INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

AN IMPLEMENTATION OF INTRUSION DETECTION SYSTEM USING REAL CODED GENETIC ALGORITHM

Asha Jenifer[1] S, Geetha S[2]

[1]*M.Tech, Information Security and cyber forensics, Department of Computer Science and engineering, Dr. MGR Educational and Research Institute,asha.jenifer91@gmail.com*
[2] *Associate Professor, Department of Computer Science and engineering, Dr. MGR Educational and Research Institute,geethasoman@gmail.com*

Abstract

In the recent era of electronic commerce and web technology, network security has become very important. Intrusion detection is a mechanism of providing security to computer networks. Although there are some existing techniques to detect intrusion there is a need to increase its efficiency. Data mining techniques give an efficient result when applied to IDS. One of the data mining techniques called genetic algorithm when applied to intrusion detection system will give a much efficient Intrusion Detection System (IDS) by increasing the detection rate and reducing the false positive. One of the types of GA called Real coded Genetic Algorithm (RCGA) is applied to our IDS.

Keywords: Intrusion Detection System, Genetic Algorithm, Electronic commerce, Data mining.

1. Introduction

There has been a great advancement in internet technologies in the recent years like virtualization, cloud technologies and so on. Network is a backbone for any of the internet technologies. The data stored in the system and the data transmitted through the network is not safe. Data is a valuable asset for any individual or organization. Unauthorized access into the system is host based attack. Unauthorized access into the network is network based attack. There is a need to detect and prevent such unauthorized access. The attackers are smarter that they apply new attack patterns for penetrating into the system. Intrusion detection system is necessary to detect intrusions. There are intrusion detection systems that apply various techniques. Data mining techniques are applied to IDS for efficient detection. Among other data mining techniques genetic algorithm proves to be proficient. The application of real coded real coded genetic algorithm is found to give high detection rate when applied to IDS.

2. Real Coded Genetic Algorithm

Real coded Genetic Algorithm (RCGA) possesses a lot of advantages than its binary coded counterpart when dealing with continuous search spaces with large dimensions and a great numerical precision is required. In RCGA, each gene represents a variable of the problem, and the size of the chromosome is kept the same as the length of the solution to the problem. Therefore, RCGA can deal with large domains without sacrificing precision as the binary implementation did (assuming a fixed length for the chromosomes). Furthermore, RCGA possesses the capacity for the local tuning of the solutions; it also allows integrating the domain knowledge so as to improve the performance of Genetic Algorithm (GA). But RCGA is still harassed by the requirement of population diversity and the frequent computation of fitness, and may become very time-consuming. As a result, its inherent parallelism is inhibited and its application field is restricted by the speed bottleneck as its binary implementation did.
The RCGA operates on a population of chromosomes (or individuals, creatures, etc.) simultaneously. It starts from an initial population, generated randomly within the search space. Once the initialization is completed, the population enters the main RCGA loop and performs a global optimization for searching the optimum solution of the problem. In a RCGA loop, preprocessing, three genetic operations, and post processing are carried out in turn. The RCGA loop continues until the termination conditions are fulfilled.

Table 1 Chromosome Representation of a Rule

S/no	Feat Name	Format	Number of Genes
1	Duration	H:M:S	3
2	Protocol	Numeric	1
3	Source Port	Numeric	1
4	Destination Port	Numeric	1
5	Source IP	a.b.c.d	4
6	Source IP	a.b.c.d	4
7	Attack Name and type	String	1

3. Training Data

KDD dataset is used in our system to train the algorithm. The algorithm will input the dataset and classify it based on the 4 features in the dataset: Basic Features, Content Features, Time-based Traffic Features, Host-based Traffic Features. There are 41 parameters in 4 features. The parameters are represented as in the chromosome structure. The dataset has connection patterns

4. Implementation

4.1 Dataset Loading

KDD99 data set is loaded into the database. This dataset will be given to the algorithm for training to classify the packets as attack patterns or genuine attacks. There are around 5 million connection patterns in the dataset which has patterns of all known attacks and also normal connection patterns. The loaded dataset is preprocessed to remove null values and the features are extracted from the packets.

4.2 Rule creation

The extracted features from the dataset are given to the real coded genetic algorithm. The algorithm has 3 phases: Selection, Crossover and mutation.
We use Tournament selection for the selection process. Here n random individuals are selected from the population and they will compete to get the best fit individual that can undergo crossover. For the crossover phase we propose to use cut n slice crossover. Here the crossover point in the parent can be in any order that is more suitable for the production of new offspring. Next process is the mutation. Here one or more portion of genes are changed to maintain the diversity of the population. The outcome of this module is the generated new rule that new attacks. The rules are stored in the database.

4.3 Intrusion Detection

In this module the sensor in the IDS will detect the incoming connections and analyzed in the analysis component. The analysis component will compare the packet features with the rules in the database. The analysis component will identify if the packet is genuine or attack packet.

