Volume 4 Issue XII, December 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A survey on Intrusion Detection System by Using Data Mining Based on Class-Association-Rule Mining Using GNP

Mr. Shankar L. Tambe[1], Prof. Ms.Rasna Sharma[2]
[1]*Student, CSE, CIIT Indore, RGPV, Bhopal, India*
[2]*Assistant Professor, CSE, CIIT Indore, RGPV, Bhopal, India*

Abstract— There is often the need to update an installed Intrusion Detection System (IDS) due to new attack methods or upgraded computing environments. Since many current IDSs are constructed by manual encoding of expert knowledge, changes to IDSs are expensive and slow. This paper describes a data mining framework for adaptively building Intrusion Detection (ID) models. Now security 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. Genetic algorithm is used to extract many rules which are required for anomaly detection systems.
Keywords— Data Mining, Intrusion Detection System (IDS), Genetic Algorithm (GA), Network Security, Fuzzy Logic.

I. INTRODUCTION

This document is a template. In recent years, computer security has become increasingly important and an international priority. This is due to the emergence of electronic commerce, the tremendous use of computers and the rapid growth of computer networks. Computer security is defined as the protection of computing systems against threats to confidentiality, integrity, and availability. Security threats come from different sources such as natural forces (flood), accidents (fire), failure of services (power) and people known as intruders. Two types of intruders are: the external intruders who are unauthorized users of the machines who attack by using various penetration techniques, and internal intruders, refers to those with access permission who wish to perform unauthorized activities.

When an intruder attempts to break into an information system or performs an action not legally allowed, this activity is referred to as an intrusion. Intrusion techniques may include password cracking, exploiting software bugs and system misconfiguration, sniffing unsecured traffic, or exploiting the design flaw of specific protocols. An IDS is a system for detecting intrusions and reporting to the proper authority.

Finally, these systems require large amounts of training data and are significantly more complex than traditional systems. In order to be able to deploy real time data mining-based IDSs, these issues must be addressed. These problems are independent of the actual learning algorithms or models used by IDS and must be overcome in order to implement data mining methods in a deployable system. An effective data mining-based IDS must address each of these three groups of issues. Although there are tradeoffs between these groups, each can generally be handled separately.

IDS can also be divided into two groups depending on where they look for intrusive behavior Network-based IDS (NIDS) and Host-based IDS. Network-based IDS refers to systems that identify intrusions by monitoring Traffic through network devices (e.g. Network Interface Card, NIC). Host- based IDS requires small programs to be installed on individual systems to be monitored.

II. LITERATURE REVIEW

An Shingo Mabu, Ci Chen, Nannan Lu, Kaoru Shimada, and Kotaro Hira- sawa[1] proposed “An Intrusion-Detection Model Based on Fuzzy Class-Association- Rule Mining Using Genetic Network Programming” which applies different methods as Genetic Network Programming(GNP) and Fuzzy Logic. They provided the techniques to solve the sharp boundary problem by using Fuzzy logic and used mixed data set that contain both discrete and continuous attribute. The disadvantage is that they have used traditional class association rule mining.

Swati Dhopte, N. Z. Tarapore[2] brought a method as Design of Intrusion Detection System using Fuzzy Class-Association Rule Mining based on Genetic Algorithm. Authors explained about Fuzzy Class Association and Genetic Algorithm a method which uses the tree for creating the rule without use of graph but it unintentionally increases complicity of their ideas.

Jonatan Gomez and Dipankar Dasgupta[3] deals with Evolving Fuzzy Classifiers for Intrusion Detection using technique Genetic Algorithms to evolve the simple set of fuzzy riles (Classifiers). The advantage is Genetic Algorithm can find simple and good fuzzy rules to characterized intrusion (normal and abnormal) network system. This system fails to Differentiate between the normal and abnormal activities.

