OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Intrusion Detection and Forensics based on decision tree and Association rule mining for Probe attack detection

Harishchandra Maurya[1], Swati Sharma[2]
[1]*M .Tech (C.S) (Assistant Professor, Bhagwant University, Ajmer, India)*
[2]*M .Tech (C.S) (Scholar, Bhagwant University, Ajmer, India)*
Department of Computer Sciences and Engineering Bhagwant University, Ajmer, India

ABSTRACT—*This paper present an approach based on the combination of, two techniques using decision tree and Association rule mining for Probe attack detection. This approach proves to be better than the traditional approach of generating rules for fuzzy expert system by clustering methods. Association rule mining for selecting the best attributes together and decision tree for identifying the best parameters together to create the rules for fuzzy expert system. After that rules for fuzzy expert system are generated using association rule mining and decision trees. Decision trees is generated for dataset and to find the basic parameters for creating the membership functions of fuzzy inference system. Membership functions are generated for the probe attack. Based on these rules we have created the fuzzy inference system that is used as an input to neuro-fuzzy system. Fuzzy inference system is loaded to neuro-fuzzy toolbox as an input and the final ANFIS structure is generated for outcome of neuro-fuzzy approach. The experiments and evaluations of the proposed method were done with NSL-KDD intrusion detection dataset. As the experimental results, the proposed approach based on the combination of, two techniques using decision tree and Association rule mining efficiently detected probe attacks. Experimental results shows better results for detecting intrusions as compared to others existing methods.*

Keywords—*Intrusion detection, Forensics, decision tree, Association rule mining, Probe attack.*

I. Introduction

Intrusion detection [1] is becoming an increasingly important technology that monitors network traffic and identifies network intrusions such as anomalous network behaviours, unauthorized network access, and malicious attacks to computer systems. There are two general categories of intrusion detection systems (IDSs): misuse detection and anomaly detection. Misuse detection systems detect intruders with known patterns, and anomaly detection systems identify deviations from normal network behaviours and alert for potential unknown attacks. In past years, there were only few intruders and so the user could manage them easily from the known or unknown attacks, but in recent years the security is the most serious problem. Because the intruders introduce a new variety of intrusions in the market, so that the user can't manage the computer systems and networks properly. Different approaches and algorithms are used to detect the attack. In this paper we propose the techniques which can detect network based attacks using Association rule mining [2] and decision tree [3].

Computer forensics [4]. has significant ability to make network infrastructures more integrated and capable of surviving attack. Computer Forensics is the use of computer technology, in accordance with the lawful procedures and criterion. From the evidence points to technical analysis, computer forensics technology mainly divided into static and dynamic forensics evidence. Dynamic forensics technology will be integrated into the firewall, intrusion detection. Dynamic computer forensics combined with the intrusion Detection, which can collect reliable evidence in real time when the system is invaded, complete the invasion of Testing and the evidence of the dynamics of computer forensics, becomes research focus of computer forensics.

- ***Decision tree***

Decision trees are well known machine learning techniques. A decision tree is composed of three basic Elements.

- A decision node specifying a test attributes.
- An edge or a branch corresponding to the one of the possible attribute values this means one of the test attribute outcomes.
- A leaf which is also named an answer node, contains the class to which the object belongs.

In decision trees, two major phases should be ensured:

Building the tree: Based on a given training set, is to build a decision tree. It consists of each decision node, Select the "appropriate" test properties, and define the class label of each leaf.

Classification: Order to classify a new instance; we began to determine the root of the tree, then we test the Node specified property. The test results, allowing moving down the tree relative to a given instance of the attribute value. This process is repeated until it encounters a leaf. The instance is then classified in the same class Characteristics to leaves.

Decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. More descriptive names for such tree models are classification trees or regression trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. Data comes in records of the form:

$$(x,Y) = (x_1, x_2, x_3, x_4,\ldots, x_k, Y) \quad (1)$$

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector x is composed of the input variables, x_1, x_2, x_3 etc., that are used for that task.

