of data. The key is 64-bit block, every 8th bit of which is ignored. DES is the most widely used symmetric algorithm in the world, despite the key length is too short. Triple length key consist of three 56-bit keys that are K1, K2, K3 then encryption is as follows:

- Encrypt with K1
- Decrypt with K2
- Encrypt with K3

Decryption is the reverse process:

- Decrypt with K3
- Encrypt with K2
- Decrypt with K1

2.2.2. RSA:

RSA is a public key algorithm invented by Rivest, Shamir and Adleman. The encryption key is different from the key used for decryption. The problem with choosing long keys is that RSA is very slow compared as compared to DES.RSA algorithm is used for digital signatures and for protecting DES keys.RSA involves a *public key* and a *private key*. The public key is known by everyone and is used for encryption. Messages encrypted with the public key can only be decrypted using the private key.

RSA algorithm keys are generated in following way:

1. Choose two distinct prime numbers p and q.
2. Compute $n = pq$.
3. Compute $\varphi(n) = \varphi(p)\varphi(q) = (p - 1)(q - 1) = n - (p + q - 1)$, where φ is Euler's totient function.
4. Choose an integer e such that $1 < e < \varphi(n)$ and $\gcd(e, \varphi(n)) = 1$ which means that e and $\varphi(n)$ are coprime.

- E having a short bit-length and small Hamming weight results in more efficient encryption – most commonly $2^{16} + 1 = 65,537$. Where smaller values of e (such as 3) have been shown to be less secure in some settings.

- Determine d as $d \equiv e^{-1} \pmod{\varphi(n)}$ solve for d given $d \cdot e \equiv 1 \pmod{\varphi(n)}$
- This is often computed using the extended Euclidean algorithm. And d is kept as the private key exponent

2.2.3. AES (Advanced Encryption Standard):

AES is a variant of Rijndael which has a fixed block size of 128 bits, and a key size can be of 128, 192, or 256 bits. AES operates on a 4 ×4 column-major order matrix of bytes, although some versions of Rijndael have a larger block size and have additional columns in the state.

Description of the algorithm

1. *Key Expansion*—round keys are derived from the cipher key using Rijndael's key schedule. 128-bit round key block is required for each round plus one more.
2. *Initial Round: Add Round Key*—each byte of the state is combined with a block of the round key using bitwise xor.
3. *Rounds*

    1. Sub Bytes—a non-linear substitution step where each byte is replaced with another according to a lookup table.
    2. Shift Rows—a transposition step where the last three rows of the state are shifted cyclically a certain number of steps.
    3. Mix Columns—a mixing operation which operates on the columns of the state, which combines four bytes in each column.
    4. Add Round Key

4. *Final Round (no Mix Columns)*

    1. Sub Bytes
    2. Shift Rows
    3. Add Round Key.

2.3. Optimization

In optimization of a design, the design objective could be simply to minimize the cost of production or to maximize the efficiency of production. An optimization algorithm is a method which is executed iteratively by comparing various solutions till an optimum or a satisfactory solution is found. Optimization has become a part of computer-aided design activities.

2.3.1. Genetic Algorithm

Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems with the help of using inheritance, mutation, selection ,and crossover [10] .

The steps in Genetic Algorithm are:

1. *Initialization:* Chromosomes are randomly created. Population should be diverse. Otherwise, good solutions are not produced.
2. *Evaluation*: Each chromosome is rated on how well the chromosome solves the problem then fitness value is calculated on every chromosome.

3. *Selection*: The fittest chromosomes are selected for propagation into the future generation based on how fit they are.
4. *Recombination*: Individual chromosomes and pairs of chromosomes are then recombined and modified accordingly and then again put back into the population.

3. Proposed Algorithm

In the proposed algorithm encryption techniques and optimization technique have been applied in order to check the unexpected arrival of messages or queries and the system parameter .This algorithm helps to prevent the system from Brute force attack and IP spoofing attack, when so ever unwanted messages arrives to the system .No doubt it consumes a lot of time but though it is effective

1. *State 1: Start***:** it initialize the process and call the first step.
2. Here user can select two steps that refer to their corresponding algorithm. If select 1 than brute force algorithm else it proceed IP spoofing algorithm.
3. *Check point***:** it checks the choice of user for execution of program and generates results. If choice is one than it executes stage first steps and if it is 2 than IP spoofing.
4. *Upload Log file***:** here algorithm's first level encryption implemented on log file. It works with RSA, AES and hybrid algorithm that is combination of RSA + AES. Control transfer to next step.
5. *Second level***:** It is next step of algorithm that works for transfer bits and generate assessment value for gives input to next step to complete the process.
6. *Checkpoint:* both system generated assessment value and user's assessment values are compared if they are same than access is allowed else it rejected by the system.
7. Stage 2: system generate password for given network. And encrypt session as per selection.
8. *Check point***:** if selection is one than it process for generate unique sophisticated password for network to make authentication system to improve security and process next step to compare with user input.
9. If both passwords are same than system gives authorization to access data. Else the request reject by system.
10. If phase two selected than it encrypt session and execute hybrid algorithm (RSA+AES) and generate output.

