Composing Signatures for Misuse Intrusion Detection System Using Genetic Algorithm in an Offline Environment

Mayank Kumar Goyal and Alok Aggarwal

Dept. of CSE/IT, JIIT University, Noida, UP, India
mayankrkgit@gmail.com, alok.aggarwal@jiit.ac.in

Abstract. In recent years Internet has experienced a rapid expansion and also facing increased no. of security threats. However many technological innovations have been proposed for information assurance but still protection of computer systems has been difficult. With the rapid growth of Internet technology, a high level of security is needed for keeping the data resources and equipments secure. In this context intrusion detection (ID) has become an important area of research since building a system with no vulnerabilities has not been technically feasible.

In this paper, a Genetic Algorithm based approach is presented for network misuse intrusion detection system. The proposed genetic algorithm uses a set of classification rules which are generated from a predefined intrusion behavior. From the results it could be concluded that by applying proposed rule based network intrusion detection algorithm, more no. of intrusions can be detected.

Keywords: Genetic algorithm, misuse intrusion detection, information assurance, data set.

1 Introduction

Computers store, process and retrieve the data. Data is an invaluable asset for every organization, company, enterprise or even for an individual. Availability, integrity and confidentiality are the most important requirements for data handling. Earlier computers were isolated and usually not connected with other computers and does not have a modem. During those days the most common attack to the data stored in computers was the physical use of computer system. Thus in those days security of the room where computer system is placed was enough to secure the data. But with the growth of Internet during the last one decade gave many issues to security of data. Now-a-days computer break-in and misuse has become a common feature.

Intrusion is an activity performed by a person by breaking into an information system or performing an illegal action. Such person is termed as an intruder [1]. Intruders can be classified as external and internal. External intruders are the person

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 151–157.
springerlink.com

who do not have authorized access to the system and who tries to access the system illegally by using different saturation methods. Internal intruders are the persons who have authorized access to the system but carry out the illegal/unauthorized activities. Different methods have been used by intruders whether internal or external for intrusion like password cracking, software bug exploitation, mis-configuration of the system, sniffing unsecured traffic, utilizing the specific protocols design flow etc. No. of such attacks have been increased exponentially during the last one decade.

The two generally accepted categories of intrusion detection are misuse detection and anomaly detection. Former refers to techniques that characterize known methods for penetration into the system which are characterized as a ‘pattern’ or a ‘signature’ that the intrusion detection systems look for. These signature or pattern can be a static string or a set sequence of actions. Later refers to techniques that characterize normal or acceptable behaviors of the system like CPU utilization, job execution time, system calls etc. Behaviors that deviate from the expected normal behavior are considered intrusions [2].

Genetic algorithm has been used by many researchers in different types of intrusion detection systems. Genetic algorithms apply biological evolution theory to computer systems [3,4,5]. Genetic algorithm is a method of data analysis and can be termed as analogous to Darwinian evolution [6]. These are research techniques used in computer science and are implemented as a computer simulation and the approximate to combinatorial optimization problems [3].

In this paper, a genetic algorithm based approach is presented for network misuse intrusion detection system. The proposed genetic algorithm uses a set of classification rules which are generated from a predefined intrusion behavior. By applying proposed rule based network intrusion detection algorithm, more no. of intrusions can be detected. This paper is organized as follows. Section 2 describes genetic algorithms, section 3 gives genetic algorithm based intrusion detection. In section 4, different operators of Genetic algorithm used in the proposed algorithm are presented. Section 5 describes the proposed algorithm and experimental assumptions. Finally, results and discussions are given in section 6.

2 Genetic Algorithms

Genetic algorithms are computerized search and optimization methods that work very similar to the principles of natural evolution. Based on Darwin's survival-of-the-fittest principles, GA's intelligent search procedure finds the best and fittest design solutions which are otherwise difficult to find using other techniques. Metaphor from biology is used in genetic algorithms and genetics are used to iteratively evolve a population of initial individuals to a population of high quality individuals [14]. Here each individuals is composed of a fixed number of genes and represents a solution of the problem to be solved. The implementation of Genetic algorithms starts with a population of randomly selected chromosomes. The chromosomes which represent a better solution to target problem are given more opportunities to reproduce in comparison to those provide poorer solutions. The fitness of a solution is typically defined with respect to the current population. These chromosomes are representations of the problem to be solved. According to the attributes of the problem different positions of each chromosome are encoded as numbers. These positions are referred to as genes and are changed randomly within a range during evolution. The set of chromosomes during a stage of evolution are called a population. An evaluation function is used to calculate the fitness of each chromosome.