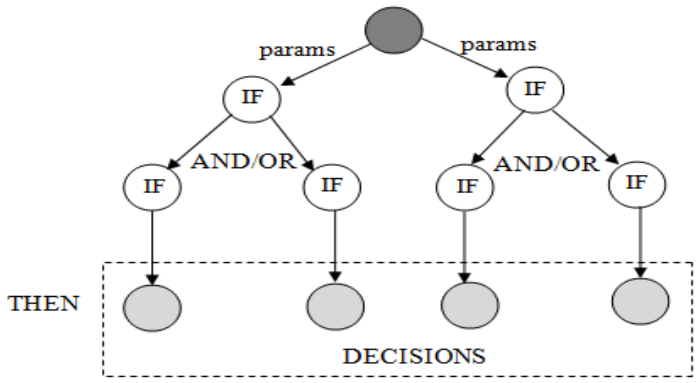

Figure 1: General structure of decision tree

- ***Association rule mining***

Association rule mining is one of the hottest research areas that investigate the automatic extraction of previously unknown patterns or rules from large amounts of data. The basic concept of association rule mining is given below:

Let $I = \{I_1, I_2, \ldots, I_m\}$ be a set of items. Let D be a database of transactions where each transaction T is a set of items such that $T \subseteq I$. Each transaction is associated to an identifier, call TID. A transaction T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$, and $A \cap B = \emptyset$. The rule $A \Rightarrow B$ holds in the transaction set D with support s, where s is the percentage of transactions in D that contain AB. The rule AB has confidence c in the transaction set D. That is,

(1)

(2)

Where A is named as the support count of the set of items A in the set of transactions D, as denoted by $sup\_count(A)$. A occurs in a transaction T, if and only if $A \subseteq T$. Rules that satisfy both a minimum support threshold (min_sup) and a minimum confidence threshold (min_conf) are called strong. A set of items referred to as an itemset. An itemset that contains k items is a k -itemset. Itemsets that satisfy min_sup is named as frequent itemsets. All strong association rules result from frequent item sets.

- ***Probing Attack and Its Type***

Probing Attack [5]. is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls. It involves discovering the algorithms and parameters of the recommender system itself. It may be necessary for an intruder to acquire this knowledge through interaction with the system itself.

- Ipsweep.
- Portsweep.
- Nmap
- Satan

The subsequent parts of this paper are organized as follows: At first, in section 2, presents the proposed approach. Next, at section 3 and 4, the proposed approach procedure of Probe attack detection is explained and experimental results are discussed respectively. Finally, section 5 makes some concluding remarks and proposes further areas for research.

II. Proposed Approach

This proposed approach uses combination of, two techniques using decision tree and Association rule mining for Probe attack detection. Decision tree based technique is applied to find out the best suitable parameters. Some of the classification algorithm that most commonly used to classify the dataset named J48. It is a collection of machine learning algorithms for data mining tasks. It contains tools for association rules. Data mining techniques have been commonly used to extract patterns from sets of data. Specifically, association rules is one of the data mining approach [6] and used for anomaly detection. Association rule find correlations between features or attributes used to describe a data set. Here, we are combining both the approaches of association rule mining for selecting the best attributes together and decision tree for identifying the best parameters together to create the rules for fuzzy expert system. The combination proves to be good technique. Figure below shows the proposed framework to generate rules for fuzzy expert system.

