An Intelligent Intrusion Detection System Using Genetic Based Feature Selection and Modified J48 Decision Tree Classifier

[A]B.Senthilnayaki, Department of IT, University College of Engineering Villupuram, India, nayakiphd@gmail.com

[B] K.Venkatalakshmi, Department of ECE, University College of Engineering Tindivanam, India, venkata_krish@gmail.com

[C]A.Kannan, Department of IST, College of Engineering, Anna University,India, kannan@annauniv.edu

1)

Abstract— **In this paper, intelligent algorithms for intrusion detection are proposed which detect the network attacks as normal or anomaly based attacks by performing effective pre-processing and classification. This system uses a new genetic algorithm approach for pre-processing and Modified J48 classification algorithm to identify the intended activities. The new genetic based feature selection algorithm proposed in this paper is helpful to identify the important features needed to classify the normal and anomaly records. The proposed intelligent IDS has been empirically tested in a simulated environment and the experimental results show that the proposed method provides higher detection accuracy than the existing methods in terms of detection rate with reduced false rate. The salient contributions of this paper are, the proposed a new processing technique for removing noisy data in the KDD cup'99 dataset, identification of the optimal features selection by applying modified genetic algorithm and finally the proposed of modified J48 decision tree algorithm for efficient classification for providing intelligent network intrusion detection.**

***Index Terms*— Genetic Based Feature Selection, Intelligent Agent, Intrusion Detection System, Modified J48.**

II. INTRODUCTION

Network security has now become a critical research issue due to the rapid development of internet applications. In the past decade, significant progress has been made the enhancement of computer system security. This is due to the fact that computers and networks are vulnerable to attacks from both unauthorized users as well as the attacks from authorized users who misuse the privileges are called as inside attackers. Intrusion Detection System (IDS) is the best technique applied to solve these problems. An IDS is a system that attempts to identify intrusions which are misuses or abuses of computer systems or networks by malicious users. Some IDSs monitor a single computer while other monitor several computers connected by a network.

Moreover, IDSs are software or hardware systems that automate the process of monitoring the events occurring in a computer system or network. Hence, IDS has become an important and widely used additional mechanism in the network security infrastructure of many organisations. An IDS enables a network administrator to deal with intrusions more efficiently but it must be noted that they cannot prevent intrusions. The usefulness of IDS comes after an intrusion has taken place, allowing an administrator to retrace the steps of the intruder and find what security measures were bypassed.

In computer networks, more time is consumed on the processing of database which grows for every packet of data. This affects the purpose of IDS, since it is necessary to provide fast response. Besides this serious hazard, there are other challenges like huge network traffic volumes, the difficulty to realize decision boundaries between normal and abnormal behaviors and requirement for continuous adaptation to a constantly changing environment.

IDS are traditionally classified as either Misuse based or Anomaly-based. Misuse based detection uses the known data to compare the unknown data with it and an analysis the attacks whereas anomaly based detection classifies the based on attack and normal patterns. Misuse based detection is considered for detection in this paper.

In the past, anomaly based detection models of normal network behaviour (profiles) were developed for automatic detection of new patterns that deviate from the profile. Deviations represent actual intrusions or simply the new behaviour that needs to be added to the profiles. However, it is difficult to precisely model all behaviours since anomaly based detection can detect only known attacks.

978-1-4799-3448-5/13/$31.00 ©2013 IEEE

Behaviour based detection in network systems are monitored and compared with pre-configured and pre-determined attack patterns known as signatures. These hand coded signatures are provided by human expects based on their extensive knowledge of IDS. If the pattern is matched, an event is signalled for which an alarm is raised to alert the admin or user. IDS are not static, therefore signature needs to be updated whenever new software versions arrive or changes in network configuration new patterns of intrusions added to the set of signature for misuse-based detection to provide high detection rate for novel intrusions. Moreover, IDS monitors target system activity as it is recorded in audit records generation by target system. The audit database is augmented with real time traffic data upon which the IDS were trained. The important deficiency in KDD dataset is the huge number of redundant records which causes the learning algorithm to be unbiased. So, the redundant and noisy records are to be removed.

AI techniques have been applied both to misuse detection and also for anomaly detection. Standford Research Institute's (SRI's) Intrusion Detection Expert System (IDES) [1] encodes an expert's knowledge of known patterns of attack and system vulnerabilities as IF-THEN rules. Time-based Inductive Machine (TIM) [2] for intrusion detection learns sequential patterns.

