Reduced Feature Selection of KDDCUP99 Dataset Using Entropy, Gain and SVM Classifier

[1]Kusum Lata, [2]Mr. Manoj Yadav, [3]Mr. Kailash Patidar

[1]M.Tech, Research Scholar
[2,3]Assistant Professor
SSSUTMS, Sehore

ABSTRACT

Information about ensuring safety in the private sector or the government has become a need. Host intrusion detection systems monitor malicious activities and the management station is a technique that generates reports. Intrusion detection system, the availability of an attack and to protect the integrity of the data used for the detection of attacks. IDS detect intrusions using data mining techniques & other software techniques. The intrusion detection technique can efficiently expand the scope of defense of system. In this work we aim to improve efficiency for intrusion detection system. There are two phases in certain ways, in the first phase, we are using decision tree and SVM classifiers for classification of data and the second phase, we boost both the decision tree and SVM classifiers, and detect intrusion more than a single class classifier system. We are using KNN (k- nearest neighbor) classifier for misclassification data sets to improve detection rate. The kddcup99 dataset is used as a simulation set. KDDCUP 1999 benchmark dataset is used for testing the proposed algorithm and the results are promising and more important, especially high sensitivity, specificity and accuracy to create a model to achieve, that outperforms the existing methods are presented. The result shows that our proposed approach achieves better precision and detection rate by using KNN.

Keywords: - *IDS, Security Threats, SVM Classifier, KDDCUP99, Entropy, KNN classifier*

1. INTRODUCTION

System Security, privacy, reliability and availability of computer systems and its resources to protect the ability of references and unauthorized access to a computer system modification and use of the refuse safely to protect data and resources. Infiltration of the security aspects of a computer system that tries to attack the type of tolerance. For host intrusion detection system, a number of researchers, the most powerful methods for extracting information hidden in large data sets from the data mining methods, implemented. To apply data mining techniques in intrusion detection, preprocessing data collected by the first step. Then, in a special format for exchanging data mining process the configuration is used for classification and clustering. Rule-based classification model: a decision tree-based, Bayesian network-based or based on the neural network.

Information about ensuring safety in the private sector or the government has become a need. Vulnerability in the system and valuable information to attract the most attention attacked. In essence, the intrusion detection searches by abnormal data from normal data to divide which is a classification problem. Intrusion Detection System (IDS) has been applied to detect intrusion [1].

Intrusion detection technology identifies and deals with the use of contaminated computers that the system can be defined as. In this work they presented boosted of j48 decision tree and support vector machine classifiers for intrusion detection based on machine learning. They used a decision tree for classification of five-class data. First of all decision tree learn based upon training data and apply on test data then learn model categories data into normal class DOS class, probe class, U2r class & R2L class. Decision tree classified data or miss-classified data. They are also conducted an experiments with support vector machines (SVM) & then boosting of multiple classifiers. The decision tree using binary classifier and SVM is a single class classifier and also using one-against-one method in SVM for removing multiple classification problems. Then, they have applied boosting on misclassification data set to improve detection rate. KDDCUP 1999 benchmark dataset is used for testing the proposed algorithm and the results are promising and

more important, especially low false alarm rate and high detection rate to create a model to achieve, that outperforms the existing methods are presented [9]. The motivation for IDS developing absolutely secure systems is not possible because most existing systems have security flaws, Abuses by privileged insiders are possible & not all kinds of intrusions are known. Quick detection of intrusions can help to identify intruders and limit damage. IDS serve as a deterrent. The goal is to improve efficiency and accuracy for intrusion detection system.

Traditional intrusion protects techniques like firewalls way to control or encryption, have failed to fully keep safe (out of danger) systems from increasingly not simple attacks and malware. As an outcome, go into discovery systems (IDS) are used to make discovery of these attacks before they give stretched wide damage [1]. When building IDS, they need to take into account many issues such as data pre-processing, data collection, intrusion recognition, reporting, and response. Intrusion recognition is most important, Out of them. Audit data are making a comparison with detection models describe the patterns of intrusive or benign behavior, so that audit data can be identified both successful and unsuccessful intrusion attempts. Since Denning first made an offer a go into discovery, design to be copied in 1987, many research efforts have been gave one's mind to an idea on how to effectively and without error make discovery models. Between the late 1980s and the early 1990s, a joining together of expert systems and to do with facts as numbers comes, goes near was very pleasing to all. Discovery models were formed (from) from the lands ruled over knowledge of safety experts. From the mid-1990s to the late 1990s, getting acquaintance of normal or not normal behavior had curved from done with the hands to automatic. The artificial intelligence and machine learning techniques were used to build unearthing the close relation models from a group of training facts. Intrusion detection systems are being developed as devices to detect attacks and thus are becoming very important. IDS are useful in detecting successful intrusion, and also in monitoring suspicious activity and the attempts to break the security. Intrusion detection is the practice of observing and examining the actions going on in a system in order to identify the attacks and susceptibilities. The organization of the paper is done as follows: Section II presents the related work about the intrusion detection system, Section III present the proposed methodology. In section IV experimental analysis of proposed work is explained. Last section presents the overall conclusion of the research work.