During the process of evaluation, two basic operators, crossover and mutation, are used to simulate the natural reproduction and mutation of genes. The selection of chromosomes for survival and combination is biased towards the fittest chromosomes. Genetic algorithm begins with a randomly generated population, evolves through selection, crossover, and mutation. Finally, the best chromosome is picked up as the final result. The working principle of a Genetic Algorithm is illustrated in figure1.

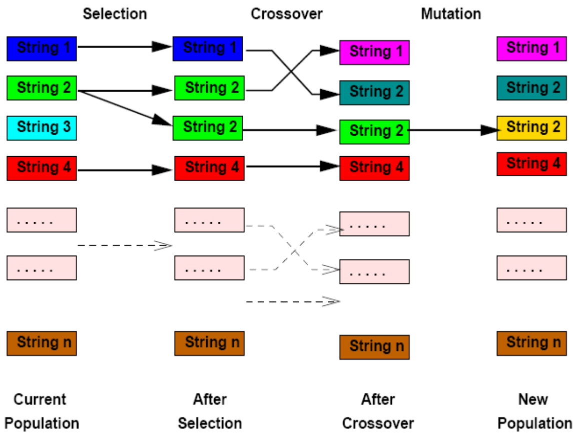

Fig. 1. Working principal of Genetic Algorithm

3 Genetic Algorithm Based Intrusion Detection

The rules stored in the rule base are in the following form:

If (condition) then (do)

The condition usually refers to a match between current network connection and the rules in intrusion detection system, such as source and destination IP addresses and port numbers, protocol, no of bytes of data sent by sender and responder indicating the probability of an intrusion. Several network features have higher possibilities that can be put in network intrusion detection identification. In our approach, six out of those features are taken to compose a classification rule. Table 1 shows the features. The first column in table 1 represents the feature name and second column represents no. of genes.

Table 1. Features and No. of Genes

Feature Name	No. of Genes
Source IP	4
Destination IP	4
Destination port	1
Protocol	1
Sender data amount	1
Responder data amount	1

For example, a rule can be defined as:

if (source-ip=9.9.12.19 and destination-ip=172.16.115.50 and destination-port=79 and protocol=http and sender byte=15000 and responder byte=15000) then (intrusion=attack A).

Each rule is encoded as a chromosome where each network features is encoded using one or more genes of different types.

The final goal of applying genetic algorithm is to generate rules that match only the anomalous connections.

4 Genetic Algorithm Operators

Encoding chromosomes and genetic operators are as follows:

Encoding Chromosomes: Encoding chromosomes in genetic algorithm means represents the set of events to the problem in one string of values.

Binary Encoding: Represent chromosome gene in binary numbers (0’s and 1’s).

Chromosome M	101010101010101011111
Chromosome N	010101111110000111100

Permutation Encoding: Represent chromosome gene in integer numbers.

Chromosome M	1 3 5 2 4 6 8 9 7
Chromosome N	9 4 1 3 2 7 8 5 6

Fitness Function: Fitness function measures the performance of all chromosomes in the population. To determine the fitness of a rule, the support and confidence framework is used. If a rule is represented as if X then Y, then the fitness of the rule is determined using the function of java genetic algorithm package in java.

Support = |X and Y| / Z

Confidence = |X and Y| / |X|

Here, Z is the total number of network connections in the audit data, |X| stands for the number of network connections matching the condition X, and |X and Y| is the number of network connections that matches the rule if X then Y.

Selection: The application of the fitness criterion to choose which individuals from a population will go on to reproduce.

Crossover: The parent's chromosomes are recombined by one of the crossover methods. It produces one or more new chromosome(s). A crossover operator is used to recombine two strings to get a better string. In crossover operation, recombination process creates different individuals in the successive generations by combining material from two individuals of the previous generation.. It takes two chromosomes and cut their strings at some randomly chosen position and swaps the tail positions.

Mutation: New genetic material is introduced into the new population through mutation process. This will increase the diversity in the population. It is an operator that introduces diversity in the population whenever the population tends to become homogeneous.

```
01101011110110
    |
01111011110110
```

5 Algorithm

Rule set generation using genetic algorithm.

Input : Population size, number of generations
Output : A set of classification rules

1. Initialize the blacklisted classification rules (population)
2. Initialize the population

```
3. N= total no of records in rule set
4. For each chromosome in the population
   4.1 calculate the fitness
5. Select the chromosome into new population
6. For each chromosome in the population
   6.1 Apply a crossover rate of 10 to the chromosome
   6.2 Apply a mutation rate of 1/50 to the chromosome
7. If number of generations is not reached, go to line 4
```

The training process begins by randomly generating an initial population of rules (Step 2). Step 3 calculates the total number of records. Step 4 calculates the fitness of each rule. Step 5 selects the best chromosome. Step 6 applies the crossover and mutation operators to each rule in the new population. Finally, step 7 checks and decides whether to terminate the process or to enter the next generation to continue.

