NETWORK INTRUSION DETECTION USING GENETIC ALGORITHM

Shrikant Vanve[1], Sarita Patil[2]
[1,2]Department of Computer Engineering,
G.H. Raisoni College of Engineering and Management Research Centre Wagholi, Pune
Email:shrkntvanve0@gmail.com[1], saritapatil555@gmail.com[2]

Abstract
As the use of internet is increasing, many times network becomes victim of various types of attack. To prevent this attacks network intrusion detection is playing important role. In this paper new approach of genetic algorithm based network intrusion detection is designed in which network rules are formulated to classify normal activity and malicious behavior of network. Due to use of genetic algorithm best rules are evaluated to yield better result and it increases the accuracy of system.

Keywords: Genetic algorithm , KDD CUP 99 Data set , network intrusion detection.

I. INTRODUCTION

Now a days in many industries data and network security is major issue considered at high priority level. Because many times network is accessed by unauthorized user which results in security of organizational database. Currently network intrusion detection systems are used to avoid such attack. This paper presents how genetic algorithm based network intrusion detection system are used to avoid attacks. Intrusion detection system is a system that keeps watch on events occurring in computer system or in network. Basically intruders are of two types. External intruders and internal intruders. External intruders are those who do not have access to system but they try to access the network or system whereas internal intruders are those who have access to system and try to behave as abnormal user. Generally network intrusion detection systems are classified into two types. 1)misuse based intrusion detection - In this type known patterns of attacks are compared with abnormal connections to find attacks. The main drawback of this type is that it considers normal connections when any new attack comes into network besides the database stored in known patterns of attacks. 2) Anomaly based intrusion detection - In such system any abnormal behavior of network which deviates from normal connections are considered as intrusions. Drawback of this system is that many times normal connections are considered as attacks. Basically there are four types of attacks which are as follows

Denial of Service Attack (DoS): In this type hacker makes the resources so busy that no any request gets executed and denies user from accessing the system or resources. e.g. smurf, Neptune, ping of death.

Remote to User Attack (R2L) : R2L is attack in which user send packet to machine which is not accessible for user and expose the vulnerabilities and exploits the privileges of local user. e.g. guest, phf, xnsnoop.

User to root attacks (U2R): In this type of attack hacker tries to get super user privileges. e.g. Perl, xterm.

Probing : Probing is attack in which hacker scans the machine and find out the weaknesses of system e.g. saint, portsweep, mscan.

This paper explores how genetic algorithm is used to detect intrusions in the network. To implement genetic algorithm effectively in the system KDD CUP 99 Dataset is used. This Training and testing dataset is provided by MIT Lincoln Lab to further study on network intrusion detection using genetic algorithm.

II. RELATED WORK

In 1995 Crossbie and Spafford used Multiple agent technology using Genetic Programming to detect attacks in the network. Each agent was used to monitor the network behavior in this system. Main drawback of this technology is the communication between different agents and training to each agent is time consuming.

L me introduced new genetic algorithm technique for misuse detection called Genetic algorithm for simplified security audit trials analysis (GASSATA).In this technique 2 dimensional matrix is constructed which represents the patterns of intrusions. Thus genetic algorithm is used to find the attacks appearing in the audit record.

Deber at al proposed Neural Network model to diagnose any deviation from normal behaviour in network intrusion detection system. Liu at al described NIDS using Neural Network. According to that system, the NNs are used to classify without consulting a domain expert; hence, this automation helped to detect both known and novel intrusions. The key part of the work was focused on the development of an adaptive resonance theory (ART) NN, and it is trained in real-time in an unsupervised way.

Xiao et al. Proposed a technique which uses information theory and genetic algorithm to find abnormal network behaviour. Based on the mutual information between network features and the types of network intrusions, a small number of network features are closely identified with network attacks. Then a linear structure rule is derived using the selected features and a GA. The use of mutual information reduces the complexity of GA, and the single resulting linear rule makes intrusion detection efficient in real-time environment. However, the approach considers only discrete features.