The fuzzy logic provides some flexibility to the uncertain problem of intrusion detection and allows much greater complexity for IDS. Most of the fuzzy IDS require human experts to determine the fuzzy sets and set of fuzzy rules. These tasks are time consuming. However, if the fuzzy rules are automatically generated, less time would be consumed for building a good intrusion classifier and shortens the development time of building or updating an intrusion classifier. The model suggested by Dokas et al [3] provides rare class prediction models for identifying known intrusions and their variations and anomaly or outlier detection schemes for detecting novel attacks whose nature is unknown. One of the approaches in fuzzy reference is to use a new Markov model. In [4] Gomez et al, Using genetic algorithm to generate fuzzy classifiers that can detect anomalies and some specific intrusions. Their main idea is to evolve two rules, one for the normal class and the other one for the abnormal class using a profile data set with information related to the computer network during the normal behavior and during intrusive (abnormal) behavior. Fuzzy preference relation is another method applied to intrusion detection based on fuzzy satisfaction function which is applied for comparison of attack signatures, where fuzzy signatures are combined by fuzzy operators [5].

Classification is a data mining technique used to predict group membership for data instances. Popular classification techniques include decision trees and neural networks. Classification tree methods were a good choice when the data mining task is classification or prediction of outcomes. Using classification tree labels records and assigns them to discrete classes and provide the measure of confidence that the classification is correct. It is built through a process known as binary recursive partitioning.

Soft computing techniques provided in artificial intelligence can be applied for developing IIDSs. Since these techniques utilize tolerance for tackling ambiguity, uncertainty, partial truth and approximation to achieve robustness at low solution cost. Many intelligent IDS which they used Support Vector Machine (SVM) for classification to accurate detection rate.

Moreover in IDS, the patterns must be stored for each and every packet that passes the IDS so the audit database size will increase. This will result in more time consumption and will generate more false positives. In addition, this database contains redundant and irrelevant features. Therefore, it is necessary to perform both feature reduction and effective classification to develop an efficient intrusion detection system.

In this paper, we propose a new intrusion detection system using soft computing techniques to offer effective security through the provision of detection accuracy, fast processing time, ability to adapt and exhibit fault tolerance. For this IDS, we propose a new genetic based feature selection algorithm which reduces the 41 features of the KDD Cup data set into 9 important features by applying the fitness value as a threshold. Moreover, we perform classification using a modified decision tree algorithm which has been developed by enhancing the existing J48 decision tree algorithm. The major advantages of the proposed IDS are reduction in false positive and fast classification. The remainder of this paper is organized as follows: Section 2 provides a survey of related works. Section 3 depicts the architecture of the system proposed in this paper. Section 4 explains the genetic based feature selection algorithm proposed in this work. Section 5 details the proposed classification algorithm and the results obtained in this work. Section 6 provides conclusions on this work and suggests some possible future works.

III. THE RELATED WORK

There are many works in the literature that discuss about the design and implementation of IDSs. Zhi-Song-Pan [6] proposed a neural network based IDS, since neural network have high performance to detect DOS and Probing attacks. They also used C4.5 algorithm since it can detect the R2L and U2R more accurately than neural network. Prema Rajeswari and Kannan [7] proposed an Active Rule Based Approach for Network Intrusion Detection with Enhanced C4.5 Algorithm. It detects misuse behaviors using classification and it is used to accurately predict intrusions. This Enhanced C4.5 decision trees provide better detection than C4.5 Generated rules to detect intrusions. Christopher Krueger et al [8] used the Bayesian Event Classification method for Intrusion Detection

that improves the aggregation of different models and it gives accurate outputs.

Alexander Hofmann et al [9] proposed an evolutionary wrapper approach for IDS which gives better classification results. This wrapper approach is addressed by means of an evolutionary algorithm for feature selection. Mukhopadhyay et al [10] proposed new IDS that use Back Propagation Neural Network (NN) approach for classification. The NN can detect abnormal packets faster. It undergoes rigorous training, validation and testing phases. Network traffic has been efficiently modelled by this using Back propagation network since the processing elements are highly interconnected. Jeich Mar et al [11] proposed an IDS based on Adaptive Neuro-Fuzzy Inference System (ANFIS) rules to minimize the detection delay on the deauthentication process while finding the Denial-of-Service (DoS) attacks on the Medium Access Control (MAC) layer of a Wireless Local Area Network (WLAN).