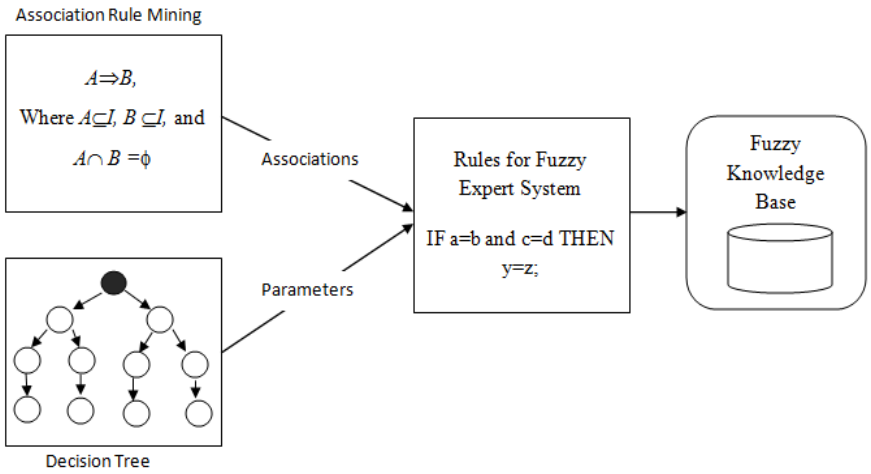

Figure 2: Generating rules for fuzzy expert system

III. Procedue Based On Decision Tree And Association Rule Mining For Probe Attack Detection

In this section, we explain the procedure of Probe attack detection. Procedure is as follows:

- In our approach we are proposing a combination of neural and fuzzy systems are two standalone techniques to efficiently detect the anomalous behaviour. In which fuzzy systems provide a powerful framework for expert knowledge representation, while neural networks provide learning capabilities and exceptional suitability for computationally efficient hardware implementations.
- Generates fuzzy rules from the uploaded dataset using subtractive clustering method. This approach of generating rules using clustering methods is limited to the problems of clustering techniques. The NSL-KDD dataset [7]. is not suitable for clustering [8]. So we are purposing the new method to generate rules for fuzzy expert system using association rule mining and decision trees.
- Our work is implemented using MATLAB 7.6.0 (R2008a) Neuro-Fuzzy Toolbox software. MATLAB provides several computing tools based GUIs to work with. In this dissertation work we are using Neuro-Fuzzy toolbox (i.e. anfisedit) and Fuzzy Logic toolbox[9] (i.e. fuzzy) to implement our work
- The NSL-KDD is used as database to train and test the system performance. Experiments are performed with NSL-KDD dataset. We have conducted experiments on Probe attack.
- Both training and testing dataset can loaded in ANFIS tool.
- Decision Trees in WEKA tool is used to analyze the NSL-KDD and identify the required parameters.
- After that Membership functions are generated using trapezoidal functions and one Gaussian function is used for Boolean type of attribute (i.e. logged_in). Nine inputs having three membership functions except loggen_in (i.e. having 2 membership functions) generated for the probe attack. Figure 3.1 shows the generated fuzzy membership functions and their attributes.

