Anomaly Based Intrusion Detection System Using Artificial Neural Network and Fuzzy Clustering

Prof. D.P. Gaikwad, Sonali Jagtap, Kunal Thakare, Vaishali Budhawant

AISSMS College of Engineering Pune

Abstract

An intrusion detection system (IDS) is a security layer used to detect ongoing intrusive activities in information systems. Artificial Neural Networks (ANN) can be used to detect the intrusion in the system but there is slight complication that ANN lacks in certain areas that are detection precision for low frequent attacks and detection stability. So we have decided to implement FC-ANN approach based on ANN and fuzzy clustering, to solve the problem. The general procedure of FC-ANN is as follows: firstly fuzzy clustering technique is used to generate different training subsets. Subsequently, based on different training subsets, different ANN models are trained to formulate different base models. Finally, a meta-learner, fuzzy aggregation module, is employed to aggregate these results. In addition to this we are going to add restore point which allows for the rolling back of system files, registry keys, installed programs and the project data base etc.

1. Introduction

It is obvious that, in today’s era of information technology, the sharing of resources and information in interconnected network is essential. But as to secure this information from unauthorized uses and manipulation, it is necessary to impose some restrictions. There are some tools developed to do the same like Firewall, Anti-virus and Intrusion detection system.

Intrusion detection system is used for detection of intrusion in the system. The use of an intrusion detection system is becoming common due to the increase in complexity of attack and that of the computer systems themselves. Generally intrusion detection system works in pre-defined manner regardless of the implementation mechanism selected. These are some common steps followed by the intrusion detection system

1. Data is captured, often in the form of IP packets.

2. The data are decoded and transformed into a uniform format, through the process of feature extraction.

3. The data are then analysed in a manner which is specific to the individual IDS, and classified as threatening or not.

4. Alerts are generated if a threatening pattern is encountered [9].

The intrusion detection system has mainly two types: Signature-*based detection, anomaly-based detection.*

Signature-Based Detection:

It relies on accurate matching of system or network activity. This system can only detect an attack if there an accurate matching behaviour found against already stored patterns, known as signatures. *Snort* uses this type of detection. Snort is an open source IDS which implements a range of pattern matching algorithms of the input data and produces alerts based on the matching of the input to a signature base [9]. This type of system can restrict the false alarms from happening but Potential threats likely to get missed as new techniques of attacks keeps on evolving. Besides that the cost of maintenance is too much as the signature set needs to keep on upgrading.

Anomaly-Based Detection:

Anomaly-based detection is the strongly recommended technique in today’s scenario where new attacks are being discovered each day. This technique evolves itself by understanding and gathering the information about the system and determines the behaviour of the system based on it [10]. There are three types of algorithms basically used to develop the system and that are fuzzy logic, genetic algorithm, artificial neural networks. Novel attacks can be detected in this type of detection system unlike the signature based detection system.

In our paper we propose to develop anomaly based detection system containing both the fuzzy logic and artificial neural networks. Among these techniques, Artificial Neural Network (ANN) is one of the widely used techniques and has been successful in solving many complex practical problems. And ANN has been successfully applied into IDS. However, the main drawbacks of ANN-based IDS exist in two aspects: (1) lower detection precision and (2) weaker detection stability. To solve these two problems, we propose a novel approach for ANN-based IDS, FC-ANN, to enhance the detection precision for low-frequent attacks and detection stability.

The fuzzy clustering is being used for segregation of the datasets into different clusters and fuzzy aggregation is used to aggregate different ANN's results. ANN (Artificial Neural Network) utilized for to learn the pattern of every subset and decide certain actions on it. These modules are explained elaborately in section on proposed model.