2. RELATED WORK

The intrusion or intimidation cracks the security or hack our private information so to thwart from such issues a multiplicity of techniques and methodologies have been proposed by different researchers. In this paper literature of the work done is discussed below:

Abraham et al. [3] efficiently introduced intrusion detection system by using Principal Component Analysis (PCA) with Support Vector Machines (SVMs) as method to choose the optimum feature subset. They substantiate the efficiencies and the practicability of the proposed IDS system by abundant experiments on NSL-KDD dataset. The reduction method has been used to trim down the number of features in order to diminish the complication of the system. The experimental results show that the proposed system is proficient to speed up the process of intrusion detection and to minimize the memory space and CPU time rate.

Soni and Sharma [4] proposed a method which uses two methods C5.0 and artificial neural network (ANN) are utilized with feature selection. Feature selection methods will dispose of some inappropriate features while C5.0 and ANN acts as a classifier to categorize the data in either normal type or one of the five types of attack.KDD99 data set is used to train and test the models ,C5.0 model with numbers of features is producing improved results with all most 100% accurateness.

Jiankun Hu [5] introduced a new host-based anomaly intrusion detection methodology using discontiguous system call patterns, in an endeavor to increase detection rates whilst plummeting false alarm rates. The main idea is to apply a semantic structure to kernel level system calls in order to replicate inherent activities hidden in high-level programming languages which can help comprehend program anomaly behavior. Outstanding results were demonstrated using a multiplicity of decision engines evaluating the KDD98 and UNM data sets and a new, modern data set. The ADFA Linux data set was created as part of this research using a recent operating system and contemporary hacking methods and is now openly available. Additionally, the new semantic method possesses an inherent flexibility to mimicry attacks and demonstrated a high level of portability between dissimilar operating system versions.

Lee et al., [6] instigated decision tree method for detection of intrusion. In intrusion detection systems (IDSs) the data mining methods are useful to notice the attack particularly in anomaly detection. Intended for the decision tree, we employ the DARPA 98 Lincoln Laboratory assessment Data Set (DARPA Set) as the training dataset and the testing data set. The KDD' 99 Intrusion Detection data set is also based on the DARPA set. These three units are comprehensively used in IDSs. Consequently, they demonstrated the total process to engender the decision tree learned from the DARPA Sets. In this paper

also guesstimate the efficient value of the decision tree as the data mining method for the IDSs and the DARPA set as the learning dataset for the decision trees.

Raghuveer et al., [7] proposed a method which is divided into four steps: initial step, k-means clustering is used to generate different training subset then based on the obtained subset, various neuro-fuzzy data model are trained. Consequently, a vector for SVM classification is obtained and in last, classification using radial SVM is applied to detect the intrusion occurred or not. To make obvious the applicability and ability of the new method, the result of KDD dataset is confirmed in which it shows that the proposed methods produce better result than the BP, multiclass SVM and other approach such as decision tree etc.

Sushant Kumar Pandey [8] developed a hybrid model, which can detect intrusion by its action. We used an NSL-KDD data set, the multiclass problem and binary problems are 25% tested. This model can be used to guess the availability of intrusion, able to determine the scope of intrusions based on the transaction of data in the network; training requires optimal features of a network transaction. The accuracy of the model is better for both binary classes for the multiclass in NSL-KDD data set. The complication of false data alarm rates is the most significant challenge in the IDS system, and it may be the low false rate or high false rate. Proposed work also addresses this problem. The first step that data will be filtered by Vote algorithm, the Information Gain will get associated with a base learner, to choose the necessary features, which directly affects the accuracy of the model. It uses the following classifier: RandomTree, REPTree, RandomForrest AdaBoostM1, Meta Pagging, DesicionStump, J48, LMT, Bagging, and Naive Bayes. On the based on the proposed model, it is observed as low false rate, high accuracy.

