International Journal of Research in Social Sciences

(ISSN: 2249-2496)

CONTENTS

Chief Patron

Patron

Chief Advisors

Editorial Board

Editorial Advisory Board

Associate Editors

Title

VARIOUS TECHNIQUES IN INTRUSION DETECTION: A SURVEY

Author(s)

Jay Kant Pratap Singh

Department of Computer Science and Engineering

K.P.Engineering College, Agra, India-282005

Abhishek Kumar Sahu

Department of Computer Science and Engineering

K.P.Engineering College, Agra, India-282005

Abstract:

With the rapid expansion of internet computer systems are facing an enormous threat from external world. There are numerous approaches described in the literature during the recent year to maintain the information intact. Therefore intrusion detection is becoming important technology that identifies various network intrusions such as anomalous network behaviour, unauthorized network access and malicious attacker to computer system. In this paper we describe a various well known Intrusion detection techniques.

INTRODUCTION:

Any activity or set of activities that attempt to compromise the integrity, confidentiality or availability of a resource is known as Intrusion. The security of a computer system is one of the major challenges for researcher. Intrusion prevention alone is not sufficient to protect the computer from emerging threats. Traditional Intrusion prevention technique such authentication by using password and information protection by using encryption have been applied to protect computer systems but there are always some mechanism to exploitable weakness in the systems due to design and programming errors in the system software due to programming errors. So the intrusion detection along with prevention is also equally important for computer security. IDS may perform may be one of misuse detection, anomaly detection or Network intrusion detection system. Misuse detection relies on matching known patterns of hostile activity against databases of past attacks. They are highly effective at identifying known attack and vulnerabilities, but rather poor in identifying new security threats. Anomaly detection will search for something rare or unusual by applying statistical models or by applying artificial intelligence methods such as neural network to compare current activity against historical knowledge. Common problems with anomaly-based systems are that, they often require extensive training data for artificial learning algorithms, frequent updates. Network-based IDSs collect audit data from the network traffic. Network-based IDSs offer several advantages. First, network-based IDSs can take advantage of the standard structure of network protocols, such as TCP/IP. This is a good way to avoid the confusion resulting from heterogeneity in a distributed system. Second, network-based IDSs usually run on a separate (dedicated) computer; thus, they do not consume the resources of the computers that are being protected and most importantly it fits most the real time applications.

Any of these IDs scheme is not a silver bullet. So some of the researchers had moved to hybrid intrusion detection schemes.

REVIEW OF VARIOUS INTRUSION DETECTION METHODS:

A. *Rule based IDS.*

Rule-Based analysis relies on sets of predefined rules that are provided by an administrator, automatically created by the system, or both. Expert systems are the most common form of rule-based intrusion detection approaches. An expert system consists of a set of rules that encode the knowledge of a human "expert"[1]. These rules are used by the system to make conclusions about the security-related data from the intrusion detection system. Expert systems permit the incorporation of an extensive amount of human experience into a computer application that then utilizes that knowledge to identify activities that match the defined characteristics of misuse and attack. Unfortunately, expert systems require frequent updates to remain current[2]. Major problem:- (a) Rule-based systems suffer from an inability to detect attacks scenarios that may occur over an extended period of time. (b) Intrusion scenarios in which multiple attackers operate in concert are also difficult for these methods to detect because they do not focus on the state transitions in an attack, but instead this concentrate on the occurrence of individual elements. (c) Rule-based systems also lack flexibility in the rule-to-audit record representation. Slight variations in an attack sequence can affect the activity-rule comparison to a degree that the intrusion is not detected by the intrusion detection mechanism.