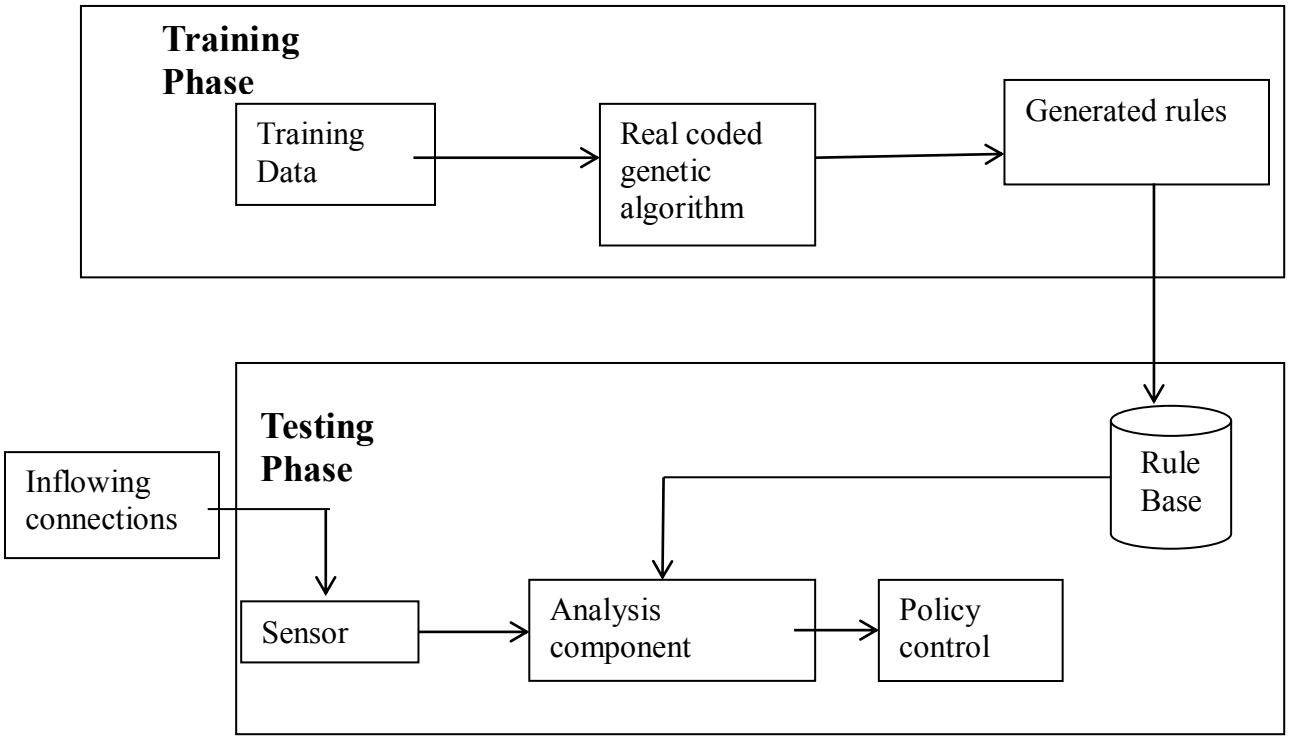

Figure 1: Architecture of IDS

4.4 Policy control

In this module the control action will be implemented. If the analysis component classifies the packet as attack then the warning is generated and the warning message is to the administrator. If the analysis component classifies the packet as genuine then the connection is let to proceed.

5. Conclusion

In this paper we designed an IDS that detects new pattern of attacks with high detection rate and less false positive rate. Although the proposed system is capable of detecting new attacks, the system takes time to get trained. But its trained it is much efficient in detecting new attacks.

REFERENCES

[1] H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood" *Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets*"
[2] A.T.Haghighat, M.Esmaeili, A.Saremi, V.R.Mousavi, "*Intrusion Detection via Fuzzy-Genetic Algorithm Combination with Evolutionary Algorithms*", in ICIS 2007, IEEE, 2007
[3] Dong Seong Kim, Ha-Nam Nguyen, Jong Sou Park, "*Genetic Algorithm to Improve SVM Based Network Intrusion Detection System*", AINA'05, IEEE, 2005
[4] KDD-cup data set http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
[5] T. Ozyer, R. Alhaji, and K. Barker, "*Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule prescreening*", Journal of Network and Computer Applications, pp. 99-113, 2007.
[6] Danilo Bruschi, Lorenzo Martignoni and Martia Monga, "*Code Normalization for Self-Mutating Malware*," IEEE Security & Privacy, Vol. 5, No. 2, 2007. pp 46-54.
[7] K. Byung-Joo and K. Il-Kon, "Kernel Based Intrusion Detection System," *Proceedings of 4th Annual ACIS International Conference on Computer and Information Science"*, 14-16 July 2005, pp. 13-18. doi:10.1109/ICIS.2005.78
[8] R. Bace and P. Mell, "*Intrusion Detection Systems*," 2001. http://csrc.nist.gov/publications/nistpubs/800-31/sp800-31.pdf
[9] S. Northcutt and J. Novak, "*Network Intrusion Detection: An Analyst's Handbook*," 2nd Edition, New Riders Publishing,Berkeley,2000.