Zohair Ihsan, Mohd Yazid Idris and Abdul Hanan Abdullaha[4] proposed a novel Attribute Normalization Techniques and Performance of Intrusion Classifiers: A Comparative Analysis. The technique used is Hybrid Normalization which gives us many advantages such as Hybrid Normalization can achieve better result than conventional Normalization but failed for more complex procedures.

Mohammad Sazza dul Hoque, Md. Abdul Mukit and Md. Abu Naser Bikas[5] mainly deals with Implementation of Intrusion Detection system Using Genetic Algorithm. In this Genetic Algorithm with KDD Data set to detect Intrusion Detection is used which gives advantage of efficiently detect various types of network Intrusion. It has failed to deal with Parameter and evolution process for GA.

As one of the most popular data mining methods for wide range of applications, association-rule mining is used to discover association rules or correlations among a set of attributes in a dataset. The relationship between datasets can be represented as association rules. An association rule is expressed by $X \Rightarrow Y$,where X and Y contain a set of attributes. This means that if a tuple satisfies X, it is also likely to satisfy Y. The most popular model for mining association rules from databases is the apriori algorithm [6]. This algorithm measures the importance of association rules with two factors: support and confidence. However, this algorithm may suffer form large computational complexity for rule extraction from a dense database. In order to discover interesting rules from a dense database, genetic algorithm (GA) [7], [8] and genetic programming (GP) [9], [10] have been applied to association-rule mining. In the GA, the method evolves the rules during generations and individuals or population themselves represent the association relationships [11]. However, it is not easy for GA to extract enough number of interesting rules, because a rule is represented as an individual of GA. GP improves the interpretability of GA by replacing the gene structures with the tree structures, which enables higher representation ability of association rules [12], [13]. Page Layout

III. LIMITATIONS OF EXISTING SYSTEMS

Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDPSes for other purposes, such as identifying problems with security policies, documenting existing threats, and deterring individuals from violating security policies. IDPSes have become a necessary addition to the security infrastructure of nearly every organization. Network Intrusion Detection System (NIDS) will be another wall for protection. Most of the existing commercial NIDS are signature-based but not adaptive. There are many have problems such as attack stealth ness: attackers try to hide their actions from either an individual in monitoring the system or a NIDS, novel intrusion: it is undetectable by signature-based NIDS; they can only be detected as anomalies by observing deviations from normal network behaviour. Whereas Anomaly detection approaches attempt to identify abnormal behaviour in patterns and can make use of supervised or unsupervised methods to detect the anomalies or attacks Does typically record information related to observed events, notify security administrators of important observed events, and produce reports. Some attacks that are aimed to be handled are:

A. Denial of Service (DoS) attack

A DoS attack is aimed at preventing authorized, legitimate users from accessing services on the network. The DoS attack is not aimed at gathering or collecting data.

B. Brute force attack

Brute force attacks simply attempt to decode a cipher by trying each possible key to find the correct one. This type of network attack systematically uses all possible alpha, numeric, and special character key combinations to find a password that is valid for a user account.

IV. PROPOSED SYSTEM

The proposed GA-based intrusion detection using data mining approach contains wastages where each works in a different stage. In

Volume 4 Issue XII, December 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

the training stage, using the GA and fuzzy-association rule mining algorithm, a set of classification rules are generated from KDD dataset. In the intrusion detection stage, the generated rules are used to classify in coming data from a test file. Once the rules are generated, the intrusion detection is simple and efficient.

Following Figure shows the proposed system architecture. Proposed method will overcome the limitation of signature based IDS also provide an additional security and In order to be able to deploy real time data mining-based IDS. Proposed method will extract many rules for anomaly detection.