B. *Fuzzy Logic Based IDS*

Rule based system used restrict the natural ability of attribute to occurs in more than one cluster. For this reason and with the aid of fuzzy logic, fuzzy clustering can be employed to overcome the weakness. The membership of a pattern in a given cluster can be any be any value between 0 and 1[3]. In this model a data object belongs to the cluster where it has the highest membership value. The final output is probabilistic in nature. Step in Fuzzy based IDS are as follows (i) *Data collection:* - choosing the best data elements to monitor in the network stream is critical to the effectiveness of the intrusion detection system. Since Fuzzy is intended to on the network packet header data rather than the contents of network packets. Fuzzy concentrates on the three main internet protocols: TCP, UDP, and ICMP. Data reduction is critical when monitoring network data over a lengthy period. (ii) *Data Analysis and Profile Generation: - Once* enough data is collected over a two-week collection period, fuzzy inputs sets from each metric are produced [4]. In general, the extents and midpoints of the membership functions were determined with a fuzzy C-means algorithm. Though, as we shall see, some metrics produce sparse variation that may require more simple statistical models to define the sets. There are five membership functions in each input set: LOW, MED-LOW, MEDIUM, MED-HIGH, and HIGH.

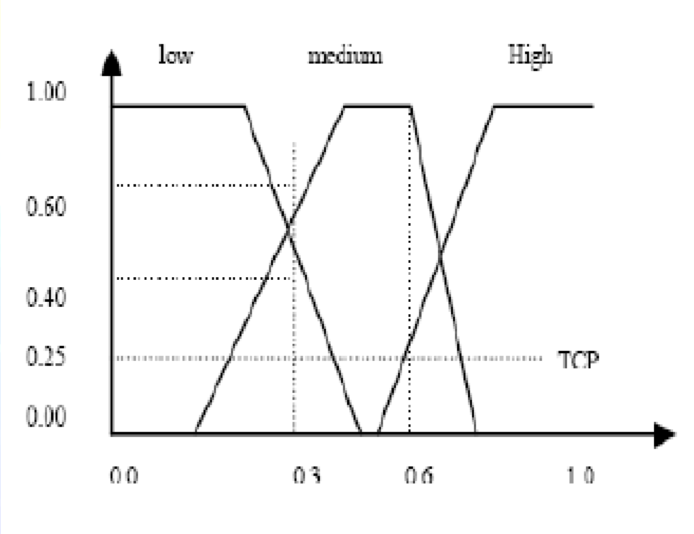

Figure: Fuzzy Membership function

(iii*) Fuzzy Rules*: - With the fuzzy input sets defined, the next step is to write the rules for detecting each type of attack. A collection of fuzzy rules with the same input and output variables is called a fuzzy system. We assume that the security administrator can use their expert knowledge to help create a set of rules for each attack the rules are created using the fuzzy system editor contained in the Matlab Fuzzy Toolbox.

Neural Network Based IDS:

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. An increasing amount of research in the last few years has investigated the application of Neural Networks to intrusion detection. [5]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, we 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 behaviour 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. 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. The

system is composed of eight modules, as shown in figure

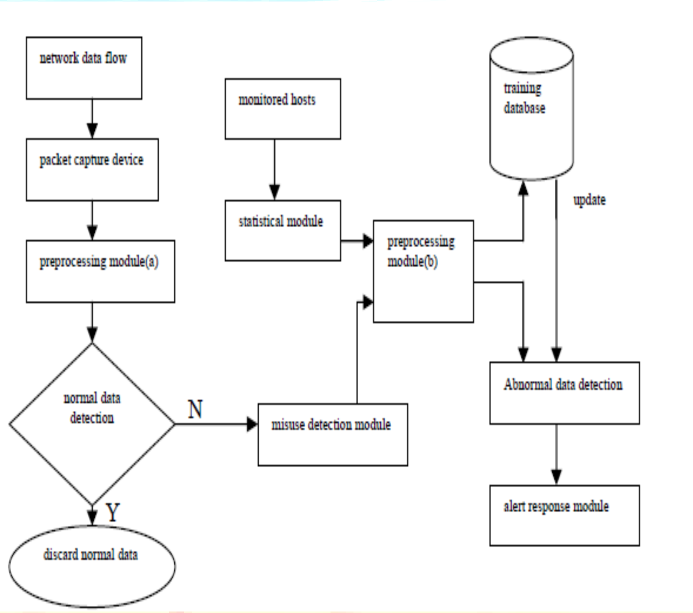

Figure: Neural Network based Ids[6]

