ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

Intrusion Detection System using Fuzzy Genetic Approach

B.Ben Sujitha[1], R.Roja Ramani[2], Parameswari[3]

Professor, Department of IT, Ponjesly College of Engineering, Nagercoil, Tamilnadu, India[1]

Assistant Professor, Department of IT, Sethu Institute of Technology, Kariapatti, Tamilnadu, India[2, 3]

Abstract : Network security is of primary concerned now days for large organizations. The intrusion detection systems (IDS) are becoming indispensable for effective protection against attacks that are constantly changing in magnitude and complexity. With data integrity, confidentiality and availability, they must be reliable, easy to manage and with low maintenance cost. Various modifications are being applied to IDS regularly to detect new attacks and handle them. This paper proposes a fuzzy genetic algorithm (FGA) for intrusion detection. The FGA system is a fuzzy classifier, whose knowledge base is modeled as a fuzzy rule such as "if-then" and improved by a genetic algorithm. The method is tested on the benchmark KDD'99 intrusion dataset and compared with other existing techniques available in the literature. The results are encouraging and demonstrate the benefits of the proposed approach.

Keywords- genetic algorithm, fuzzy logic, classification, intrusion detection, DARPA data set

I. INTRODUCTION

Today's network security infrastructure promisingly depends upon Network Intrusion Detection System (NIDS). NIDS provides safety from known intrusion attacks. It is not possible to stop intrusion attacks, so organizations need to be ready to handle them. IDS is a defensive mechanism whose primary purpose is to keep work going on considering all possible attacks on a system. Intrusion detection is a process used to detect suspicious activity both at network and host level. Two main ID techniques available are anomaly detection and misuse detection.

The anomaly detection model describes the usual behavior of a user to detect this user's anomalous or unaccustomed action. Among methods proposed to construct profiles, we mention: the statistical methods where the profile is calculated from variables taken randomly and sampled at regular intervals [1]. These variables can be, for example, the number of connections, the number of erroneous passwords, etc. The expert systems [3] and neural networks [2] are two well-known methods used to calculate a user profile.

The misuse detection model defines some anomalous behavior to analyze data susceptible to be attacks. The approach often uses known attacks called signatures. Among these methods, we mention: the expert systems [5], the genetic algorithm [4] and the pattern matching method that provides signatures of attacks. Various algorithms are used to localize these signatures in the audit trail [6]. Recently, several systems have been built to detect intrusions [7]. Various techniques have been applied extensively for intrusion detection such as agents-based detection intrusion [8] which can provide many advantages for the existing solutions due to the mobility of agents and their cooperative aspects, the Data mining approaches [9], the clustering techniques [10] and the fuzzy evolutionary algorithms [11]. Fuzzy logic [12] is an intelligent method that has been successfully employed for many IDSs. In this work, we focus on fuzzy genetic algorithms for intrusion detection. The methodology is a combination of the genetic algorithm with the fuzzy logic concepts. Genetic algorithms provide a natural tool to solve several problems in the field of applied

mathematics and science in general. Thus by combining genetic algorithms with fuzzy logic formalism we obtain complete and consistent enough for the acquisition, representation and use of knowledge by computers. We used

the concept of fuzzy logic in solving the problem of intrusion detection because fuzzy logic is an effective tool for introducing the concept of membership degree that

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

determines the "strength" in which an object belongs to different classes.

II. DATA-PREPROCESSING AND NORMALIZATION

The dataset used in the experimental study of this work are those of KDD'99 [13].As shown in TABLE I, the KDD99 dataset contains 22 different attack types which could be classified into four main categories namely Denial of Service (DOS), Remote to User (R2L), User to Root (U2R) and Probing. The full DARPA dataset contains 4885950 lines of connections.

