www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 3 March 2015, Page No. 10815-10818

Survey on Threat management system

Aniket Tare[1], Mrunal Funde[2], Shraddha zade[3],Vaishnavi khati[4]

[1]Rajiv Gandhi college of Engg & Research
Nagpur

[2]Rajiv Gandhi college of Engg & Research
Nagpur

[3]Rajiv Gandhi college of Engg & Research
Nagpur

[4]Rajiv Gandhi college of Engg & Research
Nagpur

Abstract: — With the rapid change and development in the sector of Information Technology and in Network technologies; the value of data and information is also increased. Today lot of valuable data is generated using many computers based application and stored back to the company database. But unfortunately, the threat to the same data is also increasing rapidly. So, development of a proper Intrusion
Detection System which provides a right alarm is a hot topic today. There are many areas which helps to build such devices and software applications like Data mining techniques, network protocol system, decision tree, clustering, SNORT, Genetic Algorithm etc. This paper presents a technique of applying evolutionary algorithm i.e. Genetic Algorithm to Intrusion Detection System. It also provides a brief and how to implement it in real IDS.

Keywords: Data mining, DDOS, Evolutionary algorithm Genetic Algorithm, Intrusion, IDS, SNORT, Threats

1. Introduction

. The main problem with current intrusion detection systems is high rate of false alarms triggered off by attackers. Effective protecting the network against malicious attacks remains problem in both research and the computer network managing professionals. Improved monitoring of malicious attacks will require integration of multiple monitoring systems. A series of analytical and mathematical models are used to acquire potential benefits of multiple sensors for reducing false alarms. Today, the number of attacks against large computer systems or networks is growing at a rapid pace. When an intruder attempts to break into an information system or performs an action which is not allowed, we refer this activity as an intrusion. Intruders can be classified into two groups, external intrusion and internal intrusion. The former refers to those who do not have an authorized access to the system and who attack by using various penetration techniques. The latter refers to those with access permission who wish to perform unauthorized activities. An Intrusion Detection System is a system for detecting intrusions and reporting them accurately to the proper authority. In 1980, James Anderson first introduced the concept of Intrusion Detection. Since then, Intrusion detection techniques are considered as the second gate for providing networks security behind firewalls. The purposed of Intrusion Detection Systems (IDS) is designed to detect attacks against computer systems over insecure networks by this way that detects attempts by legitimate users to abuse their privileges or to exploit security vulnerabilities for comprising the computers. Existing IDS systems can be divided into two categories according to the detection approaches: anomaly detection and misuse detection or signature detection. Anomaly detection is also called as Behaviour detection. Anomaly detection is an approach to detect intrusions by first learning the characteristics of normal activity of users. Then the system uses such characteristics to judge whether the user's activity is normal or not. Misuse detection systems are the approach that tries to match user activity to stored signatures of known exploits or attacks. That is to say, such detection system uses a previous defined knowledge to check whether the new activity is in that knowledge database. If yes, the IDS considers this activity may be as a possible attack and then blocks it.

The central theme of this paper is to explore parameters and evolution process of Genetic Algorithm which helps to detect malicious packet on the network and ultimately helps to block the respective IP addresses. Genetic algorithm is an evolutionary algorithm which is helpful for search and optimization purpose. They incorporate the concept of Darwin's theory of survival. Many researchers have introduced the use of GA in intrusion Detection and reported very high success rates. We have used GA based approach to find and detect the malicious packets and IP addresses on the network. The main reason behind selecting GA for this task is due to inherent evolutionary treatment in the algorithm which allows us to define our own fitness function based on which only those members or rules are selected that satisfy our fitness criterion.

2. PROPOSED APPROACH

In this threat detection system to detect the infected IP addresses. We used two algorithm which help us to identify the infected IP address as the our project is divided in to the two parts one part

is identifying infected IP address and second part is to block that infected IP address

3. **Recent Information Security Technologies**

. Security researchers developed various security technologies to protect the system from evolving attacks. Typical solutions are firewall, WAF (Web Application Firewall), ESM (Enterprise Security Management), IPS/IDS.

4.FIREWALL

Firewall is a regulation device that controls the network traffic between separated networks and hosts. It is a security technology which is based on access control. It decides whether to grant an access to the internal IP addresses and port numbers. Administrator sets these access control rules at initial level. The firewall is located at border of the network and can be used as a defender for the internal network. Also, firewall is used as a primary security solution to this day. Firewall has a simple protocol system that allows administrators to control firewall easily. However, firewall fails to detect and analyse threats in the network but just blocks accesses according to IP addresses and port numbers defined by administrators. Therefore a firewall only provides limited protection from threat attacks.

