Intrusion Detection System with FGA and MLP Algorithm

Miss. Madhuri R. Yadav
Department Of Computer Engineering
Siddhant College Of Engineering, Sudumbare
Pune, India

Prof. Prashant Kumbharkar
Department Of Computer Engineering
Siddhant College Of Engineering, Sudumbare
Pune, India

***Abstract*— Normally, The Internet grows rapidly and most useful in each domain but network vulnerability and intrusions are still an important issue that causes attacks. Attacks can immediately cause system down. Therefore, it is necessary to detect network attacks before they damage the whole system for that we used firstly dataset. Generally, Intrusion detection system can be deployed to detect network threats and attacks.**

A good system to detect the illegal user is to monitoring the packets and using the different algorithms, methods and applications which are created and implemented to solve the problem of detecting the attacks in intrusion detection systems. Most methods detect attacks and categorize in two groups, normal or threat. We consider network intrusion detection using fuzzy genetic algorithm to classify attacks in the datasets. Fuzzy rule is a machine learning algorithm that can classify network attack data and protect the system from damage, while a genetic algorithm is an optimization algorithm that can help finding appropriate fuzzy rule and give the optimal solution. And also a new approach of intrusion detection system based on neural network. In this paper, we have a Multi-Layer Perceptron (MLP) is used for intrusion detection system, which is better solution for the intrusion detection using weka tool. This algorithm uses the number of layers so it is more secure from the hacker.

We consider both well-known KDD99 dataset and our own network dataset. The KDD99 dataset is a benchmark dataset which is already stored. While our network dataset is an online network data captured in actual network environment. We evaluate our IDS in terms of detection speed, detection rate and false alarm rate.

***Keywords*— Fuzzy genetic algorithm; Multilayer perceptron algorithm; Intrusion detection; Network security**

I. INTRODUCTION

Nowadays, Internet grows rapidly but intrusion detection is an important issue in computer security. Information technology has become a critical component in the organization that manages the huge amount of data. Securing those systems from different actions should be the main goal when applying security, but the evolution of technologies makes this task very difficult.

Information security is dependent on the things like "Protection, Detection, Reaction, and Recovery". Intrusion detection is a crucial part of information security. IDS are the software or hardware tools that automatically scan the events that take place in network, looking for intrusion. Any activity aimed at disrupting a service or making resource unavailable or gaining unauthorized access can be termed as an intrusion.

IDS can be deployed to detect the attack. Generally, IDS necessary to detect the network attacks before they damage the whole system. Organization uses IDS system for different purpose, such as identifying problems with security policies, threats and individual from violating security policies. IDS typically record the information related to observed events, Notify security administrators of important observed events, and produce report and uses the several response techniques which involve the IDS stopping the attack and protect the system.

A. Definition and Categories

There is no standard and actual definition of intrusion detection. Intrusion detection is the detection of network behaviors that violate network security. Intrusion detection is distinguishing between network attacks and normal network behaviors or further distinguishing between different categories of attacks. Different types of IDS are available like "Host-based, Network-based, Stack-based Intrusion Detection System [1].

1) Host-based intrusion detection

It consists of an agent on a host that identifies intrusions by analyzing system calls, application logs, file-system modifications are binaries, password files, capability databases, Access control lists.[1] and other host activities and state. In a HIDS, sensors usually consist of a software agent. Some application-based IDS are also part of this category.

2) Stack-based intrusion detection

This type of system consists of an evolution to the HIDS systems. The packets are examined as they go through the TCP/IP stack and, therefore, it is not necessary for them to work with the network interface in promiscuous mode. This fact makes its implementation to be dependent on the Operating System that is being used.

3) Network-based intrusion detection

Network intrusion detection system is an independent platform that identifies intrusions by examining network traffic and monitors multiple hosts, developed in by Pete R. Network intrusion detection systems gain access to network traffic by connecting to a network hub, network switch configured for port mirroring, or network tap[1]. In a NIDS, sensors are located at choke points in the network to be monitored, often in the demilitarized zone (DMZ) or at network borders. Sensors capture all network traffic and analyze the content of individual packets for malicious traffic. An example of a NIDS is Snort.