3.1. Sql Injection

Sql injection is a process where you execute a query in a website in order to extract information such as login information, users etc. for either personal gain or random use from the website database.

The steps for proposed algorithm are:

1 *Start:* it is the first step that initializes the process. It calls the login form to check authorization of user.
2 *Login:* here user enters the id and password for login and its collect data from user and send it to the next step.
3 *Process***:** it checks the data that collect from user for process if it is valid than it execute next step else the request transfer to second step to re submission of data elements.
4 *Check and prevent***:** it verifies the data from the sql injection attack. If it found System prevent the attack and request for access the system is rejected.
5 *Encryption***:** apply Triple DES algorithm and MD5 algorithm for encrypt data for save it from external access. If it is accessed than original data never extracted due to encryption. So that it is a beneficial approach for prevent attacks on data.
6 *Filter:* it gets encrypted data from upper layer and processes it to decrypt data and execute it to sql server for produce results. If user authorised than it transfer process to file access mode else it shows

message to user for invalid attempt.

7 *Stop***:** Remove all garbage collection from memory and terminate process.

4. Results and Discussions

Results show the comparison between Encryption techniques using genetic algorithm in preventing Brute force attack, IP spoofing and SQL injection attack.

4.1. Graphical Representation for Brute Force Attack:

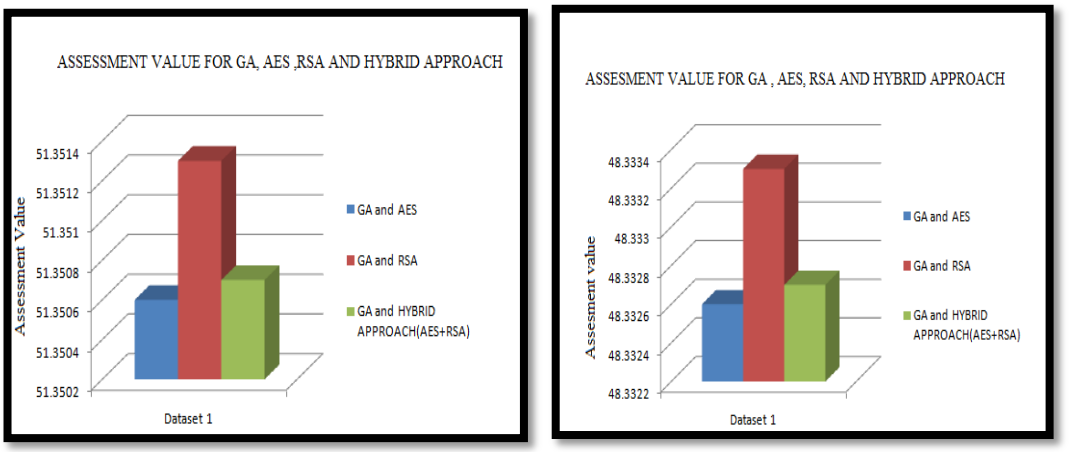

Fig.4. Comparison Shown Between Different Combinations of Encryption Algorithm with GA.

The above figure represents the assessment value on a bar chart representation. The above graph shows the assessment value for GA combined with RSA, GA Combined with AES and GA combined with HYBRID algorithm which is a production of RSA and AES with slight changes. The graph states that the assessment value for RSA is more as compared to hybrid and AES where as the hybrid assessment value is more than that of AES.

4.2. Graphical Representation for IP Spoofing:

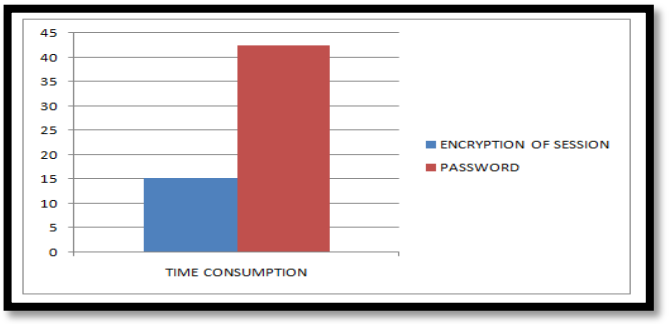

Fig.5. Comparison Shown Between Different Prevention Methods for IP Spoofing Attack on the Basis of Time Consumption.

The above graph represents the total time consumption in preventing the IP spoofing attack. It shows that time consumption for encrypting the session is less as compared to time consumption in password. It concludes that encrypting session is more effective then applying password to prevent from attacks.