6 Results and Discussions

Table 2 represents the new generated classification rules for signature based intrusion detection system using genetic algorithm which provide if implemented in existing intrusion detection system provide more efficient intrusion detection system. Obtained results are shown in Table 2.

Table 2. Newly generated classification rules for misuse intrusion detection system using genetic algorithm

Chromsome	Evolution time	Fitness value	Source IP	Destination IP	Port no	Protocol	Originator byte	Responder byte
1	1697	2526	125.15.34.137	119.127.190.239	52077	DNS	70141	697519
2	1505	26	125.119.15.58	119.141.38.12	313	SMTP	146869	431444
3	1388	26	125.81.1.146	119.81.43.93	34346	DHCP	59736	42788
4	1380	26	125.51.243.187	119.76.82.40	43980	LDAP	237241	454894
5	1399	26	125.79.240.135	119.129.132.234	51715	FTP	195534	538333

7 Conclusion

A method of genetic algorithm of rule base generation for misuse intrusion detection system is presented in this paper. Experiments have been carried out using a predefined dataset. The major advantage of using a genetic algorithm comes from the fact that in the real world the types of intrusions are dynamic. The proposed system can develop new rules to the systems so as the new intrusions become known. Therefore, it is adaptive and cost effective. Genetic algorithms are potential solutions for optimized rules sets and the determination of potential and actual network intrusions. If only mutation is used, the algorithm is very slow. Crossover makes the algorithm significantly faster.

References

[1] Satya Keerthi, N.V.L.C., Prasanna, P.L., Priscilla, B.M.: Ïntrusion Detection system Using Genetic Algorithm. Int. Journal of P2P Network Trends and Technology 1(2), 1–7 (2011)

[2] Jiang, M., Munawar, M., Reidemeister, T., Ward, P.: Efficient Fault Detection and Diagnosis in Complex Software Systems with Information–Theoretic Monitoring. IEEE Trans. on Dependable and Secure Computing (99) (2011)

[3] Owais, S.S.J., Krömer, P., Snášel, V.: Implementing GP on Optimizing Boolean and Extended Boolean Queries in IRs with Respect to Users Profiles. In: Proc. IEEE ICCES 2006, Egypt, pp. 412–417 (2006)

[4] Owais, S.: Optimization of Boolean Queries in Information Retrieval Systems Using GAs-Genetic Programming and Fuzzy Logic. In: CSIT 2006, Jordan, vol. 2, pp. 303–314 (2006)

[5] Owais, S., Krömer, P., Snašel, V.: Query Optimization by Genetic Algorithms. In: DATESO, pp. 125–137 (2005) ISBN: 80-01-03204-3

[6] Koza, J.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press (1992)

[7] Zhao, J.L., Zhao, J.F., Li, J.J.: Intrusion Detection Based on Clustering Genetic Algorithm. In: Proc. Int. Conf. on Machine Learning and Cybernetics, vol. 6, pp. 3911–3914 (2005)

[8] Diaz-Gomez, P.A., Hougen, D.F.: Three Approaches to Intrusion Detection Analysis and Enhancements. In: Proc. VI National Computer and Information Security Conference ACIS, Colombia (2006)

[9] Li, W.: Using Genetic Algorithm for Network Intrusion Detection. In: Proc. of the United States Department of Energy Cyber Security Group (2004)

[10] Gong, R.H., Zulkernine, M., Abolmaesumi, P.: A software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection. In: Proc. Int. Workshop on Self-Assembling Wireless Networks, pp. 246–253 (2005)

[11] Chen, Y., Abraham, A., Yang, B.: Hybrid Flexible Neural-Tree-Based Intrusion Detection Systems. International Journal of Intelligent Systems 22, 337–352 (2007)

[12] Abraham, Grosan, C.: Evolving Intrusion Detection Systems. Studies in Computational Intelligence (SCI) 13, 57–79 (2006)

[13] Sinclair, L.P., Matzner, S.: An Application of Machine Learning to Network Intrusion Detection. In: Proc. 15th Annual Conf. on Computer Security Applications (ACSAC), pp. 371–377 (1999)

[14] Pohlheim, H.: Genetic and Evolutionary Algorithms: Principles, Meth-ods and Algorithms, http://www.geatbx.com/docu/index.html