B. *Related Work*

Most classification techniques for intrusion detection can be based on various classification algorithms. They are classified into two groups, which are grouped into supervised learning approach and Unsupervised learning approach.

1) Supervised learning-based approach

In supervised learning approach, the instances consist of input attributes and desirable output and the algorithm would produce an inferred function, which is called a classifier or regression function. This approach has high accuracy, low false- alarm with fast computing time. Supervised learning methods for intrusion detection can only detect known intrusions. In these methods from machine learning and pattern recognition have been utilized to detect intrusions.

For supervised learning for intrusion detection, there are mainly supervised neural network (NN)-based approaches [3], [4], and support vector machine (SVM)-based approaches [5].

a) NN-based approaches

We propose an NN for distinguishing between intrusions and normal behaviors. They unify the coding of categorical fields and the coding of character string fields in order to map the network data to an NN. They use execution numbers of system calls in a host machine as the features of network behaviors to train the NN. They propose an approach for intrusion detection using hierarchical NNs and use evolutionary NNs to detect intrusions.

b) SVM-based approaches

SVMs to distinguish between normal network behaviors and intrusions and further identify important features for intrusion detection. They propose the TreeSVM and ArraySVM for solving the problem of inefficiency of the sequential minimal optimization algorithm for the large set of training data in intrusion detection. They propose an approach for online training of SVMs for real-time intrusion detection based on an improved text categorization model.

2) Unsupervised learning-based approach

Unsupervised learning methods can detect the intrusions that have not been previously learned. Examples of unsupervised learning for intrusion detection include *K*-means-based approaches and self-organizing feature map (SOM)-based approaches [6], [7].

a) K-means-based approaches

K-means-based clustering algorithm, which is named Y - means, for intrusion detection. Combine the fuzzy K-means method and a clonal selection algorithm to detect intrusions. They use the incremental clustering algorithm that is an extension of the K-means algorithm to detect intrusions.

b) SOM-based approaches

Extract features that describe network behaviors from audit data, and they use the SOM to detect intrusions. They propose a hierarchical SOM approach for intrusion detection. Specific attention is given to the hierarchical development of abstractions, which is sufficient to permit direct labeling of SOM nodes with connection type. They propose a hierarchical SOM for intrusion detection. They use the classification capability of the SOM on selected dimensions of the data set to detect anomalies.

Previous discussions review the related work. Different approaches for intrusion detection have the following:

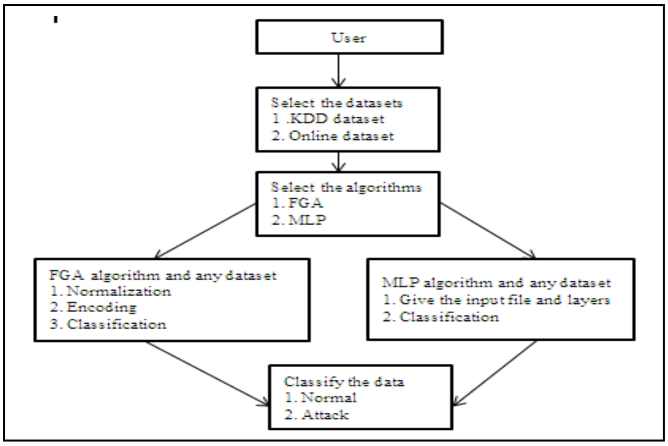
J. Gómez and E. León [8] proposed fuzzy and genetic algorithm to classify attack and normal data. The input data is KDDCup99 dataset and classify the data into 5 classes including DoS, Probe, R2L, U2R and Normal. T.P. Fries [9] proposed a fuzzy genetic algorithm approach. They used clustering algorithm and genetic algorithm to find significant attributes in KDD99 dataset and for detection used fuzzy GA algorithm. T. Komviriyavut et al [10] proposed a method to preprocess dataset in actual network environment. They used a decision tree algorithm to classify data are DoS, Probe and Normal. M.-Y. Su et al. [11] proposed a Real-time IDS for large-scale attacks by using fuzzy association rules then, each record will be sent to another computer in order to update new rule and able to detect only DoS attack. P. Kachurka and V. Golovko [12] proposed a neural network approach to real-time network intrusion detection; they collected the network traffic by using an open source intrusion detection system, able to detect unknown attack in real time. Aiming as constructing the "Fuzzy genetic algorithm and Multilayer perceptron algorithm" for the IDS which are supervised learning approach. Which has high accuracy, low false- alarm with fast computing time

