www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 10 Oct. 2016, Page No. 18249-18254

Intrusion Detection System Using Machine Learning Approach

P.Akshaya,
Dept of Computer Science,
Thiagarajar college of engineering,
Madurai.

Abstract
In this paper, we present an intrusion detection model based on genetic algorithm and neural network. The key idea is to take advantage of classification abilities of genetic algorithm and neural network for intrusion detection system. The new model has ability to recognize an attack, to differentiate one attack from another i.e. classifying attack, and the most important, to detect new attacks with high detection rate and low false negative. This approach uses evolution theory to information evolution in order to filter the traffic data and thus reduce the complexity. To implement and measure the performance of this System. We used the KDD99 benchmark dataset and obtained reasonable detection rate

Keywords: Genetic algorithm, neural network, Intrusion detection.

1 Introduction

Detection of these attacks is an important issue of Computer security. Intrusion Detection Systems (IDS) technology is an effective approach in dealing with the problems of network security. In general, the techniques for Intrusion Detection (ID) fall into two major categories depending on the modeling methods used: misuse detection and anomaly detection. Misuse detection compares the usage patterns for knowing the techniques of compromising computer security. Although misuse detection is effective against known intrusion types; it cannot detect new attacks that were not predefined. Anomaly detection, on the other hand, approaches the problem by attempting to find deviations from the established patterns of usage. Anomaly detection may be able to detect new attacks. However, it may also cause a significant number of false alarms because the normal behavior varies widely and obtaining complete description of normal behavior is often difficult. Architecturally, an intrusion detection system can be categorized into three types: host based IDS, network based IDS and hybrid IDS [1] [2]. A host based intrusion detection system uses the audit trails of the operation system as a primary data source. A network based intrusion detection system, on the other hand, uses network traffic information as its main data source. Hybrid intrusion detection system uses both methods [3]. However, most available commercial IDS's use only misuse detection because most developed anomaly detector still cannot overcome the limitations (high false positive detection errors, the difficulty of handling gradual misbehavior and expensive Computation [4]). This trend motivates many research efforts to build anomaly detectors for the

purpose of ID [5]. The main problem is the difficulty of distinguishing between natural behavior and abnormal behavior in computer networks due to the significant overlap in monitoring data. This detection process generates false alarms resulting from the Intrusion Detection based on the Anomaly Intrusion Detection System. The use of Genetic algorithm might reduce the amount of false alarm, where Genetic algorithm is used to separate this overlap between normal and abnormal behavior in computer networks.

2 IDS

2 Existing System

In particular several Neural Networks based approaches were employed for Intrusion Detection. Several Genetic Algorithms (GAs) has been used for detecting Intrusions of different kinds in different scenarios [6][7] [8] [9]. GAs used to select required features and to determine the optimal and minimal parameters of some core functions in which different AI methods were used to derive acquisition of rules [10] [11] [12]. In [13], authors presented an implementation of GA based approach to Network Intrusion Detection using GA and showed software implementation. The approach derived a set of classification rules and utilizes a support-confidence framework to judge fitness function. In [14], authors designed a GA based performance evaluation algorithm for network intrusion detection. The approach uses information theory for filtering the traffic data. In [15], authors used the BP network with GAs for enhancement of BP, they used some types of attack with some features of KDD data. A back-propagation Neural Network was used [16], authors used all features of KDD data, the classification rate for experiment result for normal traffic was 100%, known attacks were 80%, and for unknown attacks were 60%.

3 Proposed systems

Genetic Algorithm is chosen to make this intrusion detection system. This section gives an overview of the algorithm and the system. Genetic Algorithm (GA) is a programming technique that mimics biological evolution as a problem-solving strategy [17]. It is based on Darwinian's principle of evolution and survival of fittest to optimize a population of candidate solutions towards a predefined fitness [7]. GA uses an evolution and natural selection that uses a chromosome-like data structure and evolve the chromosomes using selection, recombination and mutation operators [7]. The process usually begins with randomly generated population of chromosomes, which represent all possible solution of a problem that are considered candidate solutions. From each chromosome different positions are encoded as bits, characters or numbers. These positions could be referred to as genes. An evaluation function is used to calculate the goodness of each chromosome according to the desired solution; this function is known as "Fitness Function". During the process of evaluation "Crossover" is used to simulate natural reproduction and "Mutation" is used to mutation of species [7]. For survival and combination the selection of chromosomes is biased towards the fittest chromosomes. When we use GA for solving various problems three factors will have vital impact on the effectiveness of the algorithm and also of the applications [18]. They are: i) The fitness function; ii) The representation of