5. Aggregation Algorithm –

Aggregating algorithm is a class of forecasting algorithm developed in the strand of machine learning known as prediction with expert advice. as during the use of internet we open many sites from which in it some of the sites are infected which sends the infected files to our system more than one time after accessing that site .in every computer system there is file which is could as log file which continuously keeps the record of each and every Ip address . the aggregation algorithm monitors this Ip address and predict Ip address which are continuously sending the files to our system

6.. Genetic Algorithm –

A. Introduction to Genetic Algorithm

Genetic algorithms are a branch of transmutative algorithms used in search and optimization techniques. The three dominant functions of a genetic algorithm i.e., selection, crossover and mutation correspond to the biological process: The survival of the fittest (As shown in Figure 1). In a genetic algorithm, there is a population of strings (called chromosomes or the genotype of the genome), which encode and indent solutions (called individuals, creatures, or phenotypes). Traditionally, solutions are been re-presented in binary as strings of 0s and 1s, but there is possibility of another encodings too. The beginning of evolution starts from a population of randomly generated individuals and evolves over generations.

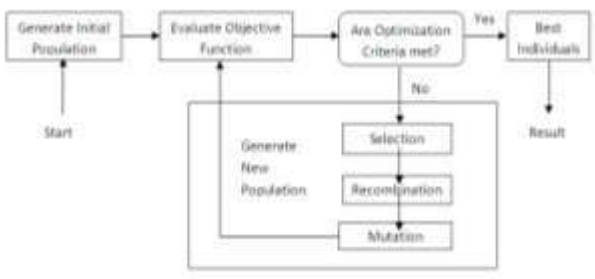

FIG: Genetic Algorithm

In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), & modified (recombined and possibly randomly mutated) to form a new population. The newly achieved population is then used in the next iteration of the algorithm. Generally, the algorithm gets terminated when either a maximum number of individuals are there in a generation, or a satisfactory fitness level has been reached for the population. If the algorithm is terminated due to maximum number of individuals, the solution may or may not be achieved.

B. Genetic Algorithm Process

GA evolves the population of chromosomes (individuals) as the process of natural selection. It generate(s) new chromosome(s) (offspring) during its process. GA process uses a set of genetic operators (selection, crossover and mutation), and evaluate chromosome using the fitness function.GA consists of population of chromosomes that reproduced over set of generations according to their fitness in an environment. Chromosomes with most fitness level are most likely to survive, mate, and bear children.GA terminate the process by define fixed maximal number of generations or as the attainment of an acceptable fitness level, or if there are no improvisations in the population for some fixed number of generations, or for any other reason. The standard GA processes is shown in figure. It contains various steps which include: encoding chromosomes, generating initial population, fitness function evaluation, and then applying one of the operators.

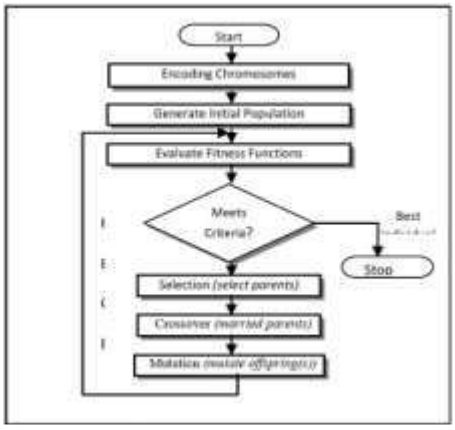

FIG: Gentic Algorithm Process

C. GA Operators

Encoding of the Chromosomes*:* In the GA process it is important to represent the data into some of the encoding formats. One outstanding problem associated with encoding is that some individuals correspond to infeasible or illegal solutions to a given problem. Various encoding methods have been created for particular problems to provide effective implementation of genetic algorithms. The encoding methods are classified as follows:

C.1 Binary Encoding

Binary encoding (i.e., the bit strings) are the most common encoding used for several of reasons. One is historical: in their earlier work, Holland and his students concentrated on such encodings and genetic algorithms practices have tended to follow this lead. Another reason for that was because much of existing GAs theories is based on the assumption of using binary encoding.

C.2 Real-number encoding

Real number encoding is best used for function optimization problems. It has been confirmed that the real number encoding performs better than binary encoding for function optimization and constrained optimizations problems. In real number encoding, the structure of genotype space is identical to that of the phenotype.
Therefore, its easy to form effective genetic operators by borrowing useful techniques from conventional methods.

C.3 Integer or literal permutation encoding

Integer or the literal permutation encoding is best used for combinational optimization problems because the essence of this kind of problems is to search for the best permutation or combination of items subject to constrains.

1) Applying fitness function: Fitness function (or objective function) defines the problem constraints; it measures the performance of all chromosomes in then population. Fitness function is the heart of all Genetic Processes. In our approach, we have used:

*Fitness = (size * weight)*

Where the size is the actual packet data size prescribed by the incoming packet data stream and weight is the vector which applied to each chromosome.

2) Selection operator: Selection Operator determines which chromosome(s) from the population will be chosen for recombination; depends on the fitness of the chromosome. The selected chromosomes are called parents. Selection methods are as follows:

- Fitness-proportion selection.
- Roulette-wheel selection.
- Rank selection.
- Local selection.
- Tournament selection
- Steady state selection

3) Crossover operator: The parent's chromosomes are recombined by one of the crossover methods. It produces one or more new chromosome(s) called offspring(s). Such methods are: Single Point Crossover, Multipoint Crossover, Uniform Crossover and Arithmetic Crossover.