In summary, there are many different algorithms for network intrusion detection. Most of them considered well-known KDD99 dataset but a few of them consider recent network dataset. Therefore, we propose to use a FGA and MLP algorithms to detect online network data and on KDD99 dataset. We captured online data traffic store in to file, and preprocess it into records. We evaluate our detection results with the results obtained from using the labeled KDD99 dataset.

The rest of this paper is presented as follows. In section II we represent the overview of the IDS system. In section III, We discuss on research methodology with algorithm, in section IV we present the primary results with dataset preparation and performance evaluation criteria. In section V we give a conclusion of this research work and future work.

II. OVERVIEW OF OUR SYSTEM

According to the characteristics of the FGA and MLP algorithms, network intrusion detection problem the overview of system".



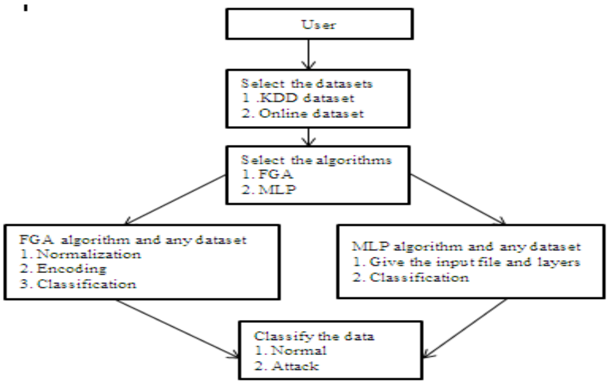


Fig. 2 Block diagram of the system

The whole system is implemented as shown in the Fig. 1, which is describing as follow:

1. User creates the online network dataset using the snort tool.
2. User selects the datasets that is KDD99 or the online dataset from the list which is given by the system.
3. Then user select the algorithm FGA or the MLP.
4. If the FGA is used then the data is prepressed into the selected attributes, then apply the normalization and the encoding on that string, create the rules using fuzzy rule and check the trained dataset with test dataset and classify the attacks.
5. If the MLP is used then then input file is given to the weka tool, and gives the some input parameter as the number of layers, then classify the data into attack and normal from used datasets.
6. Thus we classify the data sets into attacks and normal.

This is the overview of the system which is implemented using eclipse, weka tool and snort tool. Thus it is intrusion detection system using the FGA and MLP algorithm on the datasets.

III. Methodology

In this section, we explain our FGA and MLP algorithms on KDD99 dataset and our datasets and performance evaluation criteria.

A. Algorithms

1) Fuzzy Genetic Algorithm

We consider using a FGA which is similar to the work proposed by T.P. Fries [9]. The FGA can classify two classes which are normal class and attack class. The steps to use FGA are described as follows.

