ISSN 2229-3531 (Print)
JCIT Vol. 9(1), 9-13 (2018). Periodicity-2-Monthly
ISSN 2455-9997 (Online)

JOURNAL OF COMPUTER & INFORMATION TECHNOLOGY
An International Open Free Access Peer Reviewed Research Journal of Computer Science Engineering & Information Technology
website:- www.compitjournal.org
Estd. 2010

Intrusion Detection Using Data Mining Along Fuzzy Logic & Genetic Algorithms

RUCHI CHATURVEDI [1], BABITA PATHIK[2] and SHIV KUMAR[3]

[1]M. Tech Scholar, Department of Computer Science & Engineering, Lakshmi Narain College of Technology & Excellence Bhopal (M.P.), (India)
[2]Assistant Professor, Department of CSE, Lakshmi Narain College of Technology & Excellence, Bhopal (M.P) (India)
[3]Professor & Head, Department of CSE, Lakshmi Narain College of Technology & Excellence, Bhopal (M.P), (India)
[1]Corresponding Author Email: ruchichaturvedi29@gmail.com
http://dx.doi.org/10.22147/jucit/090102

Acceptance Date 31st January 2018, Online Publication Date 2nd February, 2018

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 modelled as a fuzzy rule such as "if-then" and improved by a genetic algorithm. The reasons for introducing fuzzy logic is twofold, the first being the involvement of many quantitative features where there is no separation between normal operations and anomalies. Thus fuzzy association rules can be mined to find the abstract correlation among different security features. 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.

Key word: Unsupervised Machine Learning, Network Intrusion Detection, Network Security, genetic algorithm, fuzzy logic, classification, intrusion detection, DARPA data set.

I Introduction

Intrusions can be defined as actions that attempt to bypass the security mechanisms of computer systems. Intrusions may take many forms: attackers accessing a system through the Internet or insider attackers; authorized (official) users attempting to gain and misuse non-authorized privileges. So, we say that intrusions are any set of actions that threaten the integrity, availability, or confidentiality of a network resource. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions. Intrusion detection systems (IDS) raise the alarm when possible intrusions occur[1-3].

A significant challenge in providing an effective defence mechanism to a network perimeter is having the ability to detect intrusions and implement counter measures. Components of the network perimeter defence capable of detecting intrusions are referred to as Intrusion Detection Systems (IDS). Intrusion Detection techniques have been investigated since the mid 80s and depending on the type and source of the information used to identify security breaches, they are classified as host-based or network based. A lot of research into artificial neural networks (ANNs) has been undertaken. In[4], artificial neural networks and support vector machine (SVM) algorithms were applied to intrusion

Host-based systems use local host information such as process behaviour; file integrity and system logs to detect events. Network-based systems use network activity to perform the analysis. Combinations of these two types are possible. Depending on how the intrusion is detected an IDS is further classified as signature-based (also known as misuse system) or anomaly-based[7]. Signature-based systems attempt to match observed activities against well defined patterns which also called signatures. Anomaly-based systems look for any evidence of activities that deviate from what is considered normal system use. These systems are capable of detecting attacks for which a well-defined pattern does not exist 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]. The expert systems[3] and neural networks[2] are two well-known methods used to calculate a user profile.

II Literature Survey :

2.1 B. Ben Sujitha, R.Roja Ramani, Parameswari *"Intrusion Detection System using Fuzzy Genetic Approach,"* International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 10, December 2012

In this paper Author conclude that 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 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.

***2.2* Swati Sharma, Santosh Kumar, Mandeep Kaur *"Recent trend in Intrusion detection using Fuzzy- Genetic algorithm,"* International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2014**

In this paper, Authors provided a survey of intrusion detection using fuzzy logic and GA. A brief overview of intrusion detection system, genetic algorithm, fuzzy logic and related work techniques are discussed with their advantages. GA is an optimization algorithm that can help in finding appropriate fuzzy rules and fuzzy rule is a machine learning algorithm. Fuzzy- genetic based approach provides performance better than GA based techniques.

Due to increasing incidents of attacks on network securing data is a prime goal. More hybrid techniques should be investigated in this area. In misuse detection mode signature of new intrusions should be created so that it is easy to catch the attack. This paper will prove a good starting point for newcomers in the field of GA and fuzzy logic based intrusion detection and is useful for people looking for a quick review of recent development in this field.

