Intrusion Detection using Supervised Learning with Feature Set Reduction

Yogendra Kumar Jain
Head of Department
Dept. of Computer Science & Engineering
S.A.T.I.,Vidisha,(M.P.), India

Upendra
Research Scholar (M.Tech)
Dept. of Computer Science & Engineering
S.A.T.I., Vidisha,(M.P.), India

ABSTRACT
Intrusion detection systems intend to recognize attacks with a low false positive rate and high detection rate. Many feature selection methods introduced to eliminate redundant and irrelevant features, because raw features may abbreviate accuracy or robustness of classification. In this paper we are proposing the information gain technique for the selection of the features. A feature with the highest information gain is the criteria for the selection of the features. We reduced the features of the data set than run the algorithm. Result show that drastically decreased in learning time of the algorithm without compromising the accuracy which is desirable for good IDS.We analyse two learning algorithms (NB and BayesNet) for the task of detecting intrusions and compare their relative performances. We comment on the suitability of the BayesNet algorithm for the intrusion detection task based on its high accuracy and high true positive rate. We finally state the usefulness of machine learning to the field of intrusion detection.

Keywords
Intrusion Detection, Machine Learning,BayesNet, NB, KDD 99

1. INTRODUCTION
If an intrusion is detected quickly enough, an intruder can be identified quickly and ejected from the system before any damage is done or any data are compromised. Even if the detection is not sufficiently timely to pre-empt the intruder, the sooner that the intrusion is detected, the less is the amount of damage done and more quickly that recovery can be achieved. An effective intrusion detection system can serve as a deterrent, so acting to prevent intrusion detection enables the collection of information about intrusion techniques that can be used to strengthen the intrusion prevention facility.

The significant increase of our everyday life dependency to Internet-based services has intensified the survivability of networks. On the other hand, the number of attacks on networks has dramatically increased during the recent years. Consequently, interest in network intrusion detection systems has increased among the researchers. During the past number of years, machine learning and data mining techniques have received considerable attention among the intrusion detection researchers to address the weaknesses of knowledgebase detection techniques. This has led to the application of various supervised and unsupervised techniques for the purpose of intrusion detection.

In this paper, we present the application of machine learning to intrusion detection. We analyse two learning algorithms (NB and BayesNet) for the task of detecting intrusions and compare their relative performances. There is only available data set is KDD data set for the purpose of experiment for intrusion detection. KDD data set [2] contain 42 attributes. In general intrusion detection system uses the all the attributes available in the data for the purpose of the intrusion detection. Using all the attributes of the data set causes increase in the learning time of the algorithm means late detection of the intrusion by IDS. Here we are proposing the feature reduction of the data set using information gain. After reduction of the feature learning time of the algorithm decreased drastically without compromising the accuracy of the IDS which is desirable. We then comment on the suitability of the BayesNet algorithm for the intrusion detection task based on its high accuracy and high recall. We finally state the usefulness of machine learning to the field of computer security and also comment on the security of machine learning itself.

2. RELATED WORKS
In 1980, the concept of intrusion detection began with Anderson's seminar paper [4]; he introduced a threat classification model that develops a security monitoring surveillance system based on detecting anomalies in user behavior. In 1986, Dr. Denning proposed several models for commercial IDS development based on statistics, Markov chains, time-series, etc [5]. In the early 1980's, Stanford Research Institute (SRI) developed an Intrusion Detection Expert System (IDES) that monitors user behavior and detects suspicious events. R.C. Staudemeyer [1] compares the some machine learning algorithm using feature set reduction. In this paper they proposed a minimum feature set which can easily extracted from network traffic. They compared various algorithms on the KDD 99 data set. Meng Jianliang [7] used the K Mean

algorithm to cluster and analyze the data. He used the unsupervised learning technique for the intrusion detection. Li Tian, Wang Jianwen [8] showed that because of the k-means algorithm's short comings about dependence and complexity, he puts forward an improved clustering algorithm through studying on the traditional means clustering algorithm. The new algorithm learns the strong points from the k-means and improved relations trilateral triangle theorem. Gary Stein [9] applied the genetic algorithm and the decision tree algorithm for the intrusion detection. He used the genetic algorithm technique for the feature reduction. Rung- Ching Chen [10] used the rough set theory and support vector machine for the intrusion detection. He used the Rough set theory for the reduction of the dimensions. After that features were selected by rough set theory will be sent to support vector machine to detect intrusion.Lin NI, Hong Ying Zheng [11] in 2007 done the intrusion detection based on unsupervised clustering and Chaos Simulated Annealing Algorithm. Jiong Zhang and Mohammad Zulkernine [12] done the intrusion detection using the random forest algorithms in anomaly based NIDS. Cuixio Zhang, Guobing Zhang, Shanshan Sun [13] used the missed approach for the intrusion detection. He designed the mixed combining the anomaly detection and misuse detection In this model the anomaly detection module is built using unsupervised clustering method and the algorithm is an improved algorithm of K means clustering algorithm. Juan Wang, Qiren yang and Dasen Ren [14] used the decision tree algorithm BayesNet for the intrusion detection.