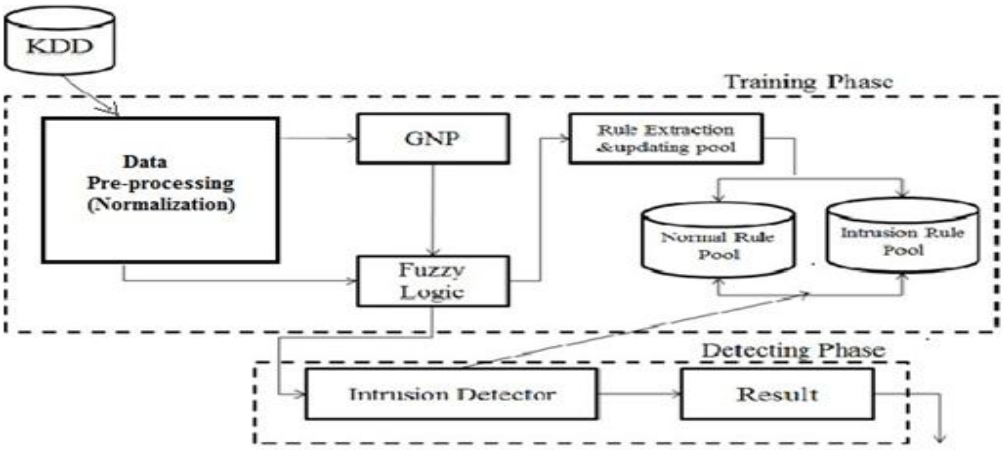

Fig.1. System Architecture

V. CONCLUSION

In this paper, we extend an end-to-end Intrusion Detection System by Using Data Mining Based on Class-Association-Rule Mining for Genetic Network Programming domain. By considering the pros and cons of different methods for Intrusion Detection System (IDS) we can choose the effective one. The important function of the proposed method will efficiently extract many rules that are statistically significant and they can be used for several purposes. Proposed method will successfully overcome the limitation of traditional IDS.

REFERENCES

[1] Shingo Mabu, Ci Chen, Nannan Lu, Kaoru Shimada, and Kotaro Hira-sawa, “An Intrusion-Detection Model Based on Fuzzy Class-Association- Rule Mining Using Genetic Network Programming”, IEEE Transactions On Systems, Man, And Cybernetics-Part C: Applications And Reviews, Volume 41, No. 1,January 2011.
[2] 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.
[3] Jonatan Gomezand Dipankar Dasgupta, ”Evolving Fuzzy Classifiers for Intrusion Detection” Proceedings of the 2002 IEEE Workshop on Infor- mation Assurance United States Military Academy, West Point, NY June2001.
[4] Zohair Ihsan, Mohd Yazid Idris and Abdul Hanan Abdullaha Life SciJ, “Attribute Normalization Techniques and Performance of Intrusion Classifiers: A Comparative Analysis”, 2013.
[5] Mohammad Sazza dulHoque, Md. Abdul Mukit and Md. Abu Naser Bikas, “Implementation of Intrusion Detection system Using Genetic Algorithm”, 2012.
[6] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in Proc. 20th VLDB Conf. , Santiago, Chile, 1994, pp. 487–499.
[7] J. H. Holland, Adaptation in Natural and Artificial Systems .Ann Arbor, MI: Univ. Michigan Press, 1975.
[8] D. E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning Reading, MA: Addison-Wesley, 1989.
[9] J. R. Koza, Genetic Programming, on the Programming of Computers by Means of Natural Selection. Cambridge, MA: MIT Press, 1992.
[10] J. R. Koza, Genetic Programming II, Automatic Discovery of Reusable Programs.. Cambridge, MA: MIT Press, 1994.
[11] A. A. Freitas, Data Mining and Knowledge Discovery with Evolutionary Algorithms New York: Springer-Verlag, 2002.
[12] M. Crosbie and G. Spafford, “Applying genetic programming to intrusion detection,” presented at the AAAI Fall Symp. Series, AAAI Press, Menlo Park, CA, Tech. Rep. FS-95-01, 1995.
[13] W. Lu and I. Traore, “Detecting new forms of network intrusion using genetic programming,” Comput. Intell. , vol. 20, no. 3, pp. 474–494, 2004.