***2.3* Y. Dhanalakshmi and Dr. I. Ramesh Babu *"Intrusion Detection Using Data Mining Along Fuzzy Logic and Genetic Algorithms,"* IJCSNS International Journal of Computer Science and Network Security, VOL. 8 No. 2, February 2008**

In this paper, a fuzzy genetic algorithm is proposed for dealing with the intrusion detection problem considering KDD99 dataset. Results are compared with the existing system which uses sequential algorithm for intrusion detection. The results show that the accuracy of detection rate of the proposed system for DoS, probe, Remote to User Attacks (R2l) and User to Root attack (U2r) are more compared to the existing systems. The time required for the training and testing of the dataset using the proposed system is less compared to the existing systems and memory allocation also requires less space for proposed system than existing systems.

***2.4* Swati Dhopte, N. Z. Tarapore *"Design of Intrusion Detection System using Fuzzy Class-Association Rule Mining based on Genetic Algorithm,"* International Journal of Computer Applications (0975 – 8887) Volume 53– No.14, September 2012**

In this paper, is considered as a major issue in networks, since the network has extended dramatically. Therefore, intrusion detection systems have attracted

attention, as it has an ability to detect intrusion accesses effectively. These systems identify attacks and react by generating alerts or by blocking the unwanted data/traffic. The proposed system includes fuzzy logic with a data mining method which is a class-association rule mining method based on genetic algorithm. Due to the use of fuzzy logic, the proposed system can deal with mixed type of attributes and also avoid the sharp boundary problem. GA-based fuzzy Class Association Rule Mining with Sub-Attribute Utilization and its application to classification, which can deal with discrete and continuous attributes at the same time? In addition, this method was applied to both misuse detection and anomaly detection. Experiments were performed with practical data provided by KDD99 Cup. The experiment results show that for misuse detection, the proposed method can provides high detection rate and low false positive rate, which are two important criteria for security systems. For anomaly detection, the method provides high detection rate and reasonable false positive rate even without prior knowledge of attack signatures, which is an important advantage over other methods.

III Background Study :

3.1 Data Set :

KDD99 CUP is the dataset prepared for the Third International Knowledge Discovery and Data Mining (KDD) Tools Competition, which was held in conjunction with KDD99 the Fifth International Conference on KDD[24]. The competition task was to build a network intrusion detector, a predictive model capable of distinguishing between “bad” connections, called intrusions or attacks, and “good” normal connections. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment. There are approximately 4,940,000 kinds of data in the training dataset, 10% of which is provided, there are 3,110,291 kinds of data in test dataset, and there are totally 41 types of network connection characteristics (characterized by continuous data and discrete data) in each kind of network connection record[9,10,11].

3.2 Data Mining :

Data mining generally refers to the process of extracting or mining knowledge from a large amount of data. This process first understands the existing data and then predicts the new data. It is the core of Knowledge Discovery and Data mining (KDD). Kind of Patterns found in Data Mining Task are specified by Data Mining Functionalities. In general, data mining tasks are categorized into two categories: predictive and descriptive. The general properties of the data in the database are characterized by Descriptive mining. Inference on the current data in order to make predictions is performed by Predictive mining [12]. Well-known data mining techniques are:

1) Classification
2) Clustering
3) Association-Rule mining

3.3 Genetics Algorithm :

Genetic algorithms (GA) are search algorithms based on the principles of natural selection and genetics, introduced by John Holland in the 1970s and inspired by the biological evolution of living beings. Genetic algorithms abstract the problem space as a population of individuals, and try to explore the fittest individual by producing generations iteratively. Individuals are represented by a string of symbols.Each individual is called a chromosome, and is composed of a predetermined number of genes[13]. The generation of new offsprings includes the operations such as crossover, mutation and selection operations[14,15]:

3.4 Fuzzy Theory :

Crisp sets do not always satisfy the needs of real world applications, because they only allow a membership of 1 or 0, i.e. member or non-member[16]. In the real world, it is not possible at all times to assign an object clearly to a certain group of objects. Rather, it might lie in between two different sets[16]. Therefore, Georg Cantor invented fuzzy sets theory which generalizes member and non-member functions by assigning values that fall in a specified range, typically 0 to 1, to the elements. Fuzzy set theory overcomes the sharp boundary problem by allowing different degrees of memberships[17,18].