4) Mutation operator*:* New genetic material could be introduced into the new population through mutation process. This will increase the diversity in the population. For each offspring mutation randomly alters some gene(s). A commonly used method for mutation is called single point mutation. Though, a special mutation types used for varies problem kinds and encoding methods. So we are having Single point mutation and multi point mutation

D.SYSTEM OVERVIEW

The proposed system overview is shown in figure no. 3 which starts from capturing firewall entries i.e. firewall data sets and then initial filtering is done on the basis of rule defined by the system. This précised data is then input to the GA based algorithm the detail proposed architecture is shown in figure 4. It starts from initial population generation from pfirewall.log file generated by the firewall system. The packets are the filtered out on the basis of rules. Then the précised data packets go through several steps namely selection, crossover and mutation operation. These processes get generate best individuals. The generated individuals are the verified by the fitness function to generate the population for next generation.

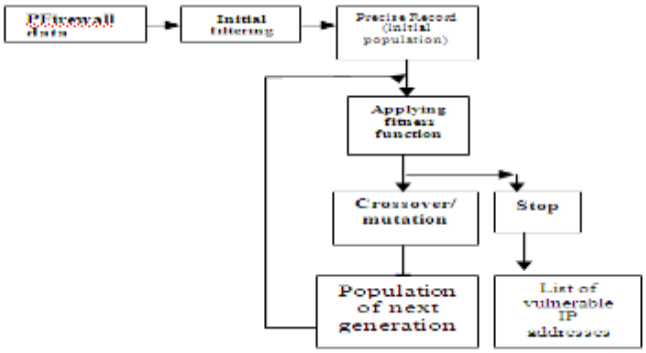

Fig: detail system Architecture

E. EXPERIMENTAL SET UP

A. OBJECTIVE-

The scope of experiment is focused to generate list of IP addresses and there packets which are vulnerable to the server or destined system. The testing is done on the entries generated by the firewall system of machine in pfirewall.log file. The training is done on the predefined data rules. The pfirewall.log file contains the entries of incoming packets with various fields like date/time, action, protocol, source port ,Destination port, source Ip, Destination Ip, size, flag, ack type and info. These entries are made available on firewall setting which are available for both successful connection and dropped packets. But for making the connection profile we have used only 5 important fields of it. These are source-Ip, Destination-ip, source-port, Destination-port and size. The size of pfirewall.log file may also vary with requirements.

B. TOOLS-

For this experiment we have used java as the frontend to make coding part and to write different algorithms and classes. The training data is stored into the wamp server which is used as the backend to the system. Wamp server is able to store the different structures of dataset tables. For this experiment we used windows based Dell computer with dual core processor system having 120 GB hard disk program.

RESULT-

From the above experiment, we have able to create a rule base that could successfully categories harmful and harmless connection types. We have shown the resultant figures below by applying 100 connection entries respectively to the proposed system. After that we were able to get around 95% of accuracy to classify the connections types.

CONCLUSIONS-

Recent unknown attacks easily bypass existing security solutions by using encryption and confusion. Therefore new detection methods for reacting to such attacks are in need. In this paper we have successfully evolved the rule set and profile of network connection which can detect existing as well as new intrusions. So now the system can be integrated with any of the IDS system to improve the efficiency and the performance of the same. The system can also be able to integrate to the input to the firewall system which can use the rule set defined and generated by the system to block Intrusion. In this paper, we have discussed the GA processes and evolution operators also discussed the overall implementation of GA into proposed system. The various operators like selection, crossover and mutation are also discussed.

REFERENCES-

[1] T. Lunt, A. Tamaru, F. Gilham, R. Jagannathan, P. Neumann, H. Javitz, A. Valdes, and T. Garvey. ―A real-time intrusion detection

expert system (IDES)‖ - final technical report. Technical report, Computer Science Laboratory, SRI International, Menlo Park, California, February 1992.
[2] K. Ilgun, R. A. Kemmerer, and P. A. Porras. ―State transition analysis: A rulebased intrusion detection approach‖. IEEE Transactions on Software Engineering, 21(3):181–199, March 1995
[3] John E. Dickerson, and Julie A. Dickerson ―Fuzzy Network Profiling for Intrusion Detection‖ Electrical and Computer Engineering
Department Iowa State University Ames, Iowa, 50011. [4] Rui Zhong, and Guangxue Yue ―DDoS Detection System Based on
Data Mining‖ ISBN 978-952-5726-09-1 (Print) Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)Jinggangshan, P. R. China,4,April.2010,pp.062-065.
[5] Dietrich, S., Long, N., and Dittrich, D. 2000. Analyzing distributed Denial of service attack tools: The shaft case. In Proceedings of 14th Systems Administration Conference. New Orleans, Louisiana, USA, 329-339.
[6] R. Magoulas and B. Lorica, "Introduction to Big Data", Release 2.0 (Sebastopol O'Reilly Media), Feb, 2009.
[7] P. Chapman. et al, "CRISP-DM 1.0 – Step-by-step data mining guide",http://www.crisp-dm.org (2000).