4.3. Graphical Representation for Sql Injection

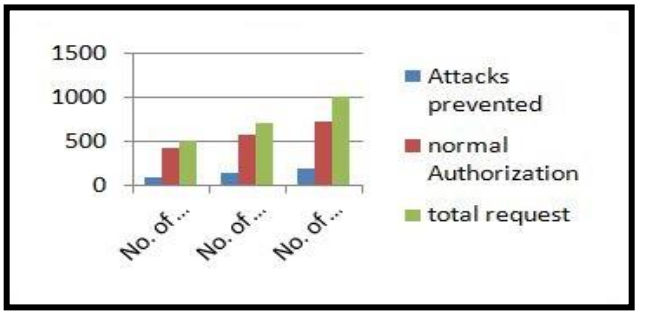

Fig.6. Results Showing the Information about the Attacks Prevented and Detected By the System

The graph shows the total number of requests send by user. Here code detects attacks from them and prevents the system from them. All results of processing are shown in the graph.

5. Conclusion and Future Scope

The current scenario of the work examines several different attacks like sql injection attack, brute force attack and IP spoofing attack. Although the results are quite satisfactory but use of combinational algorithm also increases the consumption of time with the big set of data. It is mandatory for Genetic algorithm to run several iterations to get an optimized value. If the Genetic algorithm will not possess iterations more than hundred, it would become difficult for Genetic Algorithm to propose an optimize solution

The current work is also not tested with very bulky data like about one million results. The limitation of the current work is also on with the size of data as bulky data would take one million of iterations to produce us an optimised value or solution. There has to be an algorithm which can limit the number of iterations over the size of data. In future, research workers can try their hand with back propagation neural network in which the numbers of iterations are certainly fixed with the amount of data passed to it. The accuracy would also be enhanced as back propagation model runs on hidden network modal. SQL injection prevention would also improve using encryption technique with RC5 algorithm. If the encryption scheme is stronger than we achieve more accuracy from the external attack.

References

[1] Dilpreet kaur, Sukhpreet Kaur, "A Study on User Future Request Prediction Methods Using Web Usage Mining", International Journal of Computational Engineering Research, Vol, 03, Issue, 4.

[2] S.Mirdula, D.Manivannan, "Security Vulnerabilities in Web Application- an Attack Perspective", International Journal of Engineering and Technology (IJET)**,** Vol 5 No 2 Apr-May2013.

[3] Rimmy Chuchra, Bharti Mehta, Sumandeep Kaur, "Use of web Mining in Network Security", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013).

[4] Jianli Duan, Shuxia Liu, "Research on web log mining analysis", Instrumentation & Measurement,

Sensor Network and Automation (IMSNA), 2012 International Symposium on (Volume: 2), 25-28 Aug. 2012.
[5] Daxin Jiang Jian Pei Hang Li, "Mining Search and Browse Logs for Web Search: A Survey"
[6] Diallo Abdoulaye Kindy, Al-Sakib Khan Pathan, "A Detailed Survey on Various Aspects of SQL Injection in Web Applications: Vulnerabilities, Innovative Attack, and Remedies"
[7] Sandra Sarasan, "Detection and Prevention of Web Application Security Attacks", International Journal of Advanced Electrical and Electronics Engineering, (IJAEEE) Volume-2, Issue-3, 2013.
[8] Sharmin Rashid, Subhra Prosun Paul, "Proposed Methods of IP Spoofing Detection and Prevention" International Journal of Science and Research (IJSR), Volume 2 Issue 8, August 2013.
[9] Swati Paliwal, Ravindra Gupta, "A Review of Some Popular Encryption Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 2, February 2013.
[10] K. F. Man, K. S. Tang, and S. Kwong, "Genetic Algorithms: Concepts and Applications", IEEE Transactions on Industrial Electronics, vol. 43, no. 5, october 1996.
[11] Inadyuti Dutt, Soumya Paul, Dr. S.N. Chaudhuri, "Implementation of Network Security using Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 2, February 2013.
[12] Ann Mary Jacob, Saritha S, "Survey on Various IP Spoofing Detection Techniques", International Journal of Science and Research (IJSR) ISSN: 2319-7064.
[13] Rajan.S.Jamgekar, Geeta Shantanu Joshi, "File Encryption and Decryption Using Secure RSA", International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-1, Issue-4, February 2013
[14] Mandeep Singh, Narula Simarpreet Singh, "Implementation of Triple Data Encryption Standard using Verilog", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1, January 2014.
[15] Shaneel Narayana, Michael Fitzgerald, "Empirical Network Performance Evaluation of Security Protocols on Operating Systems", IJWMT Vol. 2, No. 5, October 2012.

Author(s) Profiles

Roop Kamal Kaur has obtained her B.Tech degree in 2012 and M.Tech degree in 2014. Presently she is working as a lecturer in GGSCMT, Punjab, India.

Kamaljeet Kaur is working as an Assistant Professor in Sri Guru Granth Sahib World University, Punjab, India.

How to cite this paper: Roop Kamal Kaur, Kamaljit Kaur,"A New Technique for Detection and Prevention of Passive Attacks in Web Usage Mining", IJWMT, vol.5, no.6, pp.53-62, 2015.DOI: 10.5815/ijwmt.2015.06.07