Rupali Datti and Bhupendra verma [12] proposed a Linear Discriminate Analysis based algorithm to extract features for detecting intrusions and Back Propagation Algorithm is used for classification of attacks.

Gunes Kayacik et al [13] proposed an Information Gain based model that provides an effective facility for selecting features. Mahbod Tavallaee et al [14] proposed a detailed analysis of the KDD Cup 99 dataset using classification since KDD is mostly widely used data set for the evaluation of systems. The important deficiency in the KDD data set is the huge number of redundant records. It provides good classification rate and investigates the relevance of the 41 features with respect to dataset labels. Huan Liu et al [15] proposed new concepts and algorithms for feature selection. They used different search methods for optimal feature selection.

The work by a dynamic approach that tries to discover known or unknown intrusion patterns which uses Support Vector Machine [16]. Du Hongle et al [17] proposed an improved v-FSVM through introduction membership to each data point. They reformulated the improved v-FSVM so that different input points can make different contributions to decision hyperplane. In order to verify the performance of the improved v-FSVM, they applied it to intrusion detection. Yu-Ping Zhou et al [18] presented a hierarchical neuro-fuzzy inference intrusion detection system. In the proposed system, principal component analysis neural network has been used to reduce the input data space. An enhanced fuzzy C-Means clustering algorithm has been applied to create and extract fuzzy rules. The adaptive neural fuzzy inference system has been utilized repeatedly in their model. At last, the system was optimized by genetic algorithm. The main advantages of the HFIS model are the capability to not only misuse detection but also anomaly detection. Moreover, the proposed method has higher speed and better performance.

A hybrid intrusion detection method based on HMM and Fuzzy logic has been proposed [19]. The experiment results showed that their method is efficient to classify the anomaly profile from the normal profile.

ZihuiCheXueyunJi [20] presented a new anomaly detection model based on rough set reduction and Hidden Markov model (HMM) on the basis of the analysis of shortcomings of other detection methods these days. A new feature selection algorithm combining rough sets and genetic algorithm on the basis of clustering was proposed by YutengGuo et al [21]. Their algorithm uses the rough set theory to process selection, then uses a genetic algorithm based approach for clustering that finds the optimal subsets.

Sanjay et al [22] proposed an improved network Intrusion Detection technique based on K-Means Clustering and Naive Bayes Classification algorithms. The K-means clustering algorithm makes clusters for normal and anomalous traffic in the training dataset. Varun Chandola and Arindam Banerjee [23] proposed an Anomaly Detection model which is used for classifying sequence data because the nodes of a network have critical information such as intrusions, faults, and system failures. Adetunmbi A.Olusola et al [24] proposed a technique for the analysis of KDD'99 Intrusion Detection Dataset using feature selection and classification to detect and classify the intrudes.

Comparing with all the works present in the literature, the work proposed in this paper is more effective in many ways. First, it uses the standard bench mark dataset namely, the KDD cup dataset for carrying out the experiments. Second, it uses a new fitness function for taking decisions on feature selection using genetic algorithms. Third, it uses intelligent agents for effective decision making in classification using decision trees. Finally, it reduces false positive and error rate.

IV. THE PROPOSED SYSTEM ARCHITECTURE

The architecture of the system proposed is shown in fig.1. In this paper consists of eight major modules namely dataset, pre-processing module, optimal feature selection module, classification sub system, decision manager, rule manager, administrator and knowledge base.