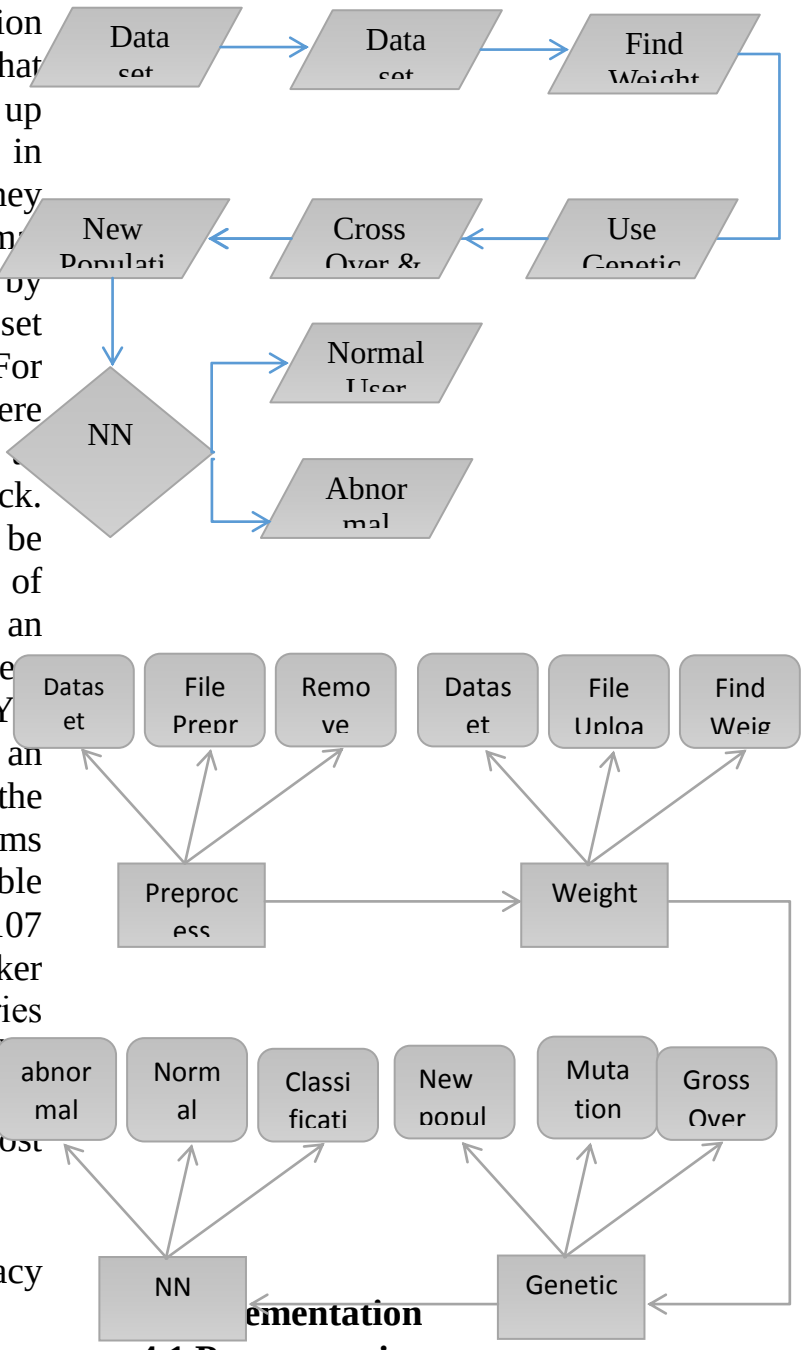
individuals iii) The GA parameters. The determination of these factors often depends on applications and/or implementation. Also all the three steps of generating new population from old population are depicted. The process of generating new population from old population includes selection, crossover, and mutation. If new population is not feasible then quit, otherwise again repeat the generation process. This system can be divided into two main phases: the pre calculation phase and the detection phase. Following are the major steps in pre calculation phase, where a set of chromosome is created using training data. This chromosome set will be used in the next phase for the purpose of comparison. C4.5 is an algorithm used to generate a decision trees and an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used as a statistical classifier, since it is best in classification. It uses the concept of information entropy and builds the decision trees similar to the ID3 algorithm. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sub lists. Neural Networks (NNs) have attracted more attention compared to other techniques. That is mainly due to the strong discrimination and generalization abilities of Neural Networks that utilized for classification purposes [19]. Artificial Neural Network is a system simulation of the neurons in the human brain [20]. It is composed of a large number of highly interconnected processing elements (neurons) working with each other to solve specific problems. Each processing element is basically a summing element followed by an active function. The output of each neuron (after applying the weight parameter associated with the connection) is fed as the input to all of the neurons in the next layer. The learning process is essentially an optimization process in which the parameters of the best set of connection coefficients (weights) for solving a problem are found [21]. An increasing amount of research in the last few years has investigated the application of Neural Networks to intrusion detection. If properly designed and implemented, Neural Networks have the potential to address many of the problems encountered by rule-based approaches. Neural Networks were specifically proposed to learn the typical characteristics of system's users and identify statistically significant variations from their established behavior. In order to apply this approach to Intrusion Detection, I would have to introduce data representing attacks and non-attacks to the Neural Network to adjust automatically coefficients of this Network during the training phase. In other words, it will be necessary to collect data representing normal and abnormal behavior and train the Neural Network on those data. After training is accomplished, a certain number of performance tests with real network traffic and attacks should be conducted [22]. Instead of processing program instruction sequentially, Neural Network based models on simultaneously explorer several hypotheses make the use of several computational interconnected elements (neurons); this parallel processing may imply time savings in malicious traffic