Mohammed et al. [10] presented various data mining classification for handling false alerts in intrusion detection as reviewed. According to the result of testing many procedures of data mining on KDD CUP 99 that is no individual procedure can reveal all attack class, with high accuracy and without false alerts. The best accuracy in Multilayer Perceptron is 92%; however, the best Training Time in Rule based model is 4 seconds. It is concluded that, various procedures should be utilized to handle several of network attacks.

Doreswamy [10] proposed a two-phase model to detect and categorize anomalies. First, we selected Random Forest based on the highest accuracy-score out of eleven commonly used algorithms tested with the same set of data. The RF is used to detect anomalies and generate an extra feature named “attack-or-not”. Secondly we fed Neural Network with the data having “attack-or-not” feature to differentiate attack categories, which will help treating each type accordingly. The model performance was good, it scored 0.99 for both Precision and Recall in anomaly detection phase and 0.93 for Precision and 0.88 for Recall in attack categorization phase. We used UNSW-NB15 data set in our study.

Venugopalan [11] presented a novel classification algorithm based on distance measure and Relief-F feature weighting. The performance measures of intrusion detection are compared with the commonly used classification algorithms such as Naı¨ve Bayes, Decision Tree and Support Vector Machine (SVM) and the proposed algorithm outperforms the above mentioned algorithms in terms of Detection Rate, Accuracy, False Alarm Rate, F-Score and Mathews Correlation Coefficient. The proposed algorithm is tested using a benchmark dataset (KDDcup99 dataset) and a real traces dataset (Kyoto 2006 ? dataset). This study also intend to compare the execution time for various classifiers and the parallel performance of NADA since NADA outperforms all the other classifiers in terms of serial execution time. The algorithm is parallelized and the results are presented in terms of execution time with various data size, speed up and efficiency.

3. PROPOSED WORK

In the classification of big data domains, sometimes concealed data possibility has been occur while the classification process. Therefore generated features contain the false correlations which are not up to the mark of finding the process of intrusion detection. The weakness of extra features is that it restrains large time for the process of computing and it impacts the precisions of IDS. Here feature selection advances the more classification precision by searching for the best features, which best classifies the training data. So in the proposed system probability has been calculated of the each independently attributes, then entropy has been deliberated and lastly information gain has been calculated for each every attributes disjointedly. And here they applied some logical implies that if calculated gain is very less (gain<0.15) then that type of attribute will not be contributed for the data preprocessing. So,in conclusion 18 attributes found whose gain was higher and that process is done in feature extraction and feature reduction.

ALGORITHM STEPS:

Step 1: Build X1reduced datasets from a database.
Step 2: Set a learning algorithm to independent pattern for test dataset.
Step 3: Set a learning algorithm to independent pattern training dataset.
SVM_struct=SVM_train(X1(train(:,i1),:),groups(train(:,1))
Step 4: Object with unknown found to do with each of the X1 classifiers predictions.
Step 5: Select the most repetitively predicted samples.

KNN steps:

Step1: Initialize population = *X1*
Step2: Apply genetic search into selected dataset
Step3: Apply KNN classifier for testing of all five data which is classified or misclassified data.
Step4: Each attribute will organize as their ranks.
Step5: Higher ranked attribute will select.
Step6: Apply KNN () on the each five subset of the attributes for enhance the accuracy level.
Step7: If KNN_classifier (class_knn)>knn_classifer (class_knn)
data_class = class_knn;
else
data_class = class_knn;
Step8: Carry out the reproduction
Step9: Apply crossover operator
Step10: Carry out mutation then produce new population *X'1*
Step11: Analyze the local maxima for each category. reiterate the steps till iteration is not finished
Step12: For each test *X'1* start all trained base models then prediction of result by combining of all trained models and separate the misclassified by optimized KNN. Classification: Majority occurrences.

Here block diagram shows that the working of proposed approach, where at initial state KDD99 dataset is selected for the processing, then into next stage entire dataset is logically separate for the moment due to it is containing string fields as well as numeric fields, so in the designing approach they developed separate macnizum for string and numeric data.