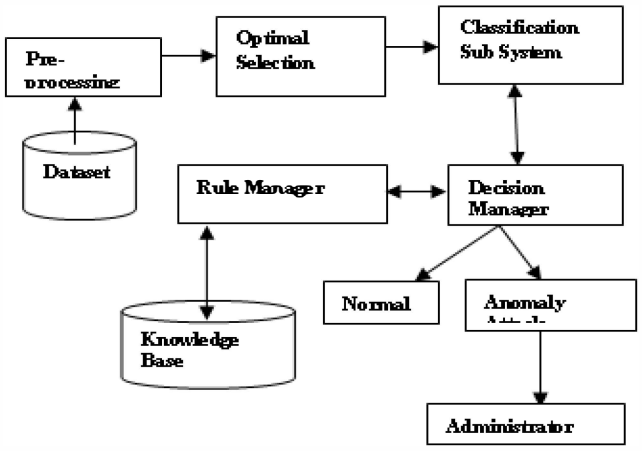

Fig.1 System Architecture

The dataset shown in the architecture is used to store all the records of the KDD cup dataset. The pre-processing module is used to select the necessary records from the KDD cup dataset. In this work, only 10% of the dataset are selected by the pre-processing module for carrying out the further experiments. The optimal feature selection algorithm selects 9 important features from the 41 features present in the dataset. The classification sub system is responsible for both training and testing in which the final decisions are taken by the decision manager. The overall control of the system is with the decision manager. It is responsible for controlling the classification sub system and the rule manager. The rule base is used to store the rules generated by the classifier during training and the rule manager is responsible for effective storage and retrieval of rules. In addition, the rule manager is responsible for rule selection, rule firing and inference in coordination with the decision manager.

V. PREPROCESSING

KDD'99 dataset suffers from major weakness due to the presence of redundant records. These to redundant records reduce the detection rate and accuracy. KDD'99 dataset has 41 features with classes labeled as either normal or anomaly with specific attack type. In Pre-processing, the redundant records are removed from the dataset. For this, the standard deviation value, maximum and minimum values for each feature are calculated. The redundant records are removed to improve detection accuracy.

A. Feature Selection Algorithm

In this work, Genetic algorithm based approach is proposed to select the optimal features from the overall 41 features. The selected features discriminate in predicting class during classification for anomaly and misuse.

The steps of the algorithm are as follows:

1. Generate random population of n chromosomes (dataset suitable solutions for the problem)
2. Evaluate the fitness f(x) = k (x) / sqrt(k+k(k-1) x) where k is a random number and x represents the chromosome in the population
3. Create a new population by repeating following steps until the new population is complete,
   a) Select two parent chromosomes from a population according to their Fitness (the better fitness, the bigger chance to be selected).
   b) With a crossover probability the parents form a new offspring (children).
      If no crossover was performed, offspring is an exact copy of parents.
   c) With a mutation probability mutate new offspring at each locus (position in chromosome).
   d) Place new offspring in a new population.
4. Use new generated population for a further run of algorithm
5. If the end condition is satisfied, stop and return the best solution in current population
6. Go to step 2.

B. Selected Features

The main reason for selecting KDD Cup 99 dataset is that currently, it is the mostly used comprehensive data set that is shared by many researchers. In this dataset, 41 attributes (Table 1) are used in each record to characterize network traffic behavior. Among this 41 attributes, 38 are numeric and 3 are symbolic. Features present in KDD data set are grouped into three categories and are discussed below.

A. Basic Features: Basic features comprises of all the attributes that are extracted from a TCP/IP connection. These features are extracted from the packet header and includes src_bytes, dst_bytes, protocol etc

B. Content Features: These features are used to evaluate the payload of the original TCP packet and looks for suspicious behavior in the payload portion. This includes features such as the number of failed login attempts, number of file creation operations etc. Moreover, most of the R2L and U2R attacks don't have any frequent sequential patterns. This is due to the fact that DoS and Probing attacks involve many connections to some host(s) in a very short duration of time but the R2L and U2R attacks are embedded in the data portions of the packets, and generally involves only a single connection. So to detect these kinds of attacks, content based features are used.

c. Traffic Features: These include features that are computed with respect to a window interval and are divided into two categories

i) "Same host" features: These features are derived only by examining the connections in the past 2 seconds that have the same destination host as the current

connection, and compute statistics related to protocol behavior, service etc.

ii) "Same service" features: These features examine only the connections in the past 2 seconds that have the same service as the current connection. The above two types are called "time based traffic features".

TABLE I
LIST OF FEATURES AVAILABLE IN KDD '99 CUP DATA SET

S.NO	Feature Name	S.No	Feature Name
1	Duration	22	Is_guest_login
2	Protocol_type	23	Count
3	Service	24	Serror_rate
4	Src_bytes	25	Rerror_rate
5	Dst_bytes	26	Same_srv_rate
6	Flag	27	Diff_srv_rate
7	Land	28	Srv_count
8	Wrong_fragment	29	Srv_serror_rate
9	Urgent	30	Srv_rerror_rate
10	Hot	31	Srv_diff_host_rate
11	Num_failed_logins	32	Dst_host_count
12	Logged_in	33	Dst_host_srv_count
13	Num_compromised	34	Dst_host_same_srv_rate
14	Root_shell	35	Dst_host_diff_srv_rate
15	Su_attempted	36	Dst_host_same_src_port_rate
16	Num_root	37	Feature Name
17	Num_file_creations	38	Dst_host_srv_diff_host_rate
18	Num_shells	39	Dst_host_serror_rate
19	Num_access_files	40	Dst_host_srv_serror_rate
20	Num_outbound_cmds	41	Dst_host_rerror_rate
21	Is_host_login		