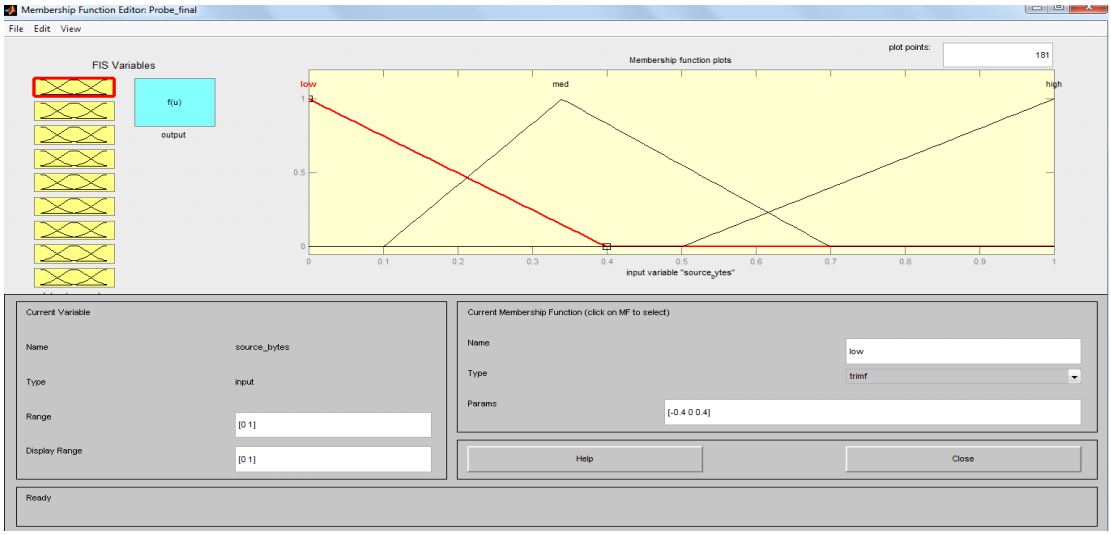

Figure 3.1: Membership functions for Probe Attack

- Based on these rules we have created the fuzzy inference system that is used as an input to neuro-fuzzy system. Figure 3.2 shows the generated FIS [10].

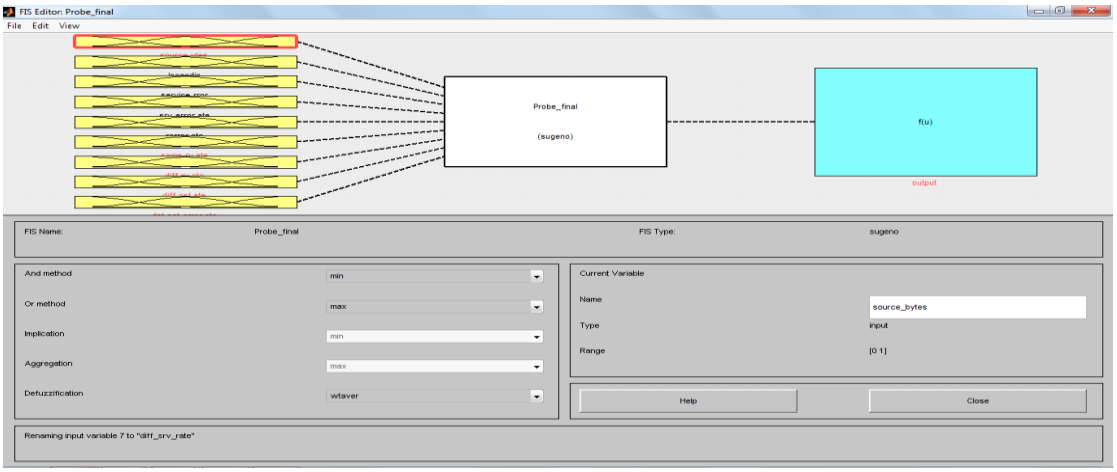

Figure 3.2: Fuzzy inference system

- Figure 3.3 shows the output of the fuzzy inference system. This system can be adjusted according to the values of the input attributes. Fuzzy Inference system shows an accuracy of 94.4 %.