3. BACKGROUND

Intrusion Detection [15] is used to detect violation of a security policy of an organization. These violations may be caused by people external to the organization (i.e. attackers) or by employees of the organization (i.e. insiders).Although progress has been made to detect violations by attackers, insider violations are difficult to detect. Intrusion detection can be divided into two types [16]: One is anomaly detection. It firstly stores users' normal behaviours into feature database, then compares characters of current behaviour with characters of feature database. If the deviation is large enough, we can say that the current behaviour is abnormal. Although having a low false negative rate and a high false alarm rate, it can detect unknown types of attacks; The other is misuse detection. It establishes a feature library according to the known attacks, and then matches the happened behaviours to detect attacks. It can only detect known types of attacks, but is unable to detect new types of attacks. Therefore misuse detection has a low false positive rate and a high false negative rate.

3.1 Data Mining Approach

Data Mining [17] is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner." During the process of data mining, many machine learning algorithms are available for choosing. Depending on whether the class labels are provided for learning, these machine learning algorithms can be classified as either supervised or unsupervised.

3.1.1 Supervised learning

Trained with data bearing class labels indicating to which subcategories they belong or what real-valued properties they have, a supervised learning algorithm [18] tries to predict the most likely labels for new test data. There are two major subcategories for supervised learning: *Classification* is to predict the class membership as one of a finite number of discrete labels. *Regression* is to predict the output value as one of a potentially infinite set of real valued points. Some generally used supervised learning algorithms:-

3.1.1.1 BayesNet

The BayesNet algorithm is performed for learning task, where a training set with target class is provided. Inference of decision trees using a set of conditions over the attributes. Classification of new examples is carried out by applying the inferred rules. Although the original algorithm contains numerous free parameters, only the number of bootstrap iterations was used in our evaluation.

3.1.2 Unsupervised Learning

In unsupervised learning,[3] the data are not labeled, which makes it hard to tell what counts as good. It is less natural, but much more revealing, to view unsupervised learning as supervised learning in which the observed data is the output and for which there is no input". The model generating the output must either be stochastic or must have an unknown and varying input in order to avoid producing the same output every time. From the perspective of machine learning, the searching for clusters is unsupervised learning. To perform clustering is to try to discover the inner nature of the data structure as a whole, and to divide the data into groups of similarity. From the viewpoint of data mining, clustering is the partitioning of a data set into groups so that the points in the group are similar as possible to each other and as different as possible from points in other groups.

4. PROBLEM STATEMENT

The significant increase of our everyday life dependency to Internet-based services has intensified the survivability of networks. On the other hand, the number of attacks on networks has dramatically increased during the recent years [16]. Consequently, interest in network intrusion detection systems has increased among the researchers. Intrusion detection systems aim to identify attacks with a high detection rate and a low false positive. If an intrusion is detected quickly enough, an intruder can be identified quickly and ejected from the system before any damage is done or any data are compromised [3].Even if the detection is not sufficiently timely to preempt the intruder, the sooner that the intrusion is detected, the less is the amount of damage done and more quickly that recovery can be achieved. Machine learning techniques have been applied

to the field of intrusion detection. They can learn normal and anomalous patterns from training data and via Feature selection improving classification by searching for the subset of features which best classifies the training data to detect attacks on computer system. The quality of features directly affects the performance of classification. Most of the earlier intrusion detection approaches make use of all the features in the data to detect the intrusion which cause take more time to detect. It degrades the performance of IDS [6]. All the feature of the data are not relevant. Selecting a minimum set of core features for automatic network intrusion detection is a challenging problem. Many feature selection methods introduced to remove redundant and irrelevant features, because raw features may reduce accuracy or robustness of classification. In this paper we are proposing the information gain concept for reduction of the feature. After using the less feature of data cause improve the performance of the IDS and extensively decreases the computer resources like memory and CPU utilization which are required to identify an attack.