Using the genetic algorithm, the following 9 features shown in Table 2 have been selected. From table 2, it is observed that the feature selection algorithm proposed in this paper has selected only the most contributing attributes from the 41 features. These 9 features are used by the classification algorithm for effective classification of the dataset.

TABLE II
LIST OF SELECTED FEATURES IN KDD '99 CUP DATA SET

S.No	S.No from KDD Table	Selected Feature
1	2	protocol_type
2	3	Service
3	4	src_byte
4	5	dst_bytes
5	6	Flag
6	27	diff_srv_rate,
7	33	dst_host_srv_count,
8	40	Dst_host_rerror_rate
9	41	dst_host_srv_rerror_rate

VI. CLASSIFICATION

The classification is based on the modified J48 Decision Tree algorithm which has been proposed in this paper to classify the anomaly and misuse classes from the selected features which were selected using genetic algorithm. It is used to classify normal and anomaly classes accurately. The modified Decision Tree is constructed using selected optimal features and intelligent agents to ensure maximum accuracy of classification from the dataset. If the classified data results as normal then it is consider as normal data and if the classification result as anomaly then it is further categorized into one of the four attack classes (DOS, Probe, U2R and R2L). The steps of the proposed classification algorithm are as follows:

Modified J48 Decision Tree Algorithm steps:

Step 1: Create the root node of the tree using the best data value.

Step 2: For selected attributes and using records in the dataset perform training as

2a) Call intelligent agent and make a binary classification based on a threshold value.

2b) For each attribute call intelligent agent while constructing the left node.

2c) Repeat this procedure for creating the right node.

Step 3: For each node in the tree, generate left and right sub trees using step 2.

Step 4: Store the IF.. THEN rules generated by the training module and intelligent agents.

Step 5: Select the intrusion data for testing using agents.

Step 6: Classify the data as normal or abnormal .

Step 7: If data is abnormal, use rules and agents and find the attack types.

Step 8: Inform the attack details to the administrator.

The results obtained from the application of the feature selection and classification algorithms proposed in this paper are used for performance analysis of the intrusion detection system proposed in this work. Table 3 shows the performance analysis for the J48 algorithm and the proposed Modified J48 algorithm based on training and testing time. From the analysis, it has been observed that the training and testing times are reduced in the proposed Modified J48 algorithm for Probe and DoS attacks. This is due to the use of intelligent agent and rules in the proposed algorithm

TABLE III
PERFORMANCE ANALYSES FOR J48 AND MODIFIED J48

Attacks	J48		Modified J48	
	Training time (sec)	Testing time (sec)	Training time (sec)	Testing time (sec)
Probe	0.65	0.34	0.60	0.29
DOS	1.76	0.67	1.65	0.62
others	0.66	0.31	0.31	0.25

Fig. 2 shows the decision tree obtained from the J48 algorithm generated using WEKA tool. Similar tree was created using the Modified J48 algorithm where selected features were used with the modified J48 algorithm.