IV Problem Identification :

A lot of research work has been done in the field. Intrusion Detection is one of the major concerns in any computer networks environment. Many techniques including that of Artificial Intelligence have been proposed and are in use presently. There are many researchers who developed intelligent Intrusion Detection Systems. The input to any Intrusion Detection System is some uncertain or fuzzy information that has to be processed. A part from being fuzzy in nature the information could be very large requiring data mining techniques for extracting the data. As the data for extracting has to follow certain rules, we need to have certain mechanism to pick up best possible rules. A genetic algorithm approach for identifying these rules is chosen. The present work has explored the possibility of integrating the fuzzy logic with Data Mining methods using Genetic Algorithms for intrusion detection. The reasons for introducing fuzzy logic is twofold, the first being the involvement of many quantitative features where there is

no separation between normal operations and anomalies. Thus fuzzy association rules can be mined to find the abstract correlation among different security features.

V Genetic Algorithm :

The genetic algorithm[5] is similar to the natural evolution; it finds good chromosomes for solving problem through acting on genes in the chromosome. It need only evaluate each chromosome which the algorithm generates, and choose chromosomes based on adaptive value, and the better adaptive values of chromosomes are, the more reproduction opportunities are. With the help of fitness functions, the each individual is evaluated, low fitness individuals are eliminated, high fitness individuals are chosen to partake in genetic operator, after genetic operator the collection of individuals to form new population of next generation, new population evolves next round. This is the basic principle of genetic algorithm. The main flow is shown in the Fig. 1.

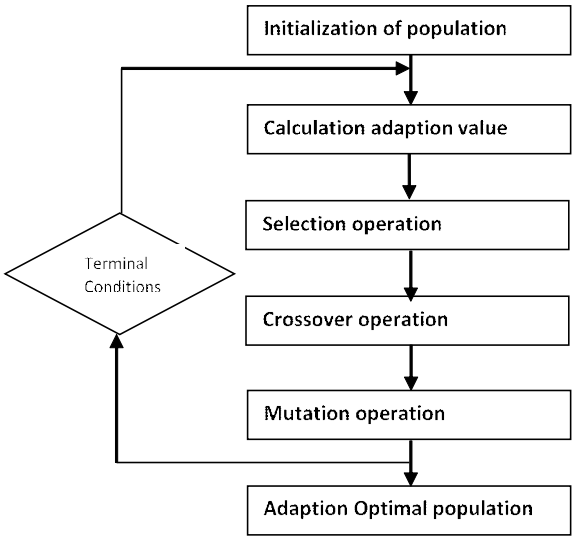

Fig. 1. Flow of genetic algorithm

Important concepts in Fuzzy Genetic algorithm are given bellow:

GA Operators- The different GA parameter are selection, mutation, and crossover. These are the most winning parts of the algorithm because they are contributing to the gen*eration of each population.*

Selection phase-In selection phase population individuals with superior fitness are selected, or else it is damaged.

Crossover- In this method a pair of individual randomly participates in exchanging their parent's genes with each other, awaiting an entire new population to be generated.

Mutation- It flips a number of the bits in an individual so that all bits are filled, There is a low probability of predicting the alter.

Fitness Function- It is defined as a function that scales the value of individual relative to the rest of the population. It generates the best likely solutions from the quantity of candidates in the population.

In the above model we will discuss number of features and the attacks out of some we will concentrate more which is given below:

Denial of Service (DoS) attack: Over usage of the bandwidth or non availability of the system resources leads to the DoS attacks. Examples: Neptune, Teardrop and Smurf.

User to Root (U2R) Attack: Initially attacker access normal user account, later gain access to the root by exploiting the vulnerabilities of the system. Examples: Perl, Load Module and Eject attacks.

Probe Attack: Have an access to entire network information before introducing an attack. Examples: ipsweep, nmap attacks.

Root to Local (R2L) Attack: By exploiting some of the vulnerabilities of the network attacker gains local access by sending packets on a remote machine. Examples: imap, guess password and ftp-write attacks.

