Prevention of DDoS and Brute Force Attacks on Web Log Files using Combination of Genetic Algorithm and Feed forward Back Propagation Neural Network

Jaspreet Kaur
Chandigarh University
Mohali (Punjab)

Rupinder Singh
Chandigarh University
Mohali (Punjab)

Pawandeep Kaur
Chandigarh University
Mohali (Punjab)

ABSTRACT
World Wide Web has become an ultimate source of information. Traditional services such as banking, education, medicine, defence, and transportation are being presented by web applications. Whenever the users make use of any web application, all the activities of the users get automatically get appended into the web log files. The web log file data helps the website owners in number of ways such as customization of web content, pre-fetching and caching, E-commerce, etc. However, the web log file data is exposed to number of attacks. In this paper, a new technique has been proposed to prevent the log file data from the two most common attacks: Brute force attack and DDoS attack.

Keywords
Web Log Files, web applications, Brute Force, DDoS

1. INTRODUCTION
Now-a-days, Internet is playing a vital role in our day-to-day life. World Wide Web has become the vast and ultimate source of information. The exceptional growth of internet has transformed the way of operation of traditional services such as banking, education, transportation, medicine and defence. These services are being presented by web-based applications. However, these web-based applications are vulnerable to number of attacks. So, securing the web data has become a must to do task. Web log files are automatically created and maintained by web servers. When the users visit any web site, their all activities are appended into web log files. Even a single click, including images, html document or other objects get logged in web log files [1]. The information in Log files includes User Name, Visiting path, Path traversed, Time Stamp, Page Last Visited, User Agent, URL, and Request type [2]. There are basically three log file formats: Microsoft IIS Log File, W3C extended Log File Format, NCSA Common Log File Format. Web Log files undergoes number of attacks such as brute force attacks, DDoS attacks, SQL injection, Information Leakage and improper error handling, Cross site request forgery, Malicious file execution, etc.

1.1 Brute Force Attacks
Brute Force attack is a password guessing attack by applying the each possible combination of letters, symbols and numbers. Password based web applications are more vulnerable to Brute Force attacks. Web sites requiring user authentication are a good target of brute force attack. But this attack mainly works for short passwords as the possible combinations for larger passwords are very difficult to generate. The various types of brute force attack includes: targeted attack, trawling attack and blind attack [3]. In targeted attack, the attacker tries to guess the password by using some brute force strategy that would help in authentication of the user. Trawling attack is the reverse of targeted attack. In this attack, the attacker tries to find out the user of a particular password [3]. In blind attack, the attacker tries to find out the account identifier name as well as the password at the same time.

1.2 DDoS Attacks
Distributed Denial of Service (DDoS) attacks are a way of preventing the legitimate users from getting services by exhausting the resources of servers. Zombie machines are created by the attackers to launch these attacks. These machines consume the resources of servers for a long time and thus making it unavailable for the legitimate users [4]. Generally there are two forms of DDoS attacks: one that crash the services and the other that flood the services. Web applications are predominantly vulnerable to denial of service attacks. The symptoms of DDoS attack could be slow network performance, unavailability of the website, remarkable increase in the number of spam emails and inability to access any web site. The web applications cannot easily figure out an ordinary traffic or an attack. These attacks can be applied by consuming computational resources (such as memory, bandwidth, processor time, or disk space), by disrupting configuration information (such as routing information, by disrupting physical network components), by disrupting state information (such as resetting of TCP sessions), and by obstructing the communication media between the victim and the intended user for their inadequate communication.

1.3 Genetic Algorithms
Genetic Algorithms were first proposed by John Holland in1975.Genetic Algorithms have biological background as they are based on Darwin theory of natural evolution [5]. This theory states that an individual who is fit among the population will survive and will reproduce to the next generation. Genetic algorithms can be applied in bioinformatics, engineering, economics, chemistry, computational science, etc.

1.3.1 Working of Genetic Algorithms

1.3.1.1 Encoding
The very first step for solving a problem using genetic algorithms includes encoding of solution. In this stage

Phenotype is mapped to genotype i.e. data is represented in genes.

1.3.1.2 Initialization

This step includes input parameter like population size, crossover probability, mutation probability, and number of generation for iteration.

1.3.1.3 Evaluation

In this step, the fitness value of each individual is calculated through fitness function.

1.3.1.4 Selection

This step eliminates bad population and selects the best fit individuals.

1.3.1.5 Crossover

Crossover is also called recombination.