5. PROPOSED APPROACH

KDD 99[19] data set contain the 41 feature .Generally all the IDS using the all 41 feature for the intrusion detection. All the feature are not relevant for the intrusion detection. For the selection of minimum number of feature for intrusion detection we have used the information gain technique without compromising the accuracy of the algorithm. After selecting the features we pass the data set to the algorithm (BayesNet and NB) for training and testing. We have used the 10 fold cross validation technique for the testing. Result shows that drastically decrease in the learning time of the algorithm and increase in accuracy and TPR, which is desirable for the good intrusion detection system. We are proposing the selection of only 11 features of the KDD99 data set using information gain for the detection. We compared the results of algorithm BayesNet and NB executing the algorithm with all features and with selected 11 feature .We have compared the result of our experiment with R.C. Staudemeyer [1] result. He had also done the experiment with selected 11 features of the KDD data set. Our experiments show the better accuracy and high TPR than his experiment. Algorithm used (BayesNet and NB) in the work and the testing method cross validation already discussed in the paper. The summary of the experiment is given below.

Table 1. Experimental summary

Algorithm	Feature Selection	Cross Validation
BayesNet	All Feature Used	10 fold
NB	All Feature Used	10 fold
BayesNet	Info gain for feature selection	10 fold
NB	Info gain for feature selection	10 fold

5.1 Feature Selection

In order to make IDS more efficient, reducing the dimensions and data complexity have been used as simplifying features. Feature selection can reduce both the data and the computational complexity. It can also get more efficient and find out the useful feature subsets. It is the process of choosing a subset of original features so that the feature space is optimally reduced to evaluation criterion. The raw data collected is usually large, so it is desired to select a subset of data by creating feature vectors that Feature subset selection is the process of identifying and removing much of the redundant and irrelevant information possible. This results in the reduction of dimensionality of the data and thereby makes the learning algorithms run in a faster and more efficient manner. It also reduces the size of hypothesis space and in some cases; it also reduces the storage requirement. Attribute Selection (also known as Feature Reduction or Feature Subset Selection) is an important task during any machine learning exercise (especially classification). Usually, the available data for machine learning analysis is multidimensional and the numbers of dimensions (i.e. features or contribute at all to the classification task. For a dataset with k attributes, the size of the hypothesis space is K2. For a small k, an exhaustive search to find out the best suited hypothesis is possible. However, this task becomes non trivial as the value of k increases .The experiments conducted in this paper use the Information Gain attribute selection method. This method is explained in the following sub-section

5.2 Information Gain

Information Gain measure is used to determine how instrumental a particular is attribute in correctly classifying the training data. Information gain is based on the concept of entropy which is widely used in the information theory domain. Given a collection of instances S, containing positive and negative examples of some target concept. The entropy of S relative to this Boolean classification is given by:

$$\text{Entropy}(S) = - P1 \log 2P1 - P2 \log 2 P2$$

Where P1 is the proportion of positive examples in S and P2 is the proportion of negative examples in S.
For a target concept with c different possible values for the classes, the entropy can be defined as:

$$\text{Entropy (S)} = \sum_{i=1} -Pi \ log2 \ Pi$$

Where pi is the proportion of S belonging to class i.Based on the above definition of entropy, information gain G of an attribute A is defined as:

$$\text{Gain (S, A)} = \text{Entropy (S)} - \sum_{Ve\,Values(A)} |Sv| \div |S| \ Entropy\ (Sv)$$

where Values (A) is the set of all possible values for attribute A and Sv is the subset of S for which A has value v. Information gain is thus the reduction in entropy caused by partitioning the examples according to an attribute In this paper, Weka's implementation of the Information gain attribute selector (called Info Gain Attribute Eval) was used to determine the effectiveness of attributes and the attributes were ranked in decreasing order of information gain values. The first 11 attributes and the Class attribute were then used in the experiment for the learning task. Using the method above for calculation of information gain, we calculate the info gain of the all the attribute of the KDD99 data set. In our proposed technique we are using the KDD99 dataset with these selected features and train and test the algorithm. For the testing we are using the 10 fold cross validation.

Table 2. Information Gain of the all attributes

Info Gain	Attribute name
0.939935	src_bytes
0.832597	service
0.807751	count
0.781945	dst_bytes
0.582982	logged_in
0.441058	dst_host_srv_diff_host_rate
0.421889	dst_host_diff_srv_rate
0.404109	dst_host_count
0.365589	srv_count
0.328112	flag
0.306328	dst_host_serror_rate
0.304958	dst_host_srv_serror_rate
0.303309	diff_srv_rate
0.297816	same_srv_rate
0.29739	serror_rate
0.296763	srv_serror_rate
0.293889	dst_host_same_srv_rate
0.278352	dst_host_diff_srv_rate
0.269263	dst_host_srv_count
0.269243	protocol_type
0.193463	srv_diff_host_rate
0.060201	dst_host_rerror_rate
0.055125	dst_host_srv_rerror_rate
0.050177	hot
0.041616	num_compromised
0.024322	srv_rerror_rate
0.023832	rerror_rate
0.021349	duration
0.008183	num_failed_logins
0.002494	wrong_fragment
0.002279	num_root
0.002095	num_access_files
0.00205	is_guest_login
0.001849	num_file_creations
0.000905	root_shell
0.000476	num_shells
0.000303	urgent
0	num_outbound_cmds
0	is_host_login
0	land
0	su_attempted

Among 42 attribute we have selected the 11 attributes with highest information gain which is given below in table. List given below in table 3 is the reduced set of the features.