(i*) Network packet capture device*: - It captures data packages from the network and decodes information. (ii) *Pre-processing module (a):*- The main aim of this module is to pre-process data based on extracted characteristics of normal data and also takes care of protocol analysis. First it converts binary audit data obtained from network into ASCII format. (iii) *Normal data detection module:* - To meet the demand of high-speed network, one class classifier is used to identify normal data. It uses back propagation algorithm and finally, the abnormal data separated are transferred to the misuse detection module. (iv) *Misuse detection module:*- It detects known attacks. First, it makes alarm when an attack is found. Then the attack packet information was detected and abnormal data are together transferred to pre-processing module (*b*). (v) *Statistical module*: - It takes care of data feature statistics based on destination hosts and services. Statistics based on destination hosts include: the number of connection same as the connected destination hosts at a time. Statistics based on services include: the number of connection same as the connected services at a time (vi) *Pre-processing module* (*b*):- In this module, data from misuse detection module and statistical module are implemented integrated processing, and finished connection record items. Then connection records containing attack information are transferred to training database of neural network, and the rest of connection records are sent to abnormal data detection module. (vii) *Abnormal data detection module*. It uses the same back propagation algorithm as the normal data detection module. The mainly difference between them are input data and training data. The former extracts only network connection feature and takes normal data as training data, but the latter processes feature statistics based on destination hosts and services, except for network connection feature and content feature, and attack data as its training data. (viii) *Alert response module*. A response and alert mechanism is implemented to give an alarm for aggressive intrusion activities detected and takes measures according to the response regulations which the users define. Pros and Cons of NNIDS is (a) High tolerance of noisy data as well as their ability to predict and classify patterns on which they have not been trained. (b) They are well-suited for continuous-valued inputs and outputs, unlike most decision tree algorithms (c) Neural network algorithms are inherently parallel; parallelization techniques can be used to speed up the computation process. (d) Because the output of a neural network is expressed in the form of a probability the neural network provides a predictive capability to the detection of instances of misuse (e) It is supervised learning contrary to GA. (f) other disadvantage of applying neural networks to intrusion detection is the "Black box" nature of the neural network (Black Box nature means -Acquired knowledge in the form of a network of units connected by weighted links is difficult for humans to interpret). (g) What could be the initial topology of neural network? (h) Accuracy of neural network depends on the choosing the initial weight. So what could be the right choice of initial weight to avoid local minima condition?

C. *Genetic Algorithm Based IDS*

It possibly over come almost all the problem that occurs in Neural network based IDS i.e. It is Unsupervised in nature and does not have problems mentioned in above (b) and (c) of Neural network based IDS[7]. The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution. Three main operator used in genetic algorithm[8]

Three main operators used in Genetic Algorithm are as follow

(1) Selection rules. Select the individuals, called parents, which contribute to the population at the next generation.

(2) Crossover rules. Combine two parents to form children for the next generation.

(3) Mutation rules. Apply random changes to individual parents to form children.

Steps of Genetic Algorithm:-

Step1):- Generate Initial individual population

Step 2):- Check the fitness of individual by using the Fitness function.

The individual which has better fitness has high probability to become into parents.

Xi is an individual having detection value f(Xi):

$$f(X_i)=\text{True Positive Rate (TPR)}-\text{False Positive Rate(FPR)}$$

Where TPR = the number of intrusion event which be detected / the number of total intrusion event.

FPR = the number of normal event which be detected / the number of total normal event.

To calculate the fitness of any individual Xi, We use the Fitness function F(Xi) as.

$$F(X_i) = f(X_i) - \sum_{i=1} F(X_i)/n +2$$

Selection of Fittest rules:- Selection Operator always ensures the best individual must be chosen i.e. the individual that has the highest selection probability. Selection Probability of Xi can be calculated as.

$$P(Xi) = F(Xi) / \sum_{i=n} F(Xi)$$

Step3) Crossover: - Crossover Operator randomly chooses a pairs individuals among those previously elected to breed and exchange substrings between them. The exchange occurs around randomly selected crossing points[9].

Let’s take an example of pair of individuals which the length is 16:

Parents 1: 1101111000110001;

Parents 2: 0011101011011111.