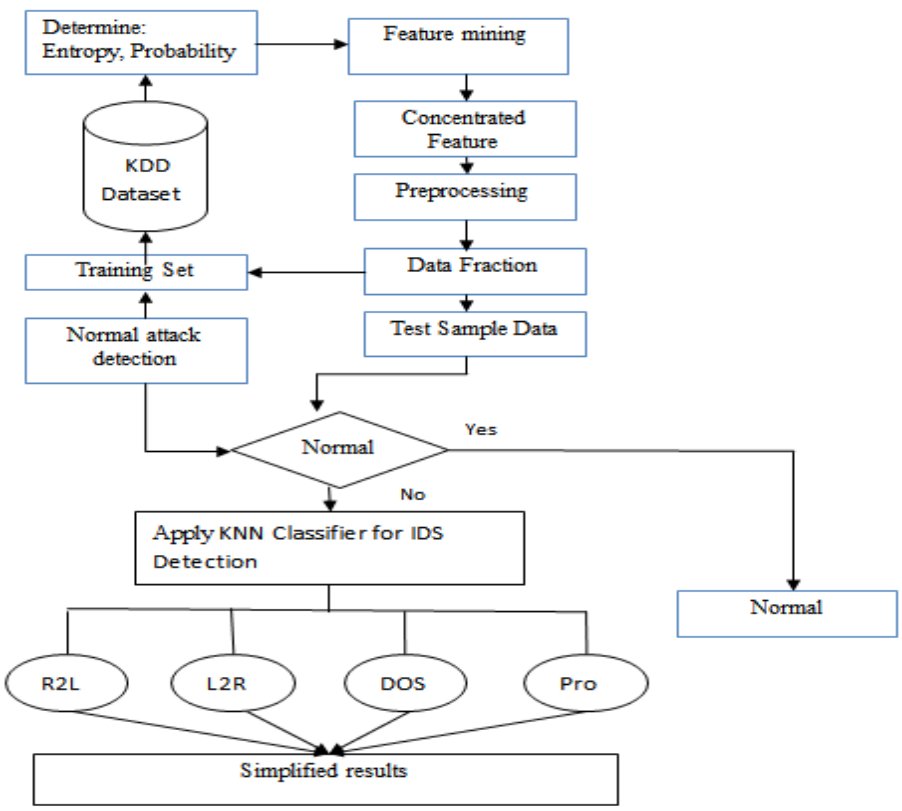

Figure 1: Block Structure of proposed work

Pre-processing: It converts the data which is more reliable for unsupervised learning by removing the labels from the dataset.

Data fraction: Preprocessed data are used to partition into training & testing sets samples.

Detection of Normal: in this step normal data is separated from the training data sample, here training process is done by SVMtrain() built in function of the MATLAB

And if the normal class has been easily detected then its goes to the separately normal class otherwise if not detected then it will go to the KNNclassify() classifier and in this process each class has been accurately predicted with their own identity, after successful prediction the result analysis approach follows for the detected intrusions.
The computation process is done accordingly:
In example we have to find out attack type of tuple given below:
X=(Source_bytes=200,dest_bytes=3000,logged_in=1,root_shell=0,count=50,dst_host_count=400,dst_host_rerror_rate= 0).
$P(C_i)$, the prior probability of each class, can be computed based on the training tuples:
P (type = normal) = 7/20 = 0.35
P (type = DOS) = 5/20 = 0.25
P (type = probe) = 5/20 = 0.25
P (type = R2L) = 3/20 = 0.15
To compute $P(\boldsymbol{X} \mid C_i)$, for i = 1, 2, 3, 4 we compute the following conditional probabilities:
P (source_bytes = 200 | type = normal) = 5/7 = 0.7143
P (source_bytes = 200 | type =DOS) = 0
P (source_bytes = 200 | type = probe) = 0
P (source_bytes = 200 | type = r2l) = 0
P (dst_bytes = 3000 | type = normal) = 4/7 = 0.57143
P (dst_bytes = 3000 | type = DOS) = 1/5 = 0.2
P (dst_bytes = 3000 | type = probe) = 1/5 = 0.2
P (dst_bytes = 3000 | type = r2l) = 0

P (logged_in = 1 | type = normal) = 5/7 = 0.7143
P (logged_in = 1 | type = DOS) = 1/5 = 0.2
P (logged_in = 1 | type = probe) = 1/5 = 0.2
P (logged_in = 1 | type = r2l) = 1/3 = 0.33
P (root_shell = 0 | type = normal) = 5/7 = 0.7143
P (root_shell = 0 | type = DOS) = 4/5 = 0.8
P (root_shell = 0 | type = probe) = 4/5 = 0.8
P (root_shell = 0 | type = r2l) = 1/3 = 0.33
P (count = 50 | type = normal) = 4/7 = 0.57143
P (count = 50 | type = DOS) = 1/5 = 0.2
P (count = 50 | type = probe) = 3/5 =0.6
P (count = 50 | type = r2l) = 0
P (dst_host_count = 400 | type = normal) = 5/7 = 0.7143