VI Conclusion

After reading number of research papers we conclude that, this algorithm can be used to detect normal data and attack data along with its 5 categories *i.e.* DOS, Probe, U2R and R2L. It is expected that result obtained from individual execution of Fuzzy or Genetic Algorithm to detect intrusion will be far better with this combination approach of genetic, fuzzy and Apriori. Also it is expected that lacunas of individual approach like low detection rate will be ruled out. Intrusion detection system can also be evaluated in terms of detection rate, detection speed, false alarm rate with help of this proposed method. The present work is the extension of in the areas of fuzzy association rules based on mining and genetic algorithms. We have proposed architecture for Intrusion Detection methods by using Data Mining algorithms to mine fuzzy association rules by extracting the best possible rules using Genetic Algorithms.

References

1. P. Jongsuebsuk, N. Wattanapongsakorn, C. Charnsripinyo "Real-Time Intrusion Detection with Fuzzy Genetic Algorithm." ©2013 IEEE.

2. T.P. Fries, "A fuzzy-Genetic approach to network intrusion detection," GECCO'08: The 10th Annual Conference on Genetic and Evolutionary Computation, pp. 2141-2146.N (2008).
3. J. Gomez and E. León, "A fuzzy set/rule distance for evolving fuzzy anomaly detectors," IEEE International Conference on Fuzzy Systems ART. No. 1682017, pp. 2286-2292.
4. N. Ngamwitthayanon and N. Wattanapongsakorn, "Fuzzy- ART in network anomaly detection with feature-reduction dataset," The 7th International Conference on Networked Computing, INC2011, Art. No. 6058956,
5. W. Li, "A Genetic Algorithm Approach to Network Intrusion Detection" SANS. Institute, USA, (2004).
5. Susan M.Bridges, Rayford B. Vaughan "Fuzzy Data mining and Genetic Algorithms Applied to Intrusion Detection", Conference onNational Information Systems Security, Oct. (2000).
6. K.C.C. Chan and W.H. Au, "Mining Fuzzy Association Rules" Proc.Of ACM CIKM, pp.209-215 (1997).
7. M. Delgada, Nicolas Marin, Daniel Sanchez and Maria Amparo Vila "Fuzzy Association Rules:General Model and Applications", IEEE Transactions on Fuzzy Systems, Vol. 11, No. 2, April, pp.214-225 (2003).
8. K.M. Faraoun, and A. Boukelif "Genetic Programming Approach for Multi-Category Pattern Classification Applied to Network Intrusions Detection", International Journal of Computational Intelligence Vol. *3*, No. 1, pp. 79-90 (2006).
9. Hoque M., Mukit M. and Bikas M., "An Implementation of Intrusion Detection System using Genetic Algorithm," International Journal of Network Security & Its Applications (IJNSA), Vol. *4*, No.2, March (2012).
10. Sathya S., Ramani R., Sivaselvi K., "Discriminant Analysis based Feature Selection in KDD Intrusion Dataset," International Journal of Computer Applications (0975 – 8887), Volume *31* No.11, October (2011).
11. Kddcup 1999data [Online]. Available: kdd.ics.uci.edu/databases/kddcup99/kddcup99 .html.
12. Han J., Kamber M., "Data Mining," Morgan Kaufmann Publishers, (2001).
13. Gong R., Zulkernine M., Abolmaesumi P., "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection," Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Dis tributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks, IEEE, (2005).
14. Shetty M. and Shekokar N., "Data Mining Techniques for Real Time Intrusion Detection Systems," International Journal of Scientific & Engineering Research Volume *3*, Issue 4, April (2012).
15. Abdullah B., Abd-alghafar I., "Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System," 13th International Conference on Aerospace Sciences & Aviation Technology, ASAT- 13, (2009).
16. Luo J., "Integrating fuzzy logic with data mining methods for intrusion detection," Master's Thesis, Department of Computer Science, Mississippi State University, Starkville, MS, (1999).
17. Helm B., "Fuzzy Association Rules: An Implementation in R," Master's Thesis, Vienna University of Economics and Business Administration Vienna (2007).
18. Mabu S., Chen C., Shimada K., "An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming," IEEE Transactions Systems, Man, Cybernetics C, Application and Reviews, volume *41*, number 1, pp. 130–139, January (2011).