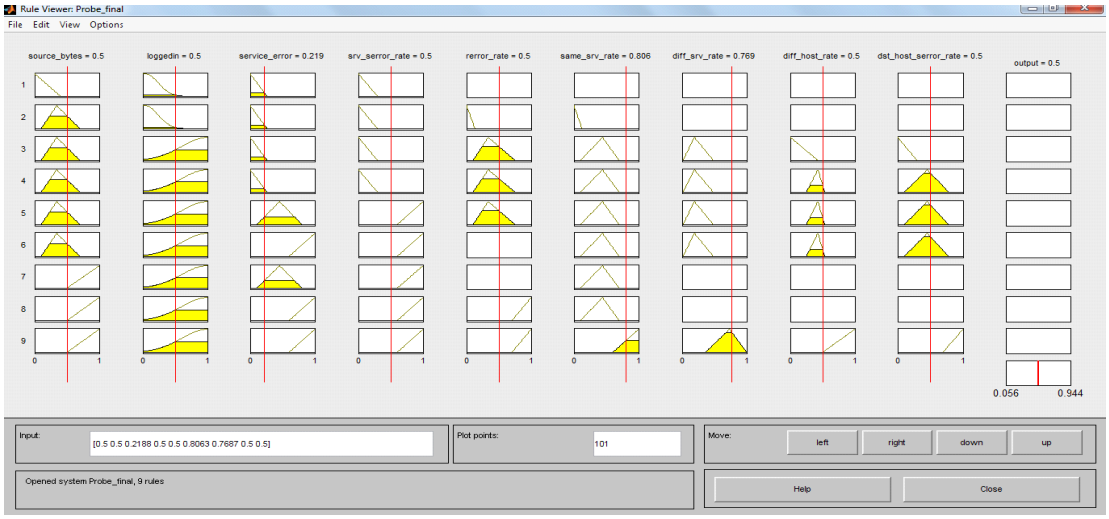

Figure 3.3: Fuzzy inference system output

- this fuzzy inference system is loaded to neuro-fuzzy toolbox as an input and the final ANFIS[11] structure is generated as shown in figure 3.4 below

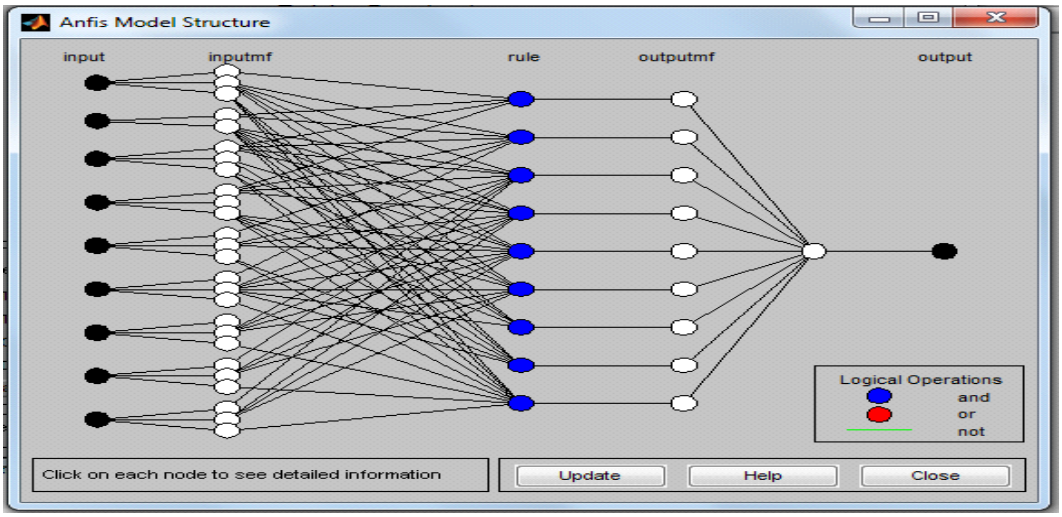

Figure 3.4: ANFIS structure

- ANFIS tool output we have implemented it for detecting probe attacks. We have implemented it for detecting probe attacks. Neuro-Fuzzy system results in 99.68 % accuracy in detecting the probe attack.

IV. Experiment And Results

In this section, we summarize our experimental results to detect Probe attack using decision tree and Association rule mining over NSL-KDD dataset. We have conducted experiments on Probe attack having nine contributing attributes having 1000 number of instances.NSL-KDD dataset consists of 1000 instances in training dataset and similar instances in testing dataset that can be loaded easily into the neuro-fuzzy toolbox. The attack types are grouped into four categories:

- ***Ipsweep*:** It probes the network to discover available services on the network. First intruder find out a machine on which he may be attacked.
- ***Portsweep*:** It probes a host to find available services on that host. If a service is known on the system so it may easily be attacked by the network intruder.
- ***Nmap*:** It is a complete and flexible tool for scanning a network either randomly or sequentially. Therefore, often intruders used this tool for scanning network parameters that may help them in attacking the system.
- ***Satan*:** It is an administration tool; it gathers information about the network. This information can be used by an attacker.