P (dst_host_count = 400 | type = DOS) = 4/5 = 0.8
P (dst_host_count = 400 | type = probe) = 1/5 = 0.2
P (dst_host_count = 400 | type = r2l) = 1/3 = 0.33
Using the above probabilities, we obtain
$P(\boldsymbol{X} \mid type = normal)$ = P (Source_bytes = 200 | type = normal) × P (dst_bytes = 3000 | type = normal) × P (logged_in = 1 | type = normal) × P (root_shell = 0 | type = normal) × P (count = 50 | type = normal) × P (dst_host_count = 400 | type = normal)
= 0.7143×0.57143× 0.7143× 0.7143 × 0.57143× 0.7143 = 0.085
To find the class, C_i, that maximizes $P(\boldsymbol{X} \mid C_i)P(C_i)$, we compute:

$\boldsymbol{P(X \mid type = normal)}$ **P (type = normal) = 0.085 × 0.35 = 0.02975.**

Therefore, the naïve Bayesian classifier predict attack type = *Normal* for tuple $\boldsymbol{X}$.
After that entropy has been calculated as follows:

Entropy H(X) = $-p_1\log_2 p_1 - p_2\log_2 p_2 - p_3\log_2 p_3 - \ldots\ldots\ldots\ldots\ldots -p_n\log_2 p_n$ *…eq(5)*

$$= -\sum_{i=1}^{m} pilog2\ pi$$

In the equation 1, the class-wise probability has been settled then entropy has been calculated of each individual attributes. Then gain was calculated as follows:

Gain = Entropy(X) - Entropy (X|Y) ……eq (6)

So as per the above process feature reduction has been done, where gain was higher than that attribute has been qualified for the process and less gain was reduced from dataset.

Table 1: Result obtained from the solution of example

Src_bytes	Dst_byte	Logged_in	Root_shell	Count	Dst_host count	Dst_host	Type
200	3000	1	0	50	400	0	Normal
400	6000	1	0	50	400	0	Normal
200	6000	1	0	50	400	0	Normal
200	3000	1	0	0	400	0	Normal
200	6000	0	1	1	50	0	Normal
400	3000	0	1	1	50	0	Normal
200	3000	1	0	50	400	0	Normal
2000	0	0	0	600	400	0	Dos
2000	0	0	0	600	400	0	Dos
2000	3000	1	1	50	50	0	Dos
2000	6000	0	0	600	400	0	Dos
2000	0	0	0	600	400	0	Dos
50	0	0	0	50	50	0	Probe
50	0	0	0	50	50	0	Probe
50	3000	0	0	600	50	0	Probe
50	6000	1	1	1	400	0	Probe
50	0	0	0	50	50	0	Probe
3000	6000	0	1	1	50	0	R2l
3000	6000	0	1	1	50	0	R2l
3000	0	1	0	0	400	0	R2l

4. EXPERIMENTAL RESULTS & ANALYSIS

The KDD99cup data set used for the purpose of experimental research analysis, as they know that KDD 99 dataset [12] has been widely used for the evaluation of signature based intrusion detection. In the novel approach they have used KDDCup'99 intrusion detection dataset, which contains 26167 records with.50:50 training ratio.
Attack types are four categories:
1. Denial of Service (DoS)
2. Remote to Local (R2l)
3. User to Root (U2R)
4. Probe

The proposed IDS has been implemented in MATLAB2012A [14] tool and the machine configuration is Intel I3 core 2.20Ghz processor, with 4GB RAM, windows 7 home basis. The proposed methodology have first used the partially ID3 algorithm for the feature reduction from the KDD, the SVMtrain function is use for training purpose of the trained sample, then kNN is use for the clustering and classification process for the classify or misclassify of the data, where GA is ensemble with KNN to enhance the best classification rate and optimized the result in very efficient manner. Here classification has five classes' data which is (normal, dos, u2r, r2l, and probe). This did classified or classified.
KNN classified data which were misclassified by alone SVM and KNN then applying KNN on multiple classifiers. This approach is focused on misclassified classifiers and putted extra efforts to optimize best classified of the category until they are not accurately classified.
Then method has been tested on full (41 attributes) dataset as well as in reduced dataset (18 attributes), and used measurement parameters are:

Sensitivity, specificity and accuracy, and method is compared with SVM, KNN and found that proposed method produced most accurate result into maximum cases.

Here figure 2(a) & figure 2(b) shows that the main GUI environment of the implemented all methods along with proposed approach, here we clearly observe that the proposed approach yield more accurate output compared to the other previous developed methods.