The Fuzzy genetic algorithm starts from a population of individuals generated randomly. Each individual is an "if then" fuzzy rule. In order to optimize the set of fuzzy rules already generated in the first stage, a genetic algorithm process which consists of selection, crossover and mutation operators are applied on the individuals. The

FGA for intrusion detection are defined in follow:

1. Generate random population of n chromosomes.

2. Normalize each attribute of data to be a real number in the range 0.0 -7.0, where the maximum and minimum values among overall attributes from the training data are set to 7.0, and 0.0, respectively. The normalization is given in Eq. 1 and applied in order to set attribute numerical values in the range [0.0, 7.0].

$$x' = \frac{x - \min a}{maxa - mina}(n\_maxa - n\_mina) + n\_mina \qquad (1)$$

Where *x*: is the numerical attribute value, *mina* is the minimum value that the attribute *x* can get and *maxa* is the maximum one and *n_maxa* and *n_mina* are the new values of attribute x.

3. Encode the each attribute value into binary format and convert it into 0.0 – 7.0 octal number.

4. Find probability of each record for each detection rule and count for true negative and true positive as shown in the pseudo code below. A data record contains information of the attribute in the dataset.

```
for each record
  {  for each rule
       {  for each attribute
            { prob = fuzzy();
                    totalprob = totalprob + prob }
                 If (totalprob > threshold)
                 {class is attack;
                         true negative ++;}
               Else
               {class is normal;
                         true positive++;}
```

We used each rule to calculate a probability of being an attack of each data record. The system will read the record data one by one and evaluate each block of data to calculate a probability to be the attack using the trapezoidal fuzzy rule shape. Then, we gather a probability of each attribute and find the average probability. If the average probability is greater than a predefined threshold, the system would classify this record as an attack. We repeat this process for every record.

We used a trapezoidal shape to measure a probability of being an attack identified by each attribute. The fuzzy logic is encoded into four parameters which are a, b, c and d. The probability is calculated as shown in Fig. 2 where its meaning described below.

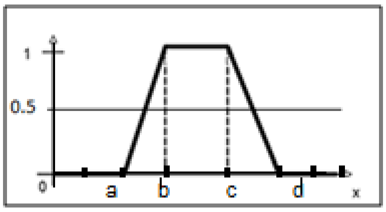


Fig. 2 A trapezoidal fuzzy rules with 4 parameters

From Fig .2, we calculate probability of being an attack from condition below

```
$if (attribute\_value\ is\ between\ "b"\ to\ "c")$
        $then\ prob = 0.0$

$else\ if\ (attribute\_value\ between\ "a"\ to\ "b"\ )$

        $then\ prob = \frac{attribute\_value\ - a}{b - a}$
$else\ if\ (attribute\_value\ between\ "c"\ to\ "d"\ )$

        $then\ prob = \frac{d - attribute\_value}{d - c}$
$else\ then\ prob = 1.0$
```

4. Fitness

Evaluate the fitness f(x) of each chromosomes x in the population. After counting all records from all rules, calculate fitness value of each detection rule using fitness Eq. 2

$$fitness = \frac{\alpha}{A} - \frac{\beta}{B} \qquad (2)$$

Where

1. A is the total number of attack records.
2. B is the total number of normal records.
3. α is the total number of attack records correctly identified as attack.
4. β is the total number of normal records incorrectly classified as attack.

5. Preserve rule that has the highest fitness value which is the best detection rule.

6. Use evolutionary GA method to find the next rules.

Thus using the "Fuzzy Genetic Algorithm" [14] we can detect the attack and classify the attack. But the algorithm is in two parts 1.Genetic algorithm 2.Rule creation. It is time consuming and open source and the dataset has detected types of DoS and Probe attack only. For the solution to this problem, we use the "Multilayer Perceptron Algorithm".

The "Fuzzy genetic algorithm" is in two parts 1.Genetic algorithm 2.Rule creation. The FGA is easier for the attacks to understand the flow of algorithm hence he can easily implement different technique to hack the system so it's not so secure and rule creation process is time consuming so it is time consuming and open source. The dataset had detected only the three types of attacks. That is DoS, Probe, and Normal attack only. For the solution to this problem, we use the **"**Multilayer Perceptron Algorithm".