III. INTRODUCTION TO GENETIC ALGORITHM

Genetic Algorithm is heuristic search technique which uses natural selection, cross over, mutation operators. According to Darwin's survival of the fittest theory, those individual having best characters can survive and characters of best individuals are passed onto next generation. This process continues till number of generations. The same theory is used in network intrusion detection using genetic algorithm technique. Now we will focus on genetic operators.

A. Natural Selection

In genetic algorithm certain number of individuals are firstly initiated called population. Among this population n number of individuals are selected according to fitness function. This selection process obeys Darwin's survival of fittest theory.

B. Crossover

In genetic algorithm individuals selected through natural selection process are combined to produce new offspring. There are different crossovers techniques which are used to combine selected individuals like one point crossover, two points cross over, multi point cross over. Here individuals are nothing but the chromosomes.

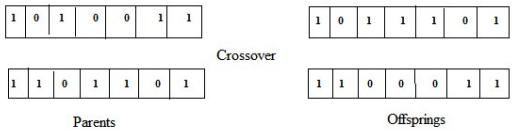

Fig. 1 Crossover

c. Mutation

Alteration in one or more gene values in chromosomes is called mutation. Mutation operator is used for maintaining genetic diversity among the individuals in population.

Fig. 2 Mutation

Steps involved in Genetic Algorithm:

1) Firstly the possible number of solutions is generated called population.

2) Then fitness of each solution is calculated which defines how much solution is better for given problem. Best fit individuals are crossover at single or multiple point.

3) After crossover if required mutation is performed so that genetically each individual will be different from each other.

4) This process continues till fixed number of generations. After this we will get best solution for given problem.

IV. APPLICATION OF GA TO NID

Genetic Algorithm is applied for evaluation of rules of network traffic. These rules are used to find any malicious behaviour of network from normal connections. We used KDD CUP 99 data set for generating these rules. This dataset contain 41 features. Out of these 41 features we

focus on 7 features for generating network rules which are involved in malicious activities in network. Numeric value of these features is taken into consideration for generating rules. The syntax for rules is as follows:

If (Condition == True)

Then (attack type == XYZ)

Here condition is formed by using 7 attributes of dataset. If the condition is true then only name of attack will follow the rule. Table 1.1 shows the list of attributes used for generating set of rules.

Name of Attribute	Description
Duration	Duration of connection
Protocol	Udp tcp icmp
Service	ftp http smtp
Source port	Source port number
Destination port	Destination port number
Src-ip	Ip address of source
Dst-ip	Ip address of destination
Type of attack	Name of attack

KDD CUP 99 Data set contain number of record having different attack types. Consider Neptune attack type of Denial of Service attack. Initially rule of this attack is defined as

If(duration == 0.0.1 and protocol == tcp and source port == 19829 and destination port == 80 and src-ip == 162.31.42.215 and dst-ip == 185.51.24.132 and service == http)

Then (type of attack == Neptune)

Above rule states that if duration is 1 second and packet is travelling from ip 162.31.42.215 port number 19829 to destination ip 185.51.24.132 and port number 80 via tcp protocol and type of service is http then connection is victim of Neptune attack. This rule is converted into number so that it becomes easy to convert it into binary system. The rule will look like as (0,0,1,1,19829,80,162,31,42,215,185,51,24,132 ,1,1). This whole string represents the chromosomes and each number in string is called genes. So it is clear that 4 genes are required to represent ip address of source and destination. In this way rules are initialised for each type of attack then genetic algorithm is applied for evaluating these rules to get best rules

The Denial-of-Service attack ”Smurf” if (duration=”0:0:1” and protocol=”finger” and source-port=19891 and destination-port=79 and source-ip=”9.9.9.9” and destinationip=”172.16.112.50” and service=”http”) then (attackname=”smurf”) The above rule expresses that if a network packet is originated from IP address 9.9.9.9 and port 18982 , and sent to IP address 172.16.112.50 and port 79 using the protocol finger, and the connection duration is 1second, then most likely it is a network attack of type smurf that may eventually cause the destination host out of service.