Main Attack Classes	22 Attacks Classes
Denial of Service (DOS)	back, land, _eptune, pod, smurt, teardrop
User to Root (U2R)	buffer_overflow,perl, loadmodule, rootkit
Remote to User (R2L	ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster
Probing	ipsweep, nmap, portsweep, satan

Table 1 Type of Attacks in KDD'99 dataset

Each line of the KDD'99 dataset called "connection" includes a set of 41 features and a label which specifies the status of connection as either normal or specific attack type. The features of a connection include the duration of the connection, the type of the protocol (TCP, UDP, etc), the network service (http, telnet, etc), the number of failed login attempts, and the service and so on. These features had all forms of continuous, discrete, and symbolic, with significantly varying ranges. Among the 41 attributes of the connection, we consider only sixteen significant attributes which are: *A8, A9, A10, A11, A13, A16, A17, A18, A19, A23, A24, A32, A33, A1, A5 and A6*. These attributes are normalized. The normalization formula given in (1) is

applied in order to set attribute numerical values in the range [0.0, 1.0].

$$X = \frac{X\text{-}MIN}{MAX\text{-}MIN} \quad \text{(1)}$$

Where X: is the numerical attribute value, MIN is the minimum value that the attribute X can get and MAX is the maximum one.

Significant attributes are the important ones that can help in classifying a connection correctly. After having analyzed the KDD'99 dataset, the MIN and MAX values of each significant attributes which we have selected and considered in the current work are given as Table 2

Attributes	Description	Range (Normalized Value)
A8	Numberof``wrong" fragments	[0.3]
A9	numberofurgent packets	[0,14]
A10	numberof``hot" indicators	[0.101]
A11	number of failed login attempts	[0.5]
A13	numberof ``compromised" conditions	[0.9]
A16	numberof``root" accesses	[0.7468]
A17	number of file creation operations	[0,100]
A18	numberof shell prompts	[0,5]
A19	number of operations on access control files	[0.9]
A23	number of connections to the same host as the current connection in the past two seconds	[0.511]
A24	number of connections to the same service as the current connection in the past two seconds	[0.511]
A32	number of connection to the same host	[0,255]
A33	number of connection to the same serves for the host	[0,255]
A1	duration is number of seconds of the connection	[0. 58329]
A5	number of data bytes from source to destination	[0.1.3 one billion]
A6	number of data bytes from destination to source	[0. 1.3 one billion]

Table 2 : Significant attributes and its value

However, for the numerical attributes A1, A5 and A6, we have observed a big value of MAX hence the need to modify the normalization formula given in (1). The logarithmic scaling (with base 10) is applied to these features to reduce the range. We used all the sixteen features as the inputs of our Local fuzzy classifier.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

III. FUZZY GENETIC ALGORITHM

Genetic algorithms [11] employ metaphor from biology and genetics to iteratively evolve a population of initial individuals to a population of high quality individuals, where each individual represents a solution of the problem to be solved and is compose of a fixed number of genes. When genetic algorithm is used for problem solving, three factors will have impact on the effectiveness of the algorithm, they are [11]:

a. The selection of fitness function
b. The representation of individuals and
c. The values of the genetic parameters

Genetic algorithm is used for evolving new rules for IDS. Using these rules normal network traffic or audit data us differentiated from abnormal traffic/data. Rules in the rule set of genetic algorithm are of type if-then. Following is general syntax for rule in genetic algorithm:

if { condition } then { act }

condition refers to the data to be verified and rule in rule set while act is the action to be performed if condition is true. A condition can check for port numbers of network protocols, protocols used, duration of connection, IP address of source and destination etc. while act refers to the action to be performed when condition is true like sending alert message, creating log messages etc.

The Fuzzy genetic algorithm (FGA) starts from a population of individuals generated randomly. Each individual is an "*ifthen*" fuzzy rule. In order to optimize the set of fuzzy rules already generated in the first stage, a genetic algorithm process which consists of selection, crossover and mutation operators are applied on the individuals. *The important components of the FGA for intrusion detection are defined in following:*

A. The fuzzy rule encoding

A fuzzy rule "*if-then*" is encoded as a string. We have used a vector of 16 bits where each bit corresponds to an attribute. Five possible linguistic values may be used for each attribute which are: S: Small, MS: Medium Small, M: Medium, ML: Medium Large and L: Large. Figure 1 draws the Membership functions of the five linguistic values.

For example: Let us consider the rule: If X1 is medium, X2 is medium small X3 is large and X4 is small, then Class= Cj with CF = CFj. Where Xi is the connection attribute, Cj is the class obtained after classification and CFj is its degree of confidence. The Corresponding code is : *"M, MS, L, S"*

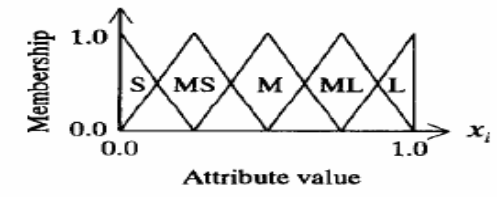

Figure 1: Membership functions of five linguistic values (S: small, MS: medium small, M: medium, ML: medium large, L: large).