1.3.1.6 Mutation

Mutation is done to add new features from outside

1.3.1..7 Iterate steps 3 to 6 until terminate the loop.

1.3.1.8 Decoding

In this step, the final solution is decoded back to phenotype.

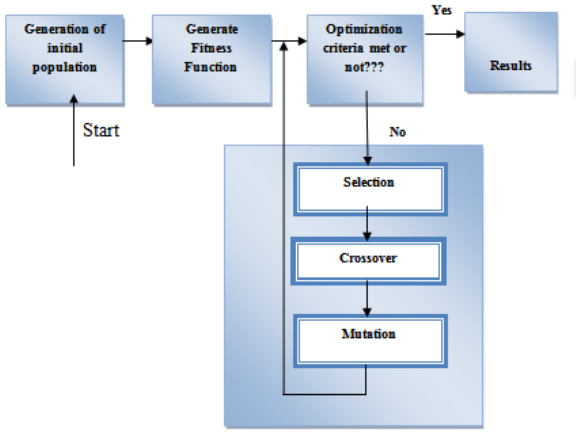

Fig 1: Concept of Genetic Algorithms

1.4 Neural Networks

Neural Networks are learning based algorithms which are used for optimization purpose. An Artificial Neural Network (ANN) is inspired by biological nervous system. An ANN can be functional in applications such as pattern recognition or data classification. Artificial Neural Networks have the advantage of computational speed and easy representation of elation between input and output.

Neural Networks consists of following components:

1.4.1 It consists of a directed graph which is also called as network topology. Arcs of these graphs represent links.

1.4.2 A state variable which is associated with every node.

1.4.3 Links associated by a real-valued weight.

1.4.4 Nodes associated with a real-valued bias.

1.4.5 A transfer function is assigned to each node that determines the state of a node as a function of bias, the weights (wi) of incoming links and the states (x,,) of the nodes connected by these links.

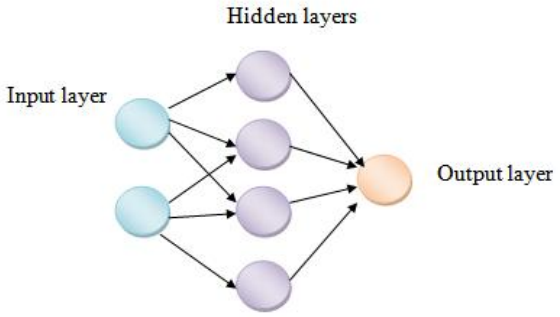

Fig 2: A simple Neural Network

2. RELATED WORK

The focus of related work is to study the possible attacks on web log files and the existing tools and techniques to identify and prevent these attacks.

Internet is playing an important role in our day-to-day life. It has become impossible to survive without it. Almost all the traditional services such as banking, education, transportation are running through online applications. The performance of these websites is very important both to the users and the website owners. Web log files are maintained to keep track of the user's interest which may help in increasing the performance of websites. Chintan R. Varnagar et al. discussed the work done so far on web usage mining [7]. As the web log files contain noisy data, preprocessing is very important to access the required information.

Web applications form a worldwide podium for network services. But these web applications suffer from various attacks. In 2012, Yi Xie et al. discussed one of the most common attack i.e. Distributed Denial of Service attack (DDoS). The authors proposed a new semi-Markov model to prevent the DDoS attack [8]. This model was used to capture the characteristics of the users that vary with time. In 2012, G.Swapna et al. proposed a new system which can be used to detect internet attacks and insider attacks by analyzing the web server log files [9]. With this system, the authors tried to protect the websites from SQL injection attacks.

Roger Meyer in his work detected various types of attacks on web log files in 2008 [10]. According to the author, there are two attack detection methods – rule based detection (static) and anomaly based detection (dynamic). The author discussed 9 different attacks on web log files such as: Cross site scripting, malicious file execution, insecure direct object reference, cross-site request forgery, etc. To detect these attacks, attack vectors have to be known in order to make detection rules and various standards and encoding variants needs to be known. Concerning intrusion detection in web log files, Shaimaa Ezzat Salama et al. [11] proposed a new method to track the attacks by combining the different log file formats into one XML format. The authors focused on the pre-processing steps of web log files and discussed the difference between preprocessing of web log files for web usage mining and web intrusion detection. For this, the authors proposed two algorithms using c#.