2) Multilayer Perceptron Algorithm

The Multilayer Perceptron algorithm is supervised learning algorithm [15]. The Multilayer Perceptron is an example of an artificial neural network that is used extensively for the solution of a number of different problems, including pattern recognition and interpolation. It is a development of the Perceptron neural network model that was originally developed with layers that are input layer, output layer, and hidden layer.

MLP for intrusion detection are defined in follow:

1. Initialize the network, with all weights set to random numbers between -1 and +1.
2. Present the first training pattern, and obtain the output.
3. Compare the network output with the target output.
4. Propagate the error backwards.
    a. Correct the output layer of weights.
    b. Correct the input weights.
5. Calculate the error, by taking the average difference between the target and the output vector.
6. Repeat from 2 for each pattern in the training set to complete one epoch.
7. Repeat from step 2 for a set number of epochs, or until the error ceases to change.

But in our project we implement the "Multilayer Perceptron Algorithm" classification using *"Weka tool"* Software.

IV. PRELIMINARY RESULTS

Building any classifier involves two phases, i.e., training and testing phases. Training phase in our approach involves learning the parameters of the model from a KDD99 dataset, and apply the algorithm on that datasets profiles this data and uses this information to test datasets. To build a classifier we need to have labeled data for training and testing. Data sets released by DARPA [13] were used to train and test our datasets.

A. Experiments

The experiments that were conducted are described as follow.

1. We are going to create IDS System for detection of attacks in that we will use KDD99 dataset which gives an input to our application.

2. After that we pass this dataset file to our Fuzzy Genetics Algorithm in that we will calculate FP, FN and DR.

3. Next applications we will develop for attack detect in the online dataset using storing the TCP, ICMP packet in the network, and create the online dataset.

4. . We extends our application for prevent the attacks coming from those system for that we kill the process of that system.

5. Additionally we will implement MLP Algorithm for detect and classify the attacks from KDD99 and online dataset.

B. Network Dataset

We consider two different sources of datasets, which are KDD99 dataset and online dataset.

1) KDD99 Dataset

The KDD99 dataset is a benchmark dataset which was simulated in military network environment in 1998 then derived to KDD99 dataset in 1999. The dataset package was gathered and preprocessed into 41 attributes. In this work, we consider using only eight important attributes from the dataset. The selected attributes are attributes "duration, src_bytes, num_failed_logins, root_shells, num_acess_files, srv_count, serror_rate, and same_srv_rate" from 41 attribute and apply the algorithms on that dataset and classify the attack and normal data.

2) *Online Data Set*

We considered online network dataset by capturing online network data to filter only TCP, ICMP protocol packets. After that, the packet information is preprocessed into attributes by counting connections between any two IP addresses (source IP and destination IP) and collected to form a record in every 2 seconds which are further divided into attributes and saved as txt file that is our online dataset file.

The attributes of the online dataset are including
# of packets
# of source ports
# of destination ports
# of ICMP packets
Using that attribute we apply the algorithms and find the attack.

3) *Performance Evaluation Criteria*

With the help of the algorithms, we detect the attack and classify them and further provide the security against the attack and calculate,

1. Detection rate (DR) is the percentage of normal and attack correctly classified from total number of data records.
2. False-negative rate (FN) is the percentage that attack is misclassified from total number of attack records.
3. False-positive (FP) is the percentage that normal data is classified as attack from total number of normal data records.
4. The speed of detection is the time that our system uses, measured right after the records arrive until the system classifies the data and gives output classes in to normal or attack.

V. Conclusion

We have proposed a Fuzzy Genetic Algorithm for intrusion detection. Fuzzy genetic algorithm for network intrusion detection using both online network dataset and well-known KDD99 dataset. Our Fuzzy genetic algorithm and online dataset has more recent attack types of DoS and Probe attack. Moreover, it has more current behaviors of network activities than those in the KDD99 dataset.