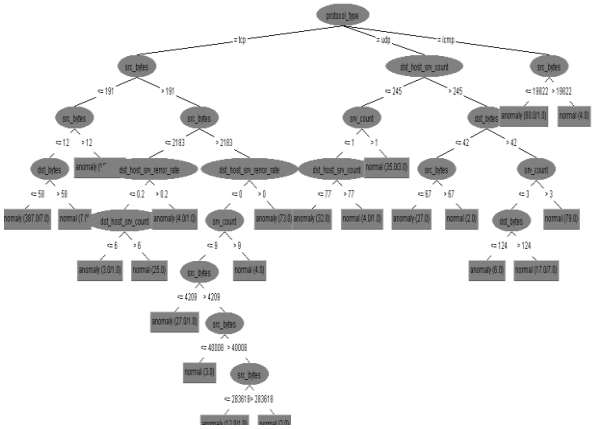

Fig. 2 Decision Tree Structure for Protocol Type

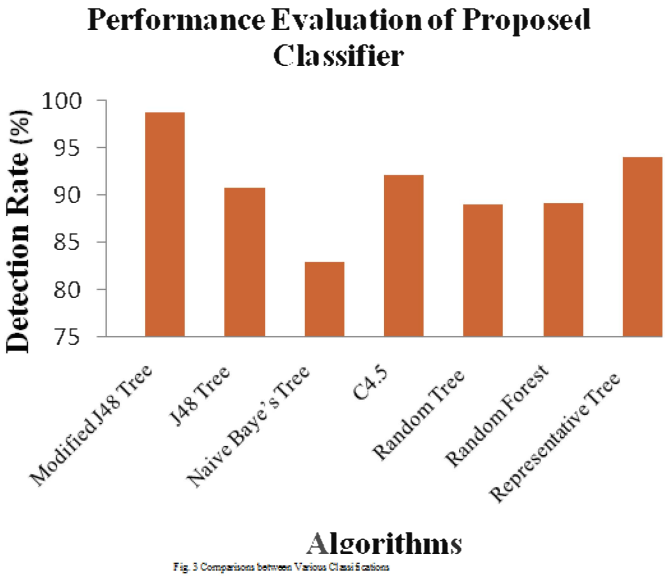

Fig. 3 Comparisons between Various Classifications

Fig. 3 shows the comparison of various classification algorithms which are available in the literature with the proposed Modified J48 classification algorithm. From this figure, it is observed that the proposed Modified J48 classification algorithm provides better accuracy than the existing algorithms.

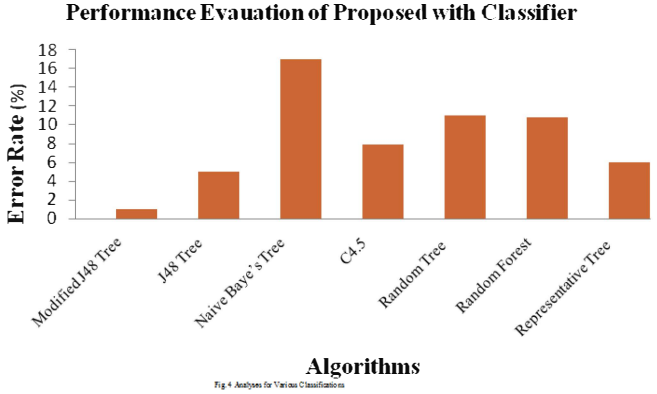

Fig. 4 Analyses for Various Classifications

Fig. 4 shows the error rate analysis for the various classifiers. From this figure, it is observed that the proposed Modified J48 algorithm provides reduced error rate when it is compared with the other existing algorithms. This is due to the fact that the rules and intelligent agents proposed in this Modified J48 algorithm are helpful for making effective decisions and hence it reduces the error rate.

VII. CONCLUSIONS

In this paper, a new IDS which uses two proposed algorithms for feature selection and classification respectively has been discussed. The feature selection algorithm uses genetic algorithm to produce the most important 9 features for effective classification. The proposed modified J48 classifier is useful to provide effective classification using intelligent agents and rules. The classification accuracy is increased in the proposed IDS and the time for classification is reduced due to the use of reduced features. Moreover, the error rate is also reduced by applying the proposed classification algorithm. Future works in this direction can be the use of support vector machine based classifier to improve the classification accuracy further.

REFERENCES

[1] Lunt T., 'Detecting Intruders in Computer Systems', Proceedings of the Conference on Auditing and Computer Technology, 1993, pp. 1-17.

[2] Teng H., Chen K. and Lu S., 'Adaptive Real Time Anomaly Detection using Inductively Generated Sequential Patterns', IEEE Computer Society Symposium on Research in Security and Privacy, California, 1990, pp278-284.

[3] Dokas P., Ertoz L., Vipin Kumar, Srivastava J. and Tan P., 'Data Mining for Network Intrusion Detection', National Science Foundation Workshop on Next Generation Data Mining, 2002, pp. 21-30.

[4] Gomez J., Fabio A. Gonzalez, MadhaviKaniganti and Dasgupta D. 'An Evolutionary Approach to Generate Fuzzy Anomaly (Attack) Signatures', Proceedings of the IEEE Workshop on Systems, Man and Cybernetics Society Information Assurance, 2003, pp.251-259.