2. Literature Survey

In paper [1], author has developed an anomaly based intrusion detection system in detecting the intrusion behavior within a network. A fuzzy decision-making module was designed to build the system more accurate for attack detection, using the fuzzy inference approach. An effective set of fuzzy rules for inference approach were identified automatically by making use of the fuzzy rule learning strategy, which are more effective for detecting intrusion in a computer network. In this paper, at first, the definite rules were for attack data as well as normal data. Then, fuzzy rules were identified by fuzzifying the definite rules and these rules were given to fuzzy system, which classify the test data. In this paper, they also used the F-test statistical methodology to find out the determine group of trial which differs significantly from an expected value. They used KDD cup 99 dataset for evaluating the performance of the proposed system and experimentation results showed that the proposed method is effective in detecting various intrusions in computer networks.

In paper [2], the author has proposed a new fuzzy tree interface algorithm for color correction. It is very simple and efficient and is suitable for adopting the center-average method for defuzzification to obtain better color quality. Therefore fast and cost-efficient implementation with the good quality correction effects for tree-based fuzzy color correction is achieved. It is simple, and suitable for using the center-average method for defuzzification, by which a better color correction effect is obtained, keeping low computational.

In the paper [3], they have designed fuzzy logic-based system for effectively identifying the intrusion activities within a network. The proposed fuzzy logic-based system can be able to detect an intrusion behavior of the networks since the rule base contains a better set of rules. Here, they used automated strategy for generation of fuzzy rules, which were obtained from the definite rules using frequent items. The experiments and evaluations of the proposed intrusion detection system are performed with the KDD Cup 99 intrusion detection dataset. The experimental results show that the proposed system achieved higher precision in identifying whether the records are normal or attack one.

In paper [4], author has shown proposed model that gives better results for DoS, Probe, U2R and R2L all types of attacks .Their proposed algorithm have higher precision and recall rate. This paper has introduced the new methods for cluster to class mapping which increases the accuracy of the model for all types of attacks. From that paper, it has been also observed from results that as value of k changes than corresponding precision and recall also increase and decreases. Accuracy of k-mean clustering depends upon the value of k.

In paper [5], author examines the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. Cluster analysis groups data so that points within a single group or cluster are similar to one another and distinct from points in other clusters. Clustering has been shown to be a good candidate for anomaly detection. The purpose of that study is to examine the use of clustering technology to automate fraud filtering during an audit. They use cluster analysis to help auditors focus their efforts when evaluating group life insurance claims. Claims with similar characteristics have been grouped together and small-population clusters have been flagged for further investigation. Some dominant characteristics of those clusters which have been flagged are large beneficiary payment, large interest payment amounts, and long lag between submission and payment.

In paper [6], author presents a novel intrusion detection system (IDS) that models normal behaviors with Hidden Markov Models (HMM) and attempts to detect intrusions by noting significant deviations from the models. Among several soft computing techniques neural network and fuzzy logic are incorporated into the system to achieve robustness and flexibility. Self-organizing map (SOM) determines the optimal measures of audit data and reduces them into appropriate size for efficient modeling by HMM. Based on several models with different measures, fuzzy logic makes the final decision of whether current behavior is abnormal or not. Experimental results with some real audit data show that the proposed fusion produces a viable intrusion detection system. Fuzzy rules that utilize the models based on the measures of system call, file access, and the combination of them produce more reliable performance.

In paper [7], the author has proposed an intrusion detection model based on hybrid fuzzy logic and neural network. The key idea of author in this paper is to take advantage of different classification abilities of fuzzy logic and neural network for intrusion detection system. The 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. Training and testing data were obtained from the Defence Advanced Research Projects Agency (DARPA) [11] intrusion detection evaluation data set. This paper addresses the problem of generating application clusters from the KDD cup 1999 network intrusion detection dataset.

3. Our Proposal

In our project, we are using the same FC-ANN approach which is based on ANN and fuzzy clustering with the addition of system restores point. System Restore is a component that allows for the rolling back of system files, registry keys, installed programs and the project data base etc. which is stored in the cloud server, to a previous state in the event of system malfunction or failure of the system or if any Intrusion is detected on the system.