TABLE1: TRAINING DATASET REPRESENTING PROBE ATTACK ATTRIBUTES

1	2	3	4	5	6	7	8	9
src_bytes	**logged_in**	**serror_rate**	**srv_serror_rate**	**rerror_rate**	**same_srv_rate**	**diff_srv_rate**	**dst_host_srv_count**	**dst_host_serror_rate**
491	0	0	0	0	1	0	25	0
146	0	0	0	0	0.08	0.15	1	0
0	0	1	1	0	0.05	0.07	26	1
232	1	0.2	0.2	0	1	0	255	0.03
199	1	0	0	0	1	0	255	0
0	0	0	0	1	0.16	0.06	19	0
0	0	1	1	0	0.05	0.06	9	1
0	0	1	1	0	0.14	0.06	15	1
0	0	1	1	0	0.09	0.05	23	1

0	0	1	1	0	0.06	0.06	13	1
0	0	0	0	1	0.06	0.06	12	0
0	0	1	1	0	0.02	0.06	13	1
287	1	0	0	0	1	0	219	0
0	0	1	1	0	0.17	0.05	2	1

TABLE 2: TESTING DATASET REPRESENTING PROBE ATTACK ATTRIBUTES

1	2	3	4	5	6	7	8	9
src_bytes	**logged_in**	**serror_rate**	**srv_serror_rate**	**rerror_rate**	**same_srv_rate**	**diff_srv_rate**	**dst_host_srv_count**	**dst_host_serror_rate**
44	0	0	0	0	0.75	0.5	254	0
0	0	1	1	0	1	0	79	0.21
53	0	0	0	0	1	0	255	0
0	0	1	1	0	1	0	1	1
54540	1	0	0	0.5	1	0	229	0
0	0	0	0	1	0.04	0.06	9	0
0	0	0	0	0	1	0	165	0.49
56	0	0	0	0	1	0	255	0
192	0	0	0	0	1	0	51	0
0	0	0.14	0	0.86	0	1	1	0.13
0	0	0.07	0	0.93	0	1	1	0.07
21	0	0	0.21	0.75	0.5	0.75	40	0

This dataset (i.e. shown in table 1 and 2) is normalized to the range of [0-1]. Standard mean deviation is used to define the mid value for each rule. The testing dataset, against which the ANFIS system is validated. The dataset for the Probe attack having 1000 number of instances is given as an input to the ANFIS tool. Both training and testing dataset can be loaded at the same time while testing dataset is used later to verify the training dataset inputs. Before creating the fuzzy membership functions, the required parameters range of parameters needs to be identified. Decision Trees is used to identify the required parameters. The minimum and maximum is set to the range 0-1. Each attribute is distributed over a range of low, mid (medium) and high. Range of parameters is shown in table 3.

TABLE 3: MEMBERSHIP PARAMETERS FOR PROBE ATTACK

Attributes	**Min**	**Low**	**Mid**	**High**	**Max**
source_bytes	0	0.2	0.34	0.7	1
logged_in	0	-	-	-	1
serror_rate	0	0.295	0.453	0.89	1
srv_serror_rate	0	0.294	0.454	0.81	1
rerror_rate	0	0.131	0.335	0.682	1
same_srv_rate	0	0.126	0.44	0.643	1
diff_srv_rate	0	0.069	0.483	0.732	1
dst_host_srv_rate	0	0	0.426	0.436	1
dst_host_serror_rate	0	0.296	0.452	0.691	1

Based on these rules we have created the fuzzy inference system. These rules are used to makes decision for neuro-fuzzy system. The final outcome is the production of these rules. Fuzzy inference system can be adjusted according to the values of the input attributes. Fuzzy inference system is used as an input to neuro-fuzzy system and shows an accuracy of 94.4 %. The last attribute shows the defuzzified output of the fuzzy system. Fuzzy inference system is loaded to neuro-fuzzy toolbox as an input and the final ANFIS structure is generated.