B. The membership function μ (X)

The membership function for each attribute X noted μ(X) is calculated by a projection on the graph of the fuzzy set . Formula (2) shows how we can calculate the μ(X) value.

$$\mu(X)=\text{Max}\ \{\ 0,1 -(\ |X-Xo|/b)\} \text{ --------- (2)}$$

where: b is the base of the triangle, b = 0.5. ,X0 = {0, 0.25, 0.5, 0.75, 1} corresponding to {S, MS, M, ML, L}. X: is the attribute value after normalization

C. An individual representation

The individual representing a fuzzy "*if-then*" rule is generated randomly. For each attribute Xi, a linguistic value (among the five values of the fuzzy set) is assigned randomly.

D. The evaluation of a fuzzy rule and the fitness function

To evaluate an "*if-then*" rule *Rj* and classify a connection *Xp* with a certain confidence degree, we have used the method introduced in [11]. To evaluate a fuzzy rule *Rj*, we give the following steps:

- Calculate the compatibility of connections with the rule

  Rj: Let us consider the fuzzy *if-then* rule *Rj* denoted "Aj1 AJ2 ... ."AJn", we calculate the compatibility of each connection *Xp* of the dataset with the rule *Rj* by using the Formula (3).

$$\mu_{Rj}(Xp)= \mu A_{j1}(X1) \text{ x } A_{j2}(X2) \text{ ..... x } A_{jn}(Xn) \text{ ----------(3)}$$

where $\mu()$ is the membership function. m: is the total number of connections.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

Xi: are the attributes. *Xp* is the current connection and n is the number of attributes which equals to 16.

- Calculate the sum of the compatibilities for each class of the five categories: for each class *h* belonging to the five classes DoS, R2L, U2R, Probing and Normal. If two classes had the same maximum value then the class is not specified (Cj = null) and CFj = 0.
- The Confidence degree and fitness value is calculated.

E. The genetic algorithm operators

The genetic algorithm we used performs for each generation:

- A random one-point crossover on two randomly selected individuals.
- A random mutation of all genes of an individual randomly selected. The mutation operator has two functions:

1) A regulation of the population explosion caused by the crossover operator.

2) The enrichment of the population by introducing new genes.

- A selection of individuals having a fitness value >0. So all individuals having a fitness value equals to zero are discarded and eliminated from the population. We consider only individuals with a fitness value superior to zero.

F. The fuzzy genetic algorithm

The different steps of the proposed fuzzy genetic algorithm for intrusion detection are depicted on Figure 2

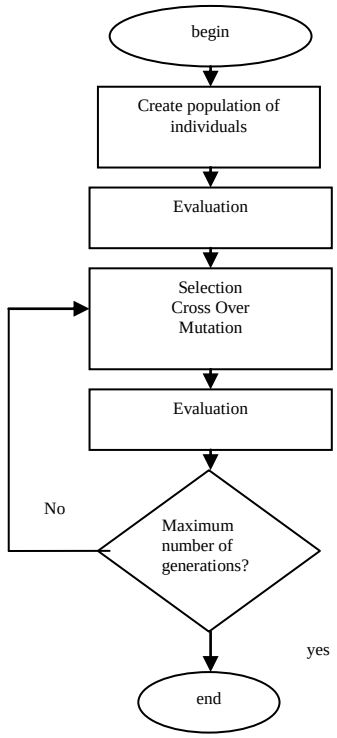

Figure 2 : Fuzzy Genetic Algorithm for ID

IV. EXPERIMENTAL RESULTS

The normalization phase is launched where the various attributes of connections of all matrices are normalized. We have obtained five normalized matrices U2R, R2L, Probing, Normal and DOS. The next step is the generation of fuzzy rules. To do this, we used the "rand" function (random number to generate random numbers that must be among the five values (1, 2, 3, 4. 5) which correspond to (Small, Medium Small, Medium, Medium Large and Large). We have applied the FGA on the five matrices *Rand* representing fuzzy rules.To evaluate the performance of the approach, we used the following measures:

- True Positives (**TP**): is the number ofnormal connections classified by the genetic approach as normal.
- False Positives (**FP):** is the number of normal connections classified as attacks by the genetic approach.
- True Negatives (**TN)**: is the number of attack connections classified as attacks by the genetic approach.
- False Negatives (**FN)**: is the number of attack connections classified as normal by the genetic approach.