Through fuzzy clustering technique, the heterogeneous training set is divided to several homogenous subsets. Thus complexity of each sub training set is reduced and consequently the detection performance is increased. The experimental results using the KDD CUP 1999 dataset demonstrates the effectiveness of our new approach especially for low-frequent attacks, i.e., R2L and U2R attacks in terms of detection precision and detection stability. Data mining techniques, such as support vector machine, evolutionary computing, outlier detection, genetic algorithm may be introduced into IDS. Comparisons of various data mining techniques will provide clues for constructing more effective hybrid ANN for detection intrusions.

3.1 Framework of FC-ANN

In this section, we elaborate our approach FC-ANN. We firstly present whole framework of the approach. Then we discuss three main modules, i.e., *FUZZY CLUSTERING* module, *ANN* module, *FUZZY AGGRETATION* module and we add Restore point for the system.

3.2 Framework of IDS based on ANN and Fuzzy Clustering:

FC-ANN firstly divides the training data into several subsets using fuzzy clustering technique. Subsequently, it trains the different ANN using different subsets. Then it determines membership grades of these subsets and combines them via a new ANN to get final results. The whole framework of FC-ANN is illustrated in figure (A). As typical machine learning framework; FC-ANN incorporates both the training phase and testing phase.

Project Overview:

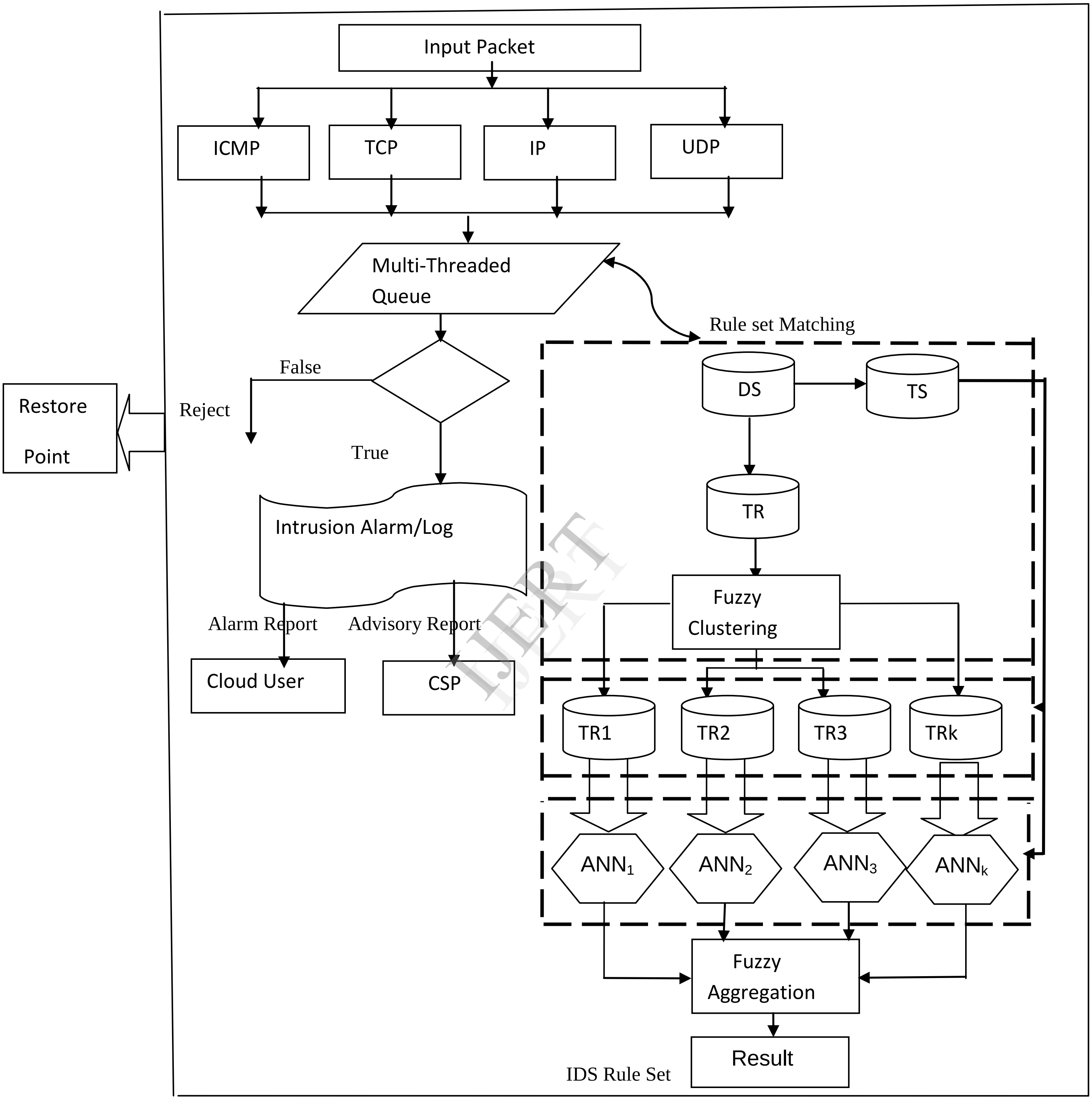

Figure (A): Framework IDS using FC-ANN and Restore Point Facility Model

3.3 Fuzzy clustering module

The aim of fuzzy cluster module is to partition a given set of data into clusters, and it should have the following properties: homogeneity within the clusters, concerning data in same cluster, and heterogeneity between clusters, where data belonging to different clusters should be as different as possible. Through fuzzy clustering module, the training set is clustered into several subsets. Due to the fact that the size and complexity of every training subset is reduced, the efficiency and effectiveness of subsequent ANN module can be improved. There are two types of clustering techniques hard clustering techniques and soft clustering techniques. Beside Partition of training set, we also need to aggregate the results for fuzzy aggregation module. Therefore, we choose one of the popular soft clustering techniques, fuzzy c-means clustering, for fuzzy clustering module.

3.4 ANN module

ANN module aims to learn the pattern of every subset. ANN is a biologically inspired form of distributed computation. It is composed of simple processing units, and connections between them. In this study, we will employ classic feed-forward neural networks trained with the back-propagation algorithm to predict intrusion. A feed-forward neural network has an input layer, an output layer, with one or more hidden layers in between the input and output layer.

3.5 Restore Point for system backup:

Our new approach in this project is adding a system restore point in the paper [8]. System Restore is a component that allows for the rolling back of system files, registry keys, installed programs and the project data base etc. which is stored in the cloud server, to a previous state in the event of system malfunction or failure of the system or if any Intrusion is detected on the system.

In System Restore interface is based on Shadow Copy technology. In prior period it was based on a file filter that watched changes for a certain set of file extensions, and then copied files before they were overwritten. Shadow Copy has the advantage that block-level changes in files located in any directory on the volume can be monitored and backed up regardless of their location.

In System Restore, the user may create a new *restore point* manually, roll back to an existing restore point, or change the System Restore configuration. Moreover, the restore itself can be undone. For the IDS this can provide restore points covering the past several weeks. ADMIN concerned with performance or space usage may also opt to disable System Restore entirely. In our project the Files stored on volumes not monitored by System Restore are never backed up or restored.

System Restore backs up system files on the server on which the overall IDS training database is saved and saves them for later recovery and use it again when the system is restored and apply the ANN for the same kind of attack is detected again on the system. It also backs up the registry and drivers which are install on the cloud server.

3.5.1 Backup strategy used for the system

Any backup strategy starts with a concept of a data repository. The backup data needs to be stored somehow and probably should be organized to a degree. It can be as simple as a sheet of paper with a list of all backup tapes and the dates they were written or a more sophisticated setup with a computerized index, catalogue, or relational database. Different repository models have different advantages.

3.5.2 Unstructured

An unstructured repository may simply be a stack of floppy disks or CD-R/DVD-R media with minimal information about what was backed up and when. This is the easiest to implement, but probably the least likely to achieve a high level of recoverability.