Some limitation in the Fuzzy genetic algorithm we proposed the Multilayer Perceptron Algorithm for intrusion detection system on the datasets. The Multilayer Perceptron algorithm is supervised learning algorithm. The Multilayer Perceptron is an example of an artificial neural network that is used into number of layer and using weka tool, so it is more secure and easily implemented than other algorithms.

Our future work will focus on the following aspects.

1. To create our own dataset means captured data and then perform the experiments, so we have to perform without storing of data.
2. . The system classifies network data into only 3 classes including DoS, Probe and Normal. So we have to detect the all types of attack using FGA and MLP with online.

Acknowledgment

Our heartfelt thanks go to Siddhant College of Engineering for providing a strong platform to develop our skills and capabilities. We would like to thank to our guide & respected teachers for their constant support and motivation for us. Last but not least, we would like to thanks all those who directly or indirectly help us in presenting the paper.

References

[1] S. Chebrolu, A. Abraham, and J. P. Thomas, "Feature deduction and ensemble design of intrusion detection systems," *Comput. Secur.*, vol. 24, no. 4, pp. 295–307, Jun. 2005.

[2] W. Lee and S. J. Stolfo, "A framework for constructing features and models for intrusion detection systems," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 4, pp. 227–261, Nov. 2000.

[3] Y.-H. Liu, D.-X. Tian, and A.-M. Wang, "Annids: Intrusion detection system based on artificial neural network," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Nov. 2003, vol. 3, pp. 1337–1342.

[4] C. Zhang, J. Jiang, and M. Kamel, "Intrusion detection using hierarchical neural networks," *Pattern Recognit. Lett.*, vol. 26, no. 6, pp. 779–791,May 2005.

[5] Z. Zhang and H. Shen, "Online training of SVMs for real-time intrusion detection," in *Proc. Int. Conf. Adv. Inf. Netw. Appl.*, 2004, vol. 1, pp. 568–573.

[6] M. O. Depren, M. Topallar, E. Anarim, and K. Ciliz, "Network-based anomaly intrusion detection system uses SOMs," in *Proc. IEEE 12th Signal Process. Commun. Appl. Conf.*, Apr. 2004, pp. 76–79.

[7] H. G. Kayacik, A. N. Zincir-Heywood, and M. I. Heywood, "On the capability of an SOM based intrusion detection system," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2003, vol. 3, pp. 1808–1813.

[8] J. Gómez and E. León, "A fuzzy set/rule distance for evolving fuzzy anomaly detectors," IEEE International Conference on Fuzzy Systems, ART. No. 1682017, pp. 2286-2292.

[9] T.P. Fries, "A fuzzy-Genetic approach to network intrusion detection," GECCO'08: The 10th Annual Conference on Genetic and Evolutionary Computation, 2008, pp. 2141-2146.

[10] T. Komviriyavut, et al., "Network intrusion detection and classification with decision tree and rule-based approaches," 9th International Symposium on Communications and Information Technology, Art.No. 5341005, pp. 1046-1050.

[11] M-Y. Su, et al., "A real-time network intrusion detection system for Large-scale attacks based on an incremental mining approach," Computers and Security 28 (5), pp. 301-309.

[12] P. Kachurka, V. Golovko., "Neural network approach to real-time Network intrusion detection and recognition," The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS'2011, Art.No. 6072781 , pp. 393-397.

[13] KDD99 dataset, a network dataset [online], http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[14] M. Saniee Abadeha, , J. Habibia, C. Lucasb, "Intrusion detection using a fuzzy genetics-based learning algorithm", Journal of Network and Computer Applications 30 (2007) 414–428.

[15] Norouzian M.R., Merati. S., "Classifying Attacks in a Network Intrusion Detection System Based on Artificial Neural Networks" Proceedings of the Advanced Communication Technology (ICACT), 2011 13th International Conference on Publication Year: 2011, Page(s): 868 - 873