Distributed Denial of Service (DDoS) attacks pose a serious threat to web applications. Yahoo, Google, Amazon are some of the companies that totally dependent on Internet. In 2007, Stefan Seufert et al. [12] discussed the effectiveness of machine learning techniques to defend against Distributed

Denial of Service attack. The authors conducted a series of experiment to ensure the capabilities of machine learning techniques. Websites that have password based authentication systems are more vulnerable to brute force attacks. Carlisle Adams et al. proposed a non-intrusive and simple security mechanism against brute force attacks using the concept of sliding window [13]. This new approach needs requires only a small amount of modification to the application code.

The aim of password based systems is to hamper the unauthorized access. Though the passwords are necessary, still they are not safe as a large number of attacks are related with passwords. Mudassar Raza et al. discussed the various authentication methods [14]. The authors suggested that authentication methods should be selected as per the scenario. Different password schemes could be combined to form a single password scheme.

Dusan Stevanovic et al. worked on the methods for separating malicious and non malicious users as detecting malicious web crawlers is one of the most dynamic research areas in the field of network security [15]. The authors analyzed this problem using unsupervised neural network learning algorithms: Self Organizing maps (SOM) and Adaptive Resonance theory (ART2). As web log files provide an easy way for intrusion detection, log files strengthen the security systems by finding out the unsuspicious behavior of users. Bhagyashree Deokar et al. [16] proposed a new intrusion detection system which overcomes the drawbacks of existing systems. To achieve this, reinforcement learning, log correlation, association rule learning, and techniques were used. This proposed system reduced the false alarm rate and improved the ability to detect unknown attacks.

3. PROPOSED METHODOLOGY

Step1. Start the process of detection and prevention of brute force and DDoS attack by opening Matlab which Contains GUI's for both the processes.

Step2. Select the log file data both for Brute Force and DDoS in GUI's.

Step3. Initialize Brute Force and DDoS attack. After this step, both the attacks can be detected. In case of Brute force attack the intruder will try all the possible combinations of the password and in DDoS the ever becomes overloaded as the number of service requests will increase than the capacity of the server.

Step4. This step involves the generation of fitness function to identify the possible malicious records

Step5. This step distinguishes the malicious users and the non-malicious users. In case of DDoS attack if the threshold value of the web log file is more than or equal to threshold of neural, then it indicates that the user is malicious else the user is not malicious. In case of Brute Force attack detection, if the assessment value matches the value generated by the server matches with the value entered by the user, then only access is guaranteed otherwise access is denied.

Step6. After the detection and prevention of both the attacks, the process should be stopped and the values of the three parameters i.e. Accuracy, Recall and Precision are represented in a tabular form by repeating the same process for minimum 10 times.

