A Novel Intrusion Detection System Using Neural-Fuzzy Classifier for Network Security

Nirmal Ratnawat¹, Prof. Anurag Jain²

¹M-Tech Research Scholar, ²HOD, Department of CSE RITS, Bhopal

Abstract—An intrusion detection system (IDS) is intelligent engine which is used to identify anomalies (abnormality) in a computer network system. Intrusions are generally unauthorized access signatures which is like a attacks for network that attempts to either prevent normal functioning of nodes in a network or to harm the nodes with malicious scripts or to sniff unauthorized packets. Intrusion detection system (IDS) reports alarms when it detects any attacks (abnormal) condition. IDS is working in dynamically changing environment. It works by detecting a signature from current session of packet transfer and matching it with database intrusion signatures. Basically IDS consists of prediction engine which analyzes data and outputs the prediction on the data. By seeing predictions system operator is able to know that data record is normal or is affected by any attack. Therefore prediction engine is the heart intrusion detection system. Conventionally IDS are complex string matching automata engine which relies on extracting string signature from set of packet fields like IP address, Port number, round trip time etc. However attacker continuously changes the attack signature which makes it difficult to detect the intrusion. Therefore knowledge based classifiers were developed for Intrusion detection system. A knowledge based classifier is one that builds a mathematical model from the signature database with a prior knowledge about the class of the signature (NORMAL, ABNORMAL) and classifies any input signature as one of the known classes by evaluating the incoming parameters in trained model. Such techniques are proven to be more accurate than string matching based automata engines. However conventional machine learning based techniques suffers from training problem. Network intrusion signatures are ever changing. Hence the training model needs to be frequently updated. Therefore any new attack signature would require entire set of the training data to be retrained. In this work we develop a unique machine learning based intrusion detection system for solving the aforementioned problem. In this work we propose a Fuzzy-Neuro system for IDS. First we create prediction model prediction engine analyzes and data according to prediction model. System operator verifies results and marks false predictions. Only false predictions are fed back to model tuner to tune model automatically. Evaluate the performance of the system with KDD-Cup dataset. Results shows that proposed system produces 100% accuracy intern of recognition and classification accuracy.

Keywords—Intrusion detection system (IDS), Fuzzy C-Means and neural classifier

I. INTRODUCTION

Due to heavy use of the Internet, the importance of information protection has been increased. As such, there is no defined method of interjection deduction. It can be considered as pattern matching problem that differentiate between n/w attack, abnormal n/w behaviors, and normal network behaviors. Any Group of activities which try to compromise on the reliability, accessibility, confidential or privacy of resources is known as interjection or interruption. An intruder is an attacker, person or group of people who initiates the activities during interruption. Attackers can be from within the trust network, person who have the access to use system with normal user rights or someone who uses a hole in some OS to escalate their access level or admin rights. It can be from outside of the system or network that is someone on another network or even in some other country who exploits a weakness, vulnerability in an insecure network service on the system to take unauthorized entry and access of the trust network.

Intrusion recognition systems are in fact a layer of security used to notified continuing interfering activities in some information systems. Conventionally, to avoid this dependency, a list of learning techniques and data mining techniques has been identified for intrusion discovery. Intrusion discovery high volume depends on widespread knowledge of safety experts, in particular, on their knowledge with the processor system that is to be sheltered.

Information security consist the following.

1) Protection- It protects systems to avoid security breaches which are generated by intruders.
2) Detection- Intrusions detected by Security violations as soon as they occur.
3) Reaction – It perform automatic alarming whenever system is intruded such as pursuit of hackers.
4) Recovery – It repairs system automatically in case of any damage happened by intrusion.

Intrusion detection systems are classified into either Model based or Automata based system. Model based systems are widely popular because of their ability to classify absolutely new data with utmost accuracy. For Example of Knowledge based classifiers are: KNN, SVM, Neural Network, Fuzzy system.
II. RELATED RESEARCH WORK ON IDS

The field of intrusion detection system in network security has been developed for 30 years. A number of methods and techniques have been proposed and many systems have been affected by variety of intrusions. The various techniques used to detect the intrusions are data mining, neural network, and statistical methods. In this related work, the various methods and techniques are discussed.

Aneetha and S. Back proposed an approach in [1] to detect an attack, which uses neural networks and k-means clustering algorithms. The proposed approach is applied to the KDD CUP’99 data set. Average detection rate was 97.5%.

Monaranjan and Sudhir Kumar proposed an approach in [2] to detect an attack, which uses artificial neural networks and Back propagation algorithms. The proposed approach is applied to the KDD CUP’98 data set. Classification rate of 98%. Denial of Service attack type called Neptune went by undetected, and this shows that more testing is needed to find out why this happened.

Ghosh and Schwartz bard used system behavior as input for the neural network in their experiment [3]. They used 4 weeks of training data and 161 sessions with testing data. Out of these, they only had 22 attack sessions. The results from this experiment showed a detection rate of 77.3% and a false positive rate of 3.6%.

In their experiments, they used software behavior for the input to the neural network, not user behavior as we used in our experiments.

III. PROPOSED METHODOLOGY

3.1 Feed Forward Neural Network (FFNN)

A multi layered feed forward network to be made up of a fixed multilayer layers and each layer hold number of units. The feed for-forward neural networks provide a basic framework for representing non-linear functional mapping between input data and output variables. Output, Input and other hidden layers receives its inputs directly from previous layer and sends its results directly to units in the next layers. Whereas, in Recurrent network contains feedback information and there are no connections from any of the units. Each and every unit only acts as an input to immediate next layer. Theoretically, this type of networks is easy to analyze compare to other general topologies because their outputs can be represented with explicit functions of the inputs and weights.