analysis . KDD 99 data set are used as the input vectors for training and validation of the tested neural network. It was created based on the DARPA intrusion detection evaluation program. MIT Lincoln Lab that participates in this program has set up simulation of typical LAN network in order to acquire raw TCP dump data. They simulated LAN operated as a norm environment, which was infected by various types of attacks. The raw data set was processed into connection records. For each connection, 41 various features were extracted. Each connection was labeled normal or under specific type of attack. There are 39 attacker types that could be classified into four main categories of attacks: DOS (Denial of Service): an attacker tries to prevent legitimate use from using a service. E.g. TCP SY Flood, Smurf (391458 record). Probe: an attacker tries to find information about the target host. For example: scanning victims in order to get Knowledge about available services, using Operating System (4107 record). U2R (User to Root): an attacker has local account on victim's host and tries to gain the root privileges (52 record R2L (Remote to Local): an attacker d not have local account on the victim host and try to obtain it (1124 records).

Advantages:

- It is used to improve Accuracy classification performance.
- It is used to predict movement accuracy.
- It is sufficient.
- It will be make new population.
- It produces low false positives.
- The model building time is low.
- It takes the advantages of the ensemble machine learning method where bagging is the added credit the system.

3.1 System Architecture



ementation

4.1 Pre processing:

This process is first convert csv file to text. Then it will be remove unwanted data from txt file like dataset The pre process is very help to pruning process because this process is delete unwanted special character and un-wanted data's This pruning techniques very help to time conception and memory conception The

after pre process we have do uploading process in database.

4.2Genetic algorithm:

In genetic algorithms, a chromosome (also sometimes called a genotype) is a set of parameters which define a proposed solution to the problem that the genetic algorithm is trying to solve. The chromosome is often represented as a simple string; although a wide variety of other data structures are also used.

4.3Neural Networks:

A typical neural network has anything from a few dozen to hundreds, thousands, or even millions of artificial neurons called units arranged in a series of layers, each of which connects to the layers on either side. Some of them, known as input units, are designed to receive various forms of information from the outside world that the network will attempt to learn about, recognize, or otherwise process. Other units sit on the opposite side of the network and signal how it responds to the information it's learned; those are known as output units.

4.4Process Analysis:

An attack pattern, which may be an attacker-specific pattern or a pattern commonly used by attackers, can be identified in the same method. Similarly, an attack pattern that an attacker frequently submits but others have seldom or never submitted will be considered as one of the attacker's representative attack patterns and will obtain a high similarity weight. Hence, signatures collected in an attacker profile can be classified into common signatures and attacker-specific signatures. The latter can be used to identify who the possible attackers are when a protected system is attacked by attacker-specific signatures.

5 Conclusions

In this paper, we implemented an Intrusion Detection System by applying genetic algorithm with Neural Network to efficiently detect various types of network intrusions. To implement and measure the performance of our system, we used the standard KDD99 benchmark dataset and obtained reasonable detection rate. The second stage of the model is Neural Network. After many experiment on the Neural Network using different training algorithms and object functions, We observed that Resilient back propagation with sigmoid function was the best one for classification therefore We used it in this work. And trail much architecture with one hidden layer and two hidden layers with different number of neurons to obtain the best performance of the Neural Network.

6 References

- [1] J., Muna. M. and Mehrotra M., "Intrusion Detection Systems : A design perspective", Proceeding of 2rd International Conference On Data Management, IMT Ghaziabad, India.,2009,265-372.
- [2] M. Panda, and M. Patra, "Building an efficient network intrusion detection model using Self Organizing Maps", Proceeding of world academy of science, engineering and technology, 38, 2009, 22-29.
- [3] M. Khattab Ali, W. Venus, and M. Suleiman Al Rababaa, "The Affect of Fuzzification on Neural Networks Intrusion Detection

System", IEEE computer society,2009, 1236-1241.

- [4] B. Mykerjee, L. Heberlein T., and K. Levitt N., "Network Intrusion Detection", IEEE Networks, 8(3), 1994, 14-26.
- [5] W. Jung K., "Integration Artificial Immune Algorithms for Intrusion Detection", dissertation in University of London, 2002, 1-5.
- [6] A. Chittur, "Model Generation for an Intrusion Detection System Using Genetic Algorithms", Technical Report, Ossining, New York, 2001.
- [7] W. Li, "Using Genetic Algorithm for Network Intrusion Detection", http://www.security.cse.msstate.edu, Department of Computer Science and Engineering, Mississippi State University, USA, 2004.
- [8] W. Lu, I. Traore, "Detecting New Forms of Network Intrusion Using Genetic Programming", Computational Intelligence, 20(3), Blackwell Publishing, Malden, 2004, 475-494.
- [9] M. M. Pillai, J. H. P. Eloff, H. S. Venter, "An Approach to Implement a Network Intrusion Detection System using Genetic Algorithms", Proceedings of SAICSIT, 2004, 221-228.
- [10] S. M. Bridges, R. B. Vaughn, "Fuzzy Data Mining And Genetic Algorithms Applied To Intrusion Detection", Proceedings of 12th Annual Canadian Information Technology Security Symposium, 2004, 109-122.