The position of crossover point is: 3; 8; 12.

After crossover operates the new individuals are:

Offspring 1: 1101101000111111;

Offspring 2: 0011111011010001.

Step 4: Mutation operator takes one string from the population and randomly alters some value within it.

Lets an an example of mutation of individual which has the length 16 as follow

Parent: 1111000011010110

Suppose algorithm randomly selects 6th bit to mutate then become.

Offspring: 1111010011010110

The above algorithm can be summarized more appropriately in terms of Flow diagram for better understanding as follow

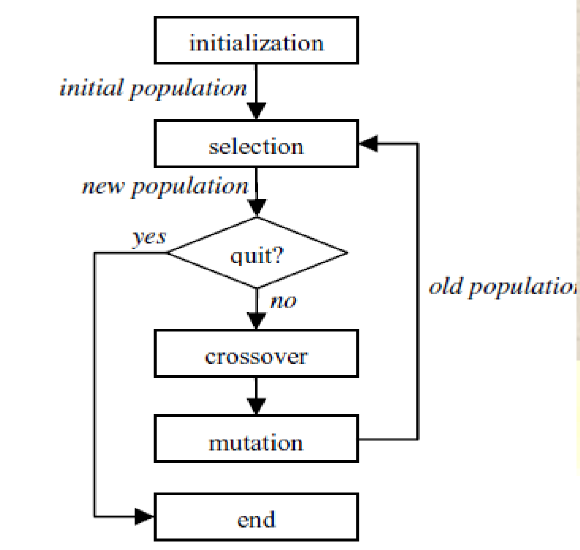

Figure: Genetic algorithm based Ids

CONCLUSION:

Rule based expert system forms rule using hard computing. Therefore slight variation in the attack cannot be detected by rule based. Moreover it require frequent updates. On the other hand Fuzzy logic generate rule that cover more vulnerability than rule formed by expert system using hard computing. Thirdly genetic algorithm is used to tune the rule and generate the necessary rules. Genetic algorithm is an unsupervised learning contrary to neural network approach. It eradicates the problem of topology of network that is inherent to neural network. In neural network we assume any network topology and initial weight also then perform operation using back propagation algorithm and check the output by some threshold value. If output is below the threshold then we again change network topology or weight or both network topology and weight. But in this approaches there is no question of assuming any network topology and weight and gives globally optimal solution.

ACKNOWLEDGMENT:

I will thank my teacher assistant professor Uday Pratap Rao, NIT Surat here for fervent help when I have some troubles in paper writing. I will also thank my class mates in laboratory for their concern and support both in study and life.

REFERENCES:

- http://en.wikipedia.org/wiki/Intrusion_detection_system
- Wang Yunwu "Using Fuzzy Expert System Based on Genetic Algorithms for Intrusion Detection System" Information Technology and Applications, IFITA , International Forum Vol 2, 2009.
- Norbik Bashah Idris, Bharanidharan Shanmugam "Novel Attack Detection Using Fuzzy Logic and Data Mining", International Conference of Soft Computing and Pattern Recognition, SOCPAR '2009. ..
- Ajith Abraham, Ravi Jain, Johnson Thomas, Sang Yong Han, "Distributed soft computing intrusion detection system", Journal of Network and Computer Applications,2007
- "Comparison of two feature selection methods in Intrusion Detection Systems," Seventh International Proceedings of the 7th IEEE International Conference on Computer and Information Technology, pp. 83-86 , 2007.
- Murkami, Honda, "Comparartive Study IDS method and Feed forward neural network": International Joint Conference on neural Network, IJCNN,2005.
- Ren Hui Gong, Mohammad Zulkernine, Purang Abolmaesumi "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/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks, 2005.
- Suhail Owais, Václav Snášel, Pavel Krömer, Ajith Abraham "Survey: Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques", 7th Computer Information Systems and Industrial Management Applications, 2008.
- B.A. Fessi, S. BenAbdallah, M. Hamdi, N. Boudriga "A New Genetic Algorithm Approach for Intrusion Response System in Computer Networks", IEEE 2009.