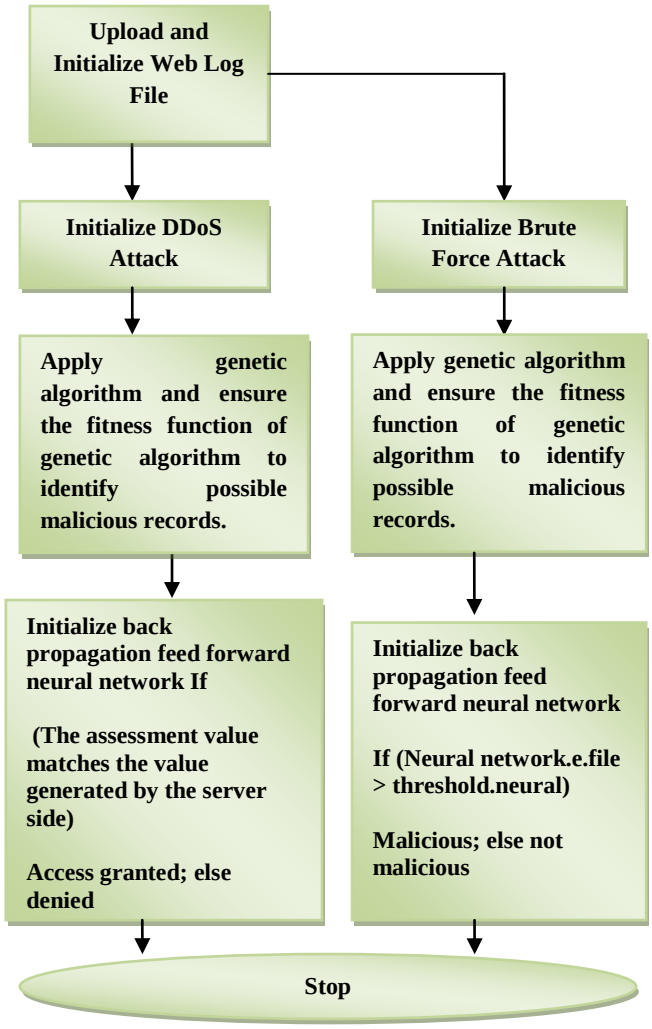

Fig 3: Methodology for DDoS and Brute Force attack prevention

4. CONCLUSION AND FUTURE SCOPE

This proposed system is basically a secure platform for sharing of files. Any user can create his/her profile on the system and then upload new data or download data that is already present on the system. The user can also request the admin for some specific data which the admin can later upload. Rule mapping can be applied at the server side to check for any abnormalities in the data being received or transmitted. This system basically concentrates on the log entries which deal with denial of service and brute force attacks using GA. This system provides a secure platform for sharing of files for individual users. As of now, whenever data is transmitted from the client side to the server side or vice-a-versa, it is done so in a plain format. In the future, this system could be improved by encrypting the data that is being transmitted on the network. As a result of this, in case of data leakage, the intruder would not be able to gain any important information

5. REFERENCES

[1] Brijesh Bakariya, Krishna K. Mohbey, G. S. Thakur, "An Inclusive Survey on Data Preprocessing Methods Used in Web Usage Mining", Proceedings of Seventh International Conference on Bio-Inspired Computing : Theories and Applications, Springer India, 2013.

[2] L.K. Joshila Grace, V.Maheswari, Dhinaharan Nagamalai, "Analysis of Web Logs and Web User in Web Mining", International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1, January 2011.

[3] Adams, C., Jourdan, G. V., Levac, J. P., & Prevost, F. (2010, August), "Lightweight protection against brute force login attacks on Web applications", In PST (pp. 181-188).

[4] Mr. Vijay, D. Katkar, Ms. Deepti, S. Bhatia, "Experiments on Detection of Denial of Service Attacks using REPTree", IEEE, 2014.

[5] Gopesh Joshi, "Review of Genetic Algorithm: An Optimization Technique", International Journal of Advanced Research in Computer Science and Software Engineering", Volume 4, Issue 4, April 2014.

[6] Sonali B. Maind, Ms. Priyanka Wankar, "Research Paper on Basic of Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication", Volume 2, January 2014.

[7] Chintan R. Varnagar, Nirali N. Madhak, Trupti M. Kodinariya, Jayesh N. Rathod, "Web Usage Mining: A Review on Process, Methods and Techniques", IEEE, 2013.

[8] Xie, Yi, Shensheng Tang, "Online anomaly detection based on web usage mining", Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International, IEEE, 2012.

[9] G.Swapna, R.Pavani Srivatsav, "Securing Web Applications By Analyzing The Logs Of The Database Server Or Web Server", International Journal of Engineering Research and Applications, Vol. 2, Issue 6, November- December 2012.

[10] Roger Meyer, "Detecting Attacks on Web Applications from Log files", SANS Institute, 2008.

[11] Shaimaa Ezzat Salama, Mohamed I. Marie, Laila M. El-Fangary, Yehia K. Helmy, "Web Server Logs Preprocessing for Web Intrusion Detection", Canadian Center of Science and Education, Vol. 4, July, 2011.

[12] Stefan Seufert, Darragh O'Brien, "Machine Learning for Automatic Defence against Distributed Denial of Service Attacks", IEEE, 2007.

[13] Carlisle Adams, Jean-Pierre Levac, François Prevost, "Lightweight protection against brute force login attacks on web applications", Eighth Annual International Conference on privacy, Security and Trust, IEEE, 2010.

[14] Mudassar Raza, Muhammad Iqbal, Muhammad Sharif, Waqas Haider "A Survey of Password Attacks and Comparative Analysis on Methods for Secure Authentication, World Applied Sciences Journal 19 (4): 439-444, 2012.

[15] Dusan Stevanovic, Natalija Vlajic, Aijun An, "Detection of malicious and non-malicious website visitors using unsupervised neural network learning", Applied Soft Computing, Elsevier, 2012.

[16] Bhagyashree Deokar, Ambarish Hazarnis, "Intrusion Detection System using Log Files and Reinforcement Learning", International Journal of Computer Applications (0975 – 8887) Volume 45– No.19, May 2012.