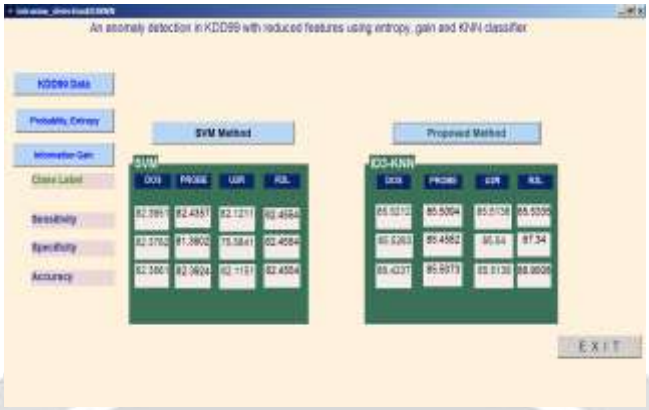

Figure 2 (a): Main Simulation GUI for 41 dataset

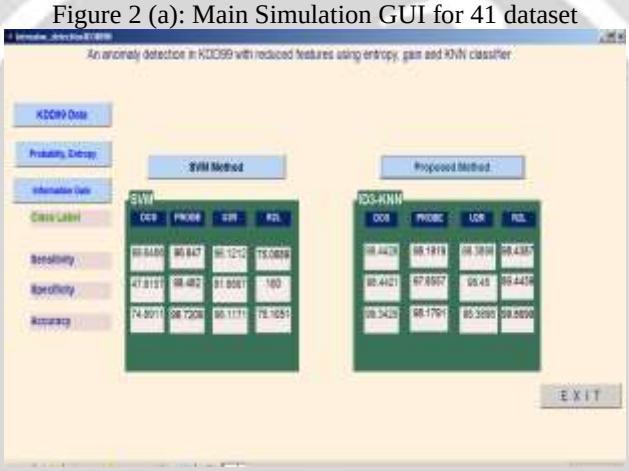

Figure 2(b): Main Simulation GUI for 18 dataset

Table 2, table 3 and table 4 illustrated that the sensitivity, specificity and accuracy comparison table of SVM, KNN and proposed ID3, KNN approaches for reduced 18 attributes, we have also examine the same scenario for the 41 full attributes, and there we found that the all methods gave the less accuracy level and taking much time as compare to reduced attribute.

Result for 18 attributes:

Table 2: Sensitivity for 18 attributes

Sensitivity (18attribute)		
Attacks	**SVM**	**Proposed**
DOS	99.6488	98.4428
PROBE	96.647	98.1919
U2R	96.1212	98.3898
R2L	75.06889	98.4387

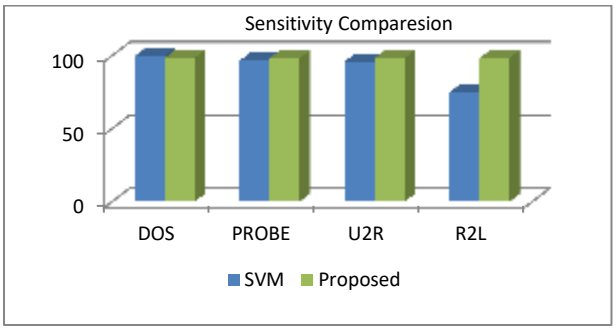

Figure 3: **Sensitivity (18attribute)**

Table 3: **Specificity (18attribute)**

Specificity(18attribute)		
Attacks	**SVM**	**Proposed**
DOS	47.8157	98.4421
PROBE	98.482	97.8557
U2R	91.6667	98.45
R2L	100	86.4439

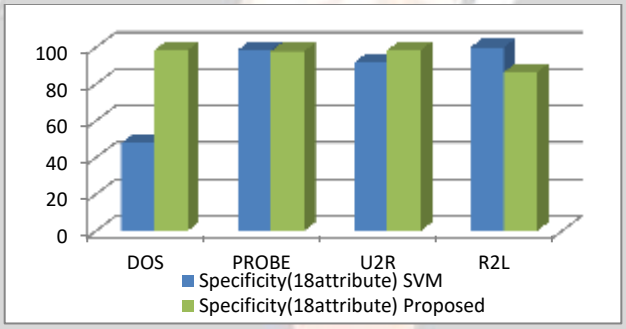

Figure 4: Specificity (18attribute)

Table 4: Accuracy (18attribute)

Accuracy(18attribute)		
Attacks	**SVM**	**Proposed**
DOS	74.8911	99.3425
PROBE	96.7209	98.1791
U2R	96.1171	98.3898
R2L	75.1051	99.8699