3.5.3 Full only / System imaging

A repository of this type contains complete system images from one or more specific points in time. This technology is frequently used by computer technicians to record known good configurations. Imaging is generally more useful for deploying a standard configuration to many systems rather than as a tool for making ongoing backups of diverse systems.

3.5.4 Incremental

An incremental style repository aims to make it more feasible to store backups from more points in time by organizing the data into increments of change between points in time. This eliminates the need to store duplicate copies of unchanged data, as would be the case with a portion of the data of subsequent full backups. Typically, a *full* backup (of all files) is made which serves as the reference point for and incremental backup set. After that, any number of *incremental* backups is made. Restoring the whole system to a certain point in time would require locating the last full backup taken previous to the data loss plus each and all of the incremental backups that cover the period of time between the full backup and the point in time to which the system is supposed to be restored. Additionally, some backup systems can reorganize the repository to synthesize full backups from a series of incremental.

3.5.5 Differential

A differential style repository saves the data since the last full backup. It has the advantage that only a maximum of two data sets are needed to restore the data. One disadvantage, at least as compared to the incremental backup method, is that as time from the last full backup (and, thus, data changes) increase so does the time to perform the differential backup. To perform a differential backup, it is first necessary to perform a *full* backup. After that, each differential backup made will contain all the changes since the last full backup. Restoring an entire system to a certain point in time would require locating the last full backup taken previous to the point of the failure or loss plus the last differential backup since the last full backup.

4. Conclusion

In this project, we propose an intrusion detection approach, called FC-ANN, based on ANN and fuzzy clustering. In fuzzy clustering technique, the heterogeneous training set is divided to several homogenous subsets. Thus complexity of each sub training set is reduced and consequently the detection performance is increased. And in this way system become more efficient and stable and we efficiently overcome the drawbacks –lower detection precision, weaker detection stability. And along with this we can successfully take backup of system using restore point facility.

5. References:

[1] *R. Shanmugavadiva, Dr. N. Nagarajan*, "Learning of Intrusion Detector in Conceptual Approach of Fuzzy Towards Intrusion Methodology", IJARCSSE, Vol. 2, May 2012.

[2] *Jer Min Jou, Shiann Rong Kuang and Ren Der Chen* "A New Efficient Fuzzy Algorithm for Color Correction", IEEE transaction on circuit and system-Fundamental Theory and applications, vol. 46, No. 6, June 1999

[3] *R. Shanmugavadiva, Dr. N. Nagarajan*," Network Intrusion Detection System Using Fuzzy Logic", R. Shanmugavadiva et al./Indian journal of Computer Science and Engineering(IJCSE11).

[4] *Kusum Kumari Bharati, Sanyam Shukla, Sweta Jain*,"Intrusion detection using clustering", International Conference [ACCTA-2010], August 2010.

[5] *Sutapat Thiprungsri, Miklos A. Vasarhelyi*, "Cluster Analysis for Anomaly Detection in Accounting Data: An Audit Approach1", The International Journal of Digital Accounting Research Vol. 11, 2011, pp.69-84

[6] *Sung Bae Cho* Member IEEE "Incorporating Soft Computing Techniques into a Probabilistic Intrusion Detection System", IEEE transaction on circuit and system-Fundamental Theory and applications—Part C Applications and Reviews, vol. 32, No. 2, MAY 2002.

[7] *Muna Mhammad T.Jawhar, Monica Mehrotra*, "Design Network Intrusion Detection System using hybrid Fuzzy-Neural Network", International Journal of Computer Science and Security, Vol. 4.

[8] *Gang Wang, Jinxing Hao, Jian Ma, Lihua*, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering", Expert System with application-2010

[9] *Julie Greensmith and Uwe Aickelin "*Firewalls, Intrusion Detection Systems and Anti-Virus Scanners" ASAP Group, University Of Nottingham, UK , June 21,2004

[10]*Zhenwei Yu, Jeffrey J. P. Tsai, Fellow, IEEE, and Thomas Weigert* "An Automatically Tuning Intrusion Detection System" ,IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 37, NO. 2, APRIL 2007 373

[11] www.darpa.mil/