3.2 Fuzzy C-Means Clustering

Fuzzy techniques can be applied to large data set that are numerical (quantitative), categorical (qualitative), or a mix-up of both. The clustering of volume data is considered. The data or information is observations of some physical process.

Many clustering algorithms are based on improving the basic c-means objective function. Hence we start discussion with presenting the fuzzy c-means functional.

\[
J_m = \sum_{k=1}^{K} \sum_{j=1}^{C} u_{kj}^m \| x_i - c_j \|^2, \quad 1 \leq m < \infty
\]

Here,

- \( m \) – Any real number greater than 1.
- \( u_{kj} \) - the degree of membership of \( x_i \) in the cluster \( j \)
- \( x_i \) - ith of d-dimensional measured data
- \( \| \cdot \| \) - Any norm expressing the similarity between any measured data and the center.

3.3 Basic Architecture Of IDS

One of the methods of developing a network protect is to describe network nature structure that point out offensive use of the system network and also look for the occurrence of those classifier patterns. So basically intrusion detection system is process based system which have intelligent engine and it is monitoring the behavior of system in given environment and detect, whether these activities are normal (legitimate) or intrusion (abnormal).

3.3.1 Data Sampling

The begging step in collecting data is to find exactly which type of data should be collected. Because of the aim of this project is going towards intrusion detection at the network level, a natural network choice for data transmission is the network transmission packet. The network system give us two types of information to study: user information and transport information, but for this only transport information is required. Transport data information have a structured pair of source ip and destination ip.
It also consists of some type of checksum on which the integrity of a packet is determined. Transport information is added to the packet as a part of the network transmission protocol. Transport information which cannot be made deceptive by fraudulent user is called as unbiased data.

As we do not have access to large scale network and IDS system for logging intrusions for generating of samples, we opt for KDD cup dataset. This dataset presents several thousand signatures which contains forty different fields from packets and the signature type (Attack or normal). Further the dataset also presents type of attacks.

3.3.2 Data Preprocessing:

It is already discussed that the fundamental principle of Machine learning based IDS system is to develop a mathematical or numerical model from string data.

Following is a sample snapshot of KDD data. As can be seen that several fields are numeric in nature. Hence our system first generates a matrix from. Some fields are repeated texts: like the protocol field (C), Transport type (B) and attack type (AP).

We first extract the unique values from the dataset for each of the columns and then replace the string names with those of their index number in the corresponding unique group.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |

The resultant set is a matrix of numbers.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

We extract 10,000 data for our experiment. 7000 data is randomly selected for training and 3000 for testing. A class contains numbers between 1-6 that represents the category of the signature.

3.3.3 Classifier system

The classifier scheme is a similar, law and policy based message passing system. All policies are of the type action (Permit, deny, rejected) form. If the rule is satisfied Then action form is receipt of the messages and the action is the sending of messages. All messages hold a tag specifying their source, destination and an extra information field.

We first cluster the dataset using Fuzzy C means clustering. We extract some sample data from every cluster, then we train neural network with this data, followed by classification of the samples.

As Neural network, SVM and other machine learning frameworks provide define result like class normal or abnormal, but for new intrusions it cannot be said with certaininity. Statement like processed data is probably x% normal, y1% smurf, y2% perl attack and so on.
Fuzzy system (classifier) defines output in probability. i.e. it classifies input as set of all probabilities belonging to all classes with highest priority given to highest percent. This data when superimposed with neural output produces a strong prediction engine.

System Design

: System Operator

\[\text{Standard KDD Dataset} \rightarrow \text{Rules set}\]

\[\text{Preprocessed dataset} \rightarrow \text{Labeled dataset}\]

\[\text{Trained dataset & weights} \rightarrow \text{Rules Set}\]

\[\text{Data set, Rules Set} \rightarrow \text{False prediction & confidence}\]

: System

\[\text{Initial Rules} \rightarrow \text{Training dataset}\]

\[\text{M/C Learning & Classifier}\]

\[\text{Prediction engine}\]

: Main UI

\[\text{Input packet from Network}\]
IV. EXPERIMENT AND RESULTS

For the implementation of the intrusion detection using fuzzy-neuro system we have used MATLAB 7.10. In our test, we have randomly import the data from KDD CUP 99 data set and dividing the dataset into two dataset, training and testing. We have first normalized the data from strings and then we have done labeling of the data with the help of fuzzy c-means algorithm. Experiments shows that proposed System produces 100% accuracy in comparison to Neural classifier (98%) and Fuzzy based system (70%).

V. COMPARISON RESULT WITH PREVIOUS WORKS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification rate</td>
<td>97%</td>
<td>95%</td>
<td>98%</td>
<td>99.66%</td>
<td></td>
</tr>
</tbody>
</table>
VI. CONCLUSION AND FUTURE SCOPE

In this work we have developed a machine learning based solution for IDS by combining fuzzy based rules with neural based classifier. Fuzzy rules provides a probability outcome of a signature which is then classified by neuro system. Experiments shows that proposed System produces 100% accuracy in comparison to Neural classifier (99.66%) and Fuzzy based system (70%). This is because a strong fuzzy rule helps neuro system to optimize the hidden layer which can then strongly classify the input signature pattern. The system can be further improved by reducing classification time through more optimization in feature selection and filtering stage.

REFERENCES


Author’s Profile

Nirmal Ratnawat is research scholar at Radharaman Institute of Technology and Science, under Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal. He is pursuing M. Tech. in Computer Science Engineering. He has keen to work on Intrusion Detection System Using Neural-Fuzzy Classifier.