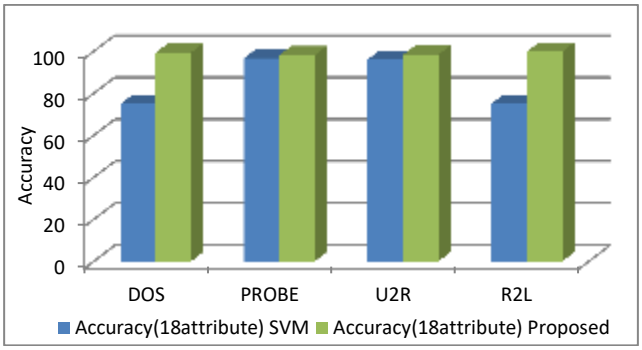

Figure 5: Accuracy (18attribute)

Result for 41 attributes:

Table 5, table 6 and table 7 illustrated that the sensitivity, specificity and accuracy comparison table of SVM, KNN and proposed ID3, KNN approaches for the 41 full attributes, and there we found that the all methods gave the less accuracy level and taking much time as compare to reduced attribute.

Table 5 Sensitivity for 41 attributes

Sensitivity(41attribute)		
Attacks	SVM	Proposed
DOS	82.3951	85.5212
PROBE	82.4357	85.5094
U2R	82.1211	85.5138
R2L	82.4554	85.5335

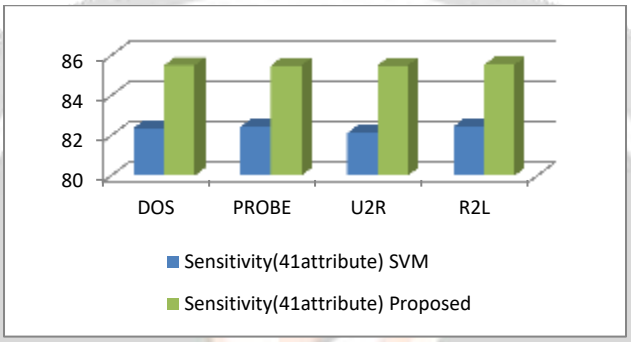

Figure 6: Sensitivity (41attribute)

Table 6 Specificity for 41 attributes

Specificity(41attribute)		
Attacks	SVM	Proposed
DOS	82.3762	85.5263
PROBE	81.3602	85.4582
U2R	75.5841	85.54
R2L	82.4554	67.34

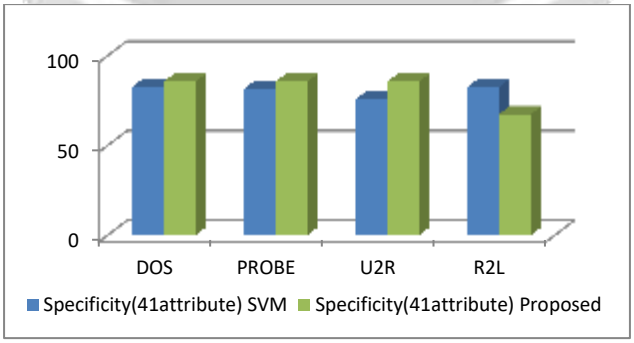

Figure 7: Specificity (41attribute)

Table 7: Accuracy (18attributes)

Accuracy(41attribute)		
Attacks	SVM	Proposed
DOS	82.3861	86.4237
PROBE	82.3924	85.5073
U2R	82.1151	85.5138
R2L	82.4554	86.9508

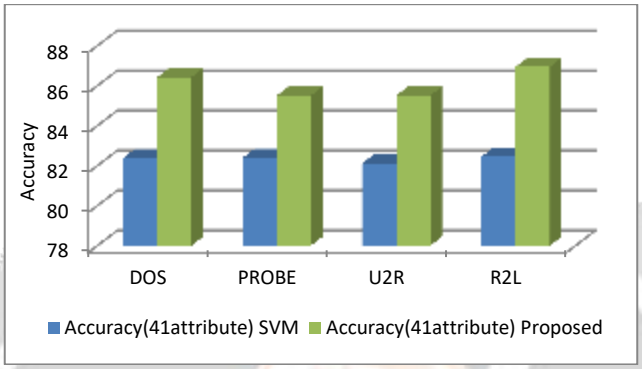

Figure 8: Accuracy (18attributes)

And in the figure 3, figure 4, figure 5, figure 6, figure 7 and figure 8 they shows the bar graph analysis of all methods of the given table, as they clearly showing that the accuracy level of novel approach is generating more improved result for 18 reduced and 41 all dataset. We have also prepared the confusion matrix for the reduced 18 dataset of the proposed method which are shown in table 8 below.