ANFIS module is trained using neural network back-propagation algorithm. While training the dataset

with this ANFIS structure it is observed that training error reduces with number of epochs. For our experiment we have set number of epochs to 100. Neural network learn through this phase while reduces the network errors. ANFIS tool output is evaluated on the basis of accuracy in predicting the attacks clearly. We have implemented it for detecting probe attacks. Results below show the outcome of neuro-fuzzy approach. The output of neuro-fuzzy toolbox is the weighted neurons that are defuzzified using centroid defuzzification method. Euro-Fuzzy system results in 99.68 % accuracy in detecting the probe attack. A slight comparison to previous literature is given as follows:

TABLE 4: COMPARISON OF RESULTS

Approach in [paper]	***Probe Detection rate (%)***
[84]	73.20
[85]	100
[86]	72.8
[87]	99
ANFIS Approach	99.68

Figure 4.1: Comparison of results

Figure 4.2: Plotting error against number of epochs

Figure 4.2 shows that error decreases with the number of epochs. Error is reduced to 0.58729 to 0.078525 in 100 epochs (iterated seven times). To detect Probe attack using decision tree and Association rule mining is implemented using MATLAB [12] 7.6.0 (R2008a) Neuro-Fuzzy Toolbox software.

V. Conclusion

In this paper, an approach based on the combination of, two techniques using decision tree and Association rule mining for Probe attack detection for Intrusion Detection and Forensics was successfully demonstrated its usefulness on the training and testing subset of NSL-KDD dataset. Well known machine learning technique called decision tree technique is applied to the NSL-KDD dataset to find out the best suitable parameters. We used WEKA tool to generate decision trees for analysing NSL-KDD dataset. One of the data mining approach called association rules is used for selecting the best attributes. Probe attacks are the kind of attack which tries to collect the data and find the network's vulnerability. Probing attacks can steel important information from computer or network and later may be used by the intruder. These attacks may also result in significant loss of time and money for many organizations. There are many methods for their prevention but has some limitations like varying nature of attacks. At last, this paper proposed an approach based on a combination of, two techniques decision tree and Association rule mining. This approach is the combination of these two technique is use to detect Probe attack category. Experimentation results showed the proposed method is effective in Probe attack detection. In the future, we will study more on improving detection accuracy of probe attacks. In addition, we will only use the most relevant features on the given probe detection task to reduce the computation time and hopefully to enhance the accuracy of detection as well.

REFERENCES

[1] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, "Network intrusion detection," *Network, IEEE*, vol. 8, no. 3, pp. 26-41, 1994.

[2] J. Luo and S. Bridges, "Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection," in *International Journal of Intelligent Systems*, pp. 687–703, John Wiley & Sons, 2000.

[3] Z. Liu and D. Feng, "Incremental Fuzzy Decision Tree-Based Network Forensic System," in *Computational Intelligence and Security*. vol. 3802: Springer Berlin Heidelberg, 2005, pp. 995-1002.

[4] G. Palmer, "A road map for digital forensic research," *First digital forensic research workshop (DFRWS'01)*, pp. 27-30, 2001.

[5] J. McHugh, "Testing Intrusion detection system: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory," *ACM Transaction on Information and system security*, vol. 3, no. 4, pp. 262-294, 2000.

[6] "WEKA – Data Mining Machine Learning Software," *[Available Online]*. *http://www.cs.waikato.ac.nz/ml/weka/*.

[7] M. Sabhnani and G. Serpen, "Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context. [Available At] http://www.eecs.utoledo.edu/~serpen/.

[8] M. B. Al-Zoubi, A. Hudaib, and B. Al-Shboul, "A fast fuzzy clustering algorithm," in *in Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases*, Corfu Island, Greece, 2007.

[9] Fuzzy Logic Toolbox, [Available At] http://www.mathworks.in/help/fuzzy/index.html

[10] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *Systems, Man and Cybernetics, IEEE Transactions*, vol. 23, no. 3, pp. 665-685, 1993.

[11] A. Al-Hmouz, S. Jun, R. Al-Hmouz, and Y. Jun, "Modeling and Simulation of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for Mobile Learning," *IEEE Transactions on Learning Technologies*, vol. 5, no. 3, pp. 226-237, 2012.

[12] C. Moler, "The Origins of MATLAB,", *Retrieved April 15, 2007, [Available Online] http://www.mathworks.in/products/matlab/*, 2004.