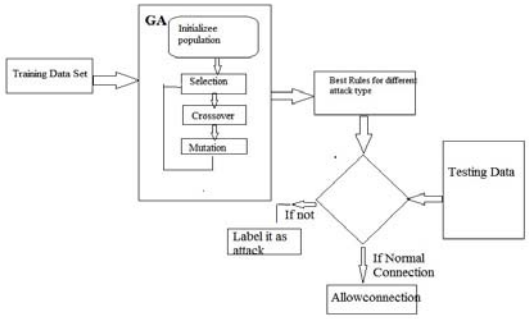

Fig. Architecture of GEdIDS

Fitness Function : To determine the fitness of each rule support confidence framework[5] is used. The rule is represented as ” if x then y”. Fitness of each rule is calculated by using following equations.

Support = | X and Y | / N

Confidence = | X and Y | / | X |

Fitness = A1 * Support + A2 * Confidence

Here N = total number of connections

| X | = total number of connections matching condition X

| X and Y | = total number of connections matching condition if X then Y

A1 and A2 are called threshold values.

V. SYSTEM PERFORMANCE AND MEASURES

As per the requirement we are using KDD Cup 99 Data set which contain records. These records are used as input. The Following Table shows the distribution of training and testing data set of KDD Cup 99

To determine the performance of system two performance measures are considered. Which are as follow.

Datatype	Normal	Dos	R2L	Probe	TOTAL
Trainingdata	97277	391458	1126	4107	494020
TestingData	60593	229853	16189	4166	311029

$$FalsePositiveRate = \frac{FalseAlarm}{TrueNegative + FalseAlarm}$$

$$DetectionRate = \frac{TruePositive}{TruePositive + FalseNegative}$$

The meaning of False Positive rate and Detection rate is defined in the following table

	Predicted Class Label	
Actual Class Label	Normal	Intrusion
Normal	True Negative	False Alarm
Intrusion	False Negative	True positive

VI. CONCLUSION

In this way the application of genetic algorithm for detecting intrusions in the network aims best results with minimum false positive rate. This approach is targeting to yield 97 percent correct result. In this approach different combinations of attribute are tested to initialise the population so that best rule are generated which are useful for differentiating malicious activity from normal connection against testing data set.

References :

[1]. M. Crosbie, G. Spafford, Applying genetic programming techniques to intrusion detection.In Proceedings of the AAAI Fall Symposium Series (AAAI Press, Nov 1995).

[2] L. M. GASSATA, ”A Genetic Algorithm as an
Alternative Tool for Security Audit Trails Analysis,” in Proceedings of the 1st International Workshop on the Recent Advances in Intrusion Detection (RAID 98), Belgium, 1998.

[3] T. Xia, G. Qu, S. Hariri, M. Yousif, ”An Efficient Network Intrusion Detection Method Based on Information Theory and Genetic Algorithm”,Proceedings of the 24th IEEE International Performance Computing and Communications Conference (IPCCC ’05), Phoenix, AZ, USA.2005.

[4] The third international knowledge discovery and data mining tools competition dataset kdd99-cup, in http://www.kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (1999)

[5] Middlemiss M and Dick G, ”Feature Selection of Intrusion detection data using a hybrid genetic of hybrid
Intelligent systems, IOS Press Amsterdam, PP.519-527, 2003.

[6] Chittur A. ”Model Generation for an Intrusion Detection System Using Genetic Algorithms, publications/gaids-thesis01.pdf, accessed in 2006.

[7] B. Uppalaiah, 2K. Anand, 3B. Narsimha, 4S. Swaraj, 5T. Bharat,”Genetic Algorithm Approach to Intrusion
Detection System”, IJCST Vol. 3, Issue 1, Jan. - March 2012.

[8] Gianluigi Folino o Clara Pizzuti oGiandomenico Spezzano, ”An ensemble-based evolutionary ramework for coping with distributed intrusion detection” Genet Program Evolvable Mach (2010) 11:131-146 DOI 10.1007/s10710010-9101-6

[9] B. Addullah, I. Abd-alghafar, Gouda I. Salama, A. Adbalhafez,”Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System”, ASAT-13-CE-14, May, pp. 26-28, 2009.