Table 8 Confusion Matrix for 18 features

SVM	ID3,KNN
Confusion Matrix of 'dos' 6810 3261 24 2988 0 0	Confusion Matrix of 'dos' 13667 1 1 12498 0 0
Confusion Matrix of 'probe' 12135 8 421 519 0 0	Confusion Matrix of 'probe' 25107 6 66 988 0 0
Confusion Matrix of 'u2r' 12564 1 507 11 0 0	Confusion Matrix of 'u2r' 26143 0 16 8 0 0
Confusion Matrix of 'r2l' 9807 0 3257 19 0 0	Confusion Matrix of 'r2l' 26123 5 3 36 0 0

5. CONCLUSION

To develop the system for exposure and detection of severe types of intrusion this may corrupt or destroy the resources used for the access. Several author have been work in the field of intrusion detection and develop the system which can reduce the true and false alarm rate but in this dissertation we develop a novel method by applying multiple classification and feature reduction techniques. In this we use ID3, SVM and KNN approach for intrusion detection and apply these methods on the benchmark KDD Cup 99 Intrusion data. We have first uses ID3 (decision tree) for feature reduction and also conducted experiments with support vector machines (SVM) & then last apply KNN as classifier for the detection of intrusions. The analysis of the methodology is done in well-known simulator MATLAB2012 using the performance metrics sensitivity, specificity and accuracy in which our method results outperform the existing methods. In future work, develop a method suing ensemble multiple classifier which can better expose the intrusion and greatly enhance the performance of the system.

REFERENCE

[1] Wu Shelly Xiaonan and Banzhaf Wolfgang "The use of computational intelligence in intrusion detection systems: A review," ELSEVIER, 2010.
[2] Aneetha A.S. and Bose S. "The Combined approach for Anomaly Detection using Neural Networks and Clustering Techniques," CSEIJ, Vol.2, No.4, pp. 37-46, 2012.
[3] Heba F. Eid, Ashraf Darwish, Aboul Ella Hassanien and Ajith Abraham Principle Components Analysis and Support Vector Machine based Intrusion Detection System", in proceeding of IEEE (2010)
[4] Pratibha Soni, Prabhakar Sharma "An Intrusion Detection System Based on KDD-99 Data using Data Mining Techniques and Feature Selection", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-4 Issue-3, July 2014
[5] Jiankun Hu, "A Semantic Approach to Host-Based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns", IEEE Transactions on Computers, vol.63, no. 4, pp. 807-819, April 2014, doi:10.1109/TC.2013.13
[6] Joong-Hee Lee, Jong-Hyouk Lee, Seon-Gyoung Sohn Jong-Ho Ryu and Tai-Myoung Chung "Effective Value of Decision Tree with KDD 99 Intrusion Detection Datasets for Intrusion Detection System",. Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on(Volume:2)Feb. 2008 Page(s):1170 - 1175 ISSN :1738-9445
[7] A. M. Chandrasekhar, K. Raghuveer " Intrusion Detection Techniques by using K-means, Fuzzy Neural network and SVM classifier", ICCCI-2013, Jan. 04-06, 2013, Coimbatore, INDIA.
[8] Sushant Kumar Pandey, "Design and performance analysis of various feature selection methods for anomaly-based techniques in intrusion detection system" Security Privacy. 2019;2:e56. wileyonlinelibrary.com/journal/spy2 © 2019 John Wiley & Sons, Ltd. , pp 1 – 14.
[9] Shaker El-Sappagh, Ahmed Saad Mohammed, Tarek Ahmed AlSheshtawy, "CLASSIFICATION PROCEDURES FOR INTRUSION DETECTION BASED ON KDD CUP 99 DATA SET", International Journal of Network Security & Its Applications (IJNSA) Vol. 11, No.3, May 2019.
[10] Mohammad Kazim Hooshmand, Doreswamy, "Machine Learning Based Network Anomaly Detection", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019.
[11] Ashok Kumar, D., & Venugopalan, S. R. (2019). A design of a parallel network anomaly detection algorithm based on classification. International Journal of Information Technology. doi:10.1007/s41870-019-00356-0.
[12] Rashmi Singh, Diwakar Singh, "A Review of Network Intrusion Detection System Based on KDD Dataset", Int J Engg Techsci Vol 5(1) 2014, 10 - 15