[5] Manic M. and Wilamowski, 'Fuzzy Preference Approach for Computer Network Attack Detection', Proceedings of IEEE Conference on Fuzzy Systems, 2001, pp 267-275.

[6] Zhi-Song Pan, Songcan Chen, Gen-Bao HU and Dao-Qiang Zhang. "Hybrid Neural Network and C4.5 for Misuse Detection". *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, 2003, pp.2463-2467.

[7] Prema Rajeswari and Kannan. "An Active Rule Approach for Network Intrusion Detection with Enhanced C4.5 Algorithm". *Journal of Communications, Network and System Sciences,* 2008, pp.285–385.

[8] Christopher Kruegerl."Bayesian Event Classification for Intrusion Detection". *IEEE conference on Computer Security Applications*, 2003.

[9] Alexander Hofmann."Feature Selection for Intrusion Detection: An Evolutionary Wrapper Approach". *IEEE transactions on systems applications*, 2004.

[10] Mukhopadhyay I, Chakraborty M, Chakrabarti S and Chatterjee T. "Back Propagation Neural Network Approach to Intrusion Detection System". *International Conference on Recent Trends in Information Systems*, 2011.

[11] Jeich Mar, Yow-Cheng Yeh, I-Fan Hsiao, "An ANFIS-IDS against Deauthentication DOS Attacks for a WLAN", ISITA2010, Taichung, Taiwan, 2010, pp. 548-553.

[12] Rupali Datti and Bhupendra verma. "Feature Reduction for Intrusion Detection Using Linear Discriminant Analysis". *International Journal on Computer Science and Engineering*, vol. 02, no. 04, 2010, pp.1072-1078.

[13] Güneş Kayacık.H, Nur Zincir-Heywood A, Malcolm and Heywood I. "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets", 2010, pp: 1-6.

[14] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani. "A Detailed Analysis of the KDD CUP 99 Data Set". *Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications*, 2009.

[15] Huan Liu and Lei Yu. "Toward Integrating Feature Selection Algorithms for Classification and Clustering". *IEEE Transactions on knowledge and Data Engineering*, Vol. 17, NO. 4, 2005, pp: 491- 501,.

[16] Chen Y. and Wang J.Z. 'Support Vector Learning for Fuzzy Rule-Based Classification Systems', IEEE Transactions on Fuzzy Systems, Vol. 11, No. 6, 2003, pp. 716-728.

[17] Du Hongle, TengShaohua, Zhu Qingfang, "Intrusion detection Based on Fuzzy support vector machines", International Conference on Networks Security, Wireless Communications and Trusted Computing, 2009, pp. 639-642,.

[18] Yu-Ping Zhou, Jian-An Fang, Yu-Ping Zhou, "Intrusion Detection Model Based on Hierarchical Fuzzy Inference System", Second IEEE International Conference on Information and Computing Science, 2009, pp.144-147.

[19] Yong Zhong Li, Rushan Wang, Jing Xu, Ge Yang, Bo Zhao, "Intrusion Detection Method Based on Fuzzy Hidden Markov Model", Sixth IEEE International Conference on Fuzzy Systems and Knowledge Discovery, 2009, , pp. 470-474.

[20] ZihuiCheXueyunJi, "An Efficient Intrusion Detection Approach based on Hidden Markov Model and Rough Set", IEEE International Conference on Machine Vision and Human-machine Interface, 2010, pp. 476-479.

[21] YutengGuo, Beizhan Wang, Xinxing Zhao, XiaobiaoXie, Lida Lin, Qingda Zhou, "Feature Selection Based on Rough Set and Modified Genetic Algorithm for Intrusion Detection", IEEE International Conference on Computer Science & Education, 2010, pp. 1441-1446.

[22] Sanjay et al. "An Improved Network Intrusion Detection Technique based on k-Means Clustering via Nalve Bayes Classification". *IEEE-International Conference on Advances in Engineering, Science and Management*, 2012.

[23] Varun Chandola and Arindam Banerjee. "Anomaly Detection for Discrete Sequences: A Survey". *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, 2012.

[24] Adetunmbi Olusola A., Adeola Oladele S and Daramola O.Abosede. "Analysis of KDD '99 Intrusion Detection Dataset for Selection of Relevance Features". *Proceedings of the World Congress on Engineering and Computer Science*, vol.1, 2010, pp. 1-7,.