·Specificity: It describes the ability to identify negative results (test the reliability of the given method).

$$\text{Specificity} = \frac{TN}{TN+FP}$$

The Success rate is as follows:

- Dos class - 99%
- Normal Class -92.5. %
- R2L Class – 95 %
- U2R Class -96%
- Probe class -94%

V. CONCLUSION

In this paper, we proposed and implemented a fuzzy genetic algorithm for solving the intrusion detection problem. The results showed the effectiveness of this classification in the field of intrusion detection. In future work will be planned to minimize the computation time consuming by the FGA algorithm.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

REFERENCES

[1] H.S. Javitz, A. Valdes, T.F. Lunt, A. Tamaru, M. Tyson, and J. Lowrance. "Next generation intrusion detection expert system (NIDES)". Technical Report A016- Rationales, SRI, 1993.

[2] Duanyang Zhao, Qingxiang Xu, Zhilin Feng, "Analysis and Design for Intrusion Detection System Based on Data Mining", 2010 Second International Workshop on Education Technology and Computer Science

[3] H.S. Vaccaro and G.E. Liepins. "Detection of anomalous computer session activity". In Proceedings of the IEEE Symposium on Security and Privacy, May 1989.

[4] Ludovic Mé. "GASSATA, A genetic algorithm as an alternative tool forsecurity audit trails analysis". In First international workshop on the Recent Advances in Intrusion Detection, 1998.

[5] T.F. Lunt and R. Jagannathan. "A prototype real-time intrusion-detection expert system". In Proceedings of the IEEE Symposium on Security and Privacy, pages 59- 66, 1988.

[6] S. Kumar and E.H. Spafford. "A pattern-matching model for misuse instrusion detection". In Proceedings of the national computer security conference, pages 11- 21, 1994.

[7] S Terry Brugger, "Data Mining Methods for Network Intrusion Detection" University of California, Davis, June 9, 2004.

[8] Theodoros Lappas and Konstantinos Pelechrinis, "Data Mining Techniques for (Network) Intrusion Detection Systems", Department of Computer Science and Engineering, UC Riverside, Riverside CA 92521.

[9] W. Lee, S. Stolfo, and K. Mok. "Mining Audit Data to build Intrusion Detection Models". In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, pages 66–72. AAAI Press, 1998.

[10] H. Shah, J. Undercoffer, and A. Joshi. "Fuzzy Clustering for Intrusion Detection". In Proceedings of the 12th IEEE International Conference on Fuzzy Systems, pages 1274–1278. IEEE Press, Vol (2), 2003.

[11] S. Selvakani and R.S. Rajesh, "Genetic Algorithm for Framing Rules for Intrusion Detection" IJCSNS International Journal of Computer Science and Network Security, Vol. 7 No. 11, November 2007.

[12] Zadeh L.A., 1965, "Fuzzy sets". Information and Control 8: 338–353.

[13] The data set can be downloaded as (http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html).

[14] M. Saniee Abadeha, , J. Habibia, C. Lucasb, "Intrusion detection using a fuzzy genetics-based learning algorithm", Journal of Network and Computer Applications 30 (2007) 414–428.

Biography

B.Ben Sujitha has the degree of B.E., M.Tech in Information Technology currently pursuing Ph.D in the area of Network Security. She is a Professor in the Department of Information Technology at Ponjesly College of Engineering, Nagercoil, Tamilnadu, India.

R.Roja Ramani has the degree of B.Tech, M.Tech in Information Technology, her area of interest is Network Security. She is a Assistant Professor in the Department of Information Technology at Sethu Institute of Technology,Virudhunagar, Tamilnadu, India.

Parameswari has the degree of B.Tech, M.Tech in Information Technology, her area of interest is Network Security. She is a Assistant Professor in the Department of Information Technology at Sethu Institute of Technology,Virudhunagar , Tamilnadu, India.