Table 3. Selected 11 attribute with highest Information gain

S No.	Feature name
1	service
2	src_bytes
3	dst_bytes
4	logged_in
5	count
6	dst_host_diff_srv_rate
7	dst_host_srv_diff_host_rate
8	dst_host_count
9	flag
10	dst_host_serror_rate
11	dst_host_srv_serror_rate
12	Class

5.3 Steps of the Proposed Method

Step 1 Select the KDD99 data set with all features
Step 2 Pass the data set to the detection algorithm for training and testing
Step 3 Use 10 fold cross validation for testing of the algorithm
Step 3 Find the result accuracy, TPR and FPR
Step 4 Calculate the information gain of all the features of the KDD99 data set
Step 5 Select the 11 features of the data set with highest information gain
Step 6 Pass the data set with reduced feature to algorithm
Step 7 Repeat step 3 and 4
Step 8 Compare the result when run with all features and with reduced features

The experiments done in light of the information explained in the previous sections of this paper consist of the evaluation of the performance of NB and BayesNet algorithms for the task of classifying novel intrusions. The KDD 99 dataset described in was used in the experiments. Weka , a machine learning toolkit was used for the implementation of the algorithms.

6. RESULT ANALYSIS

6.1 Performance Parameters

In order to analyses and compare the performance of the above mentioned algorithms, metrics like the classification accuracy, learning time, True Positive Rate, False Positive rate were used. These metrics are derived from a basic data structure called as the confusion matrix. A sample confusion matrix for a two-class problem can be represented as:

Table 4: A sample confusion matrix

	Predicted Class Positive	Predicted Class Negative
Actual Class Positive	a	b
Actual Class Negative	c	d

In this confusion matrix, the value a is called a true positive and the value d is called a true negative. The value b is referred to as a false negative and c is known as false positive.

6.1.1 True Positive Rate, False Positive Rate

In the context of intrusion detection, a true positive is an instance which is normal and is also classified as normal by the intrusion detector. For a good IDS TP rate should be high. False positive means no attack but IDS detect the attack. For a good IDS FPR should be low.

6.1.2 Accuracy

This is the most basic measure of the performance of a learning method. This measure determines the percentage of correctly classified instances. From the confusion matrix, we can say that:

$$\text{Accuracy} = \frac{a + d}{a + b + c + d}$$

This metric gives the number of instances from the dataset which are classified correctly i.e. the ratio of true positives and true negatives to the total number of instances.

6.1.3 Precision, Recall and F-Measure:

Precision and recall are terms used widely in the information retrieval domain. They are usually defined as:

Precision = ratio of number of documents retrieved that are relevant to the total number of documents that are retrieved Referring from the confusion matrix, we can define precision and recall for our purposes as

$$\text{Precision} = \frac{a}{a + c}$$

Recall = ratio of number of documents retrieved that are relevant to the total number of documents that are relevant

$$\text{Recall} = \frac{a}{a + b}$$

The precision of a intrusion detection learner would thus indicate the proportion of total number of correctly classified positive instances to the total number of predicted positive instances and recall would indicate the proportion of correctly classified positive instances to the total number of actual positive instances. Thus, a high precision and high recall are desirable for an IDS.

F– Measure Accuracy alone cannot be considered as sole reliable measure for classification. This is because in a case where there are 10 instances, out of which 9 are negative and 1 is positive, if the classifier classifies all of them as negative, the accuracy would be 90%. However, it would result in ignoring all the positive instances. The F-measure is therefore defined as the weighted harmonic mean of precision and recall to address this problem which may be present in any classification scenario.

$$\text{F Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

6.2 Result

The results of the experiments are discussed in this paper. A comparison between NB and BayesNet methods is also made based on the values of the metrics defined in 10-fold cross-validation was used for all the experiments. These results are then interpreted and conclusions are drawn based on this analysis as to which of the methods is best suited to solve the intrusion detection problem.

6.2.1 Result of BayesNet:-

BayesNet was evaluated by taking into account all features of the dataset. The results of this evaluation are summarized in the table below.

Table 5: Result of BayesNet with all attribute

Parameter	Value
Accuracy	96.5624 %
Learning Time	13.02 Sec
Error Rate	3.4376 %
Average True Positive Rate	0.996
Average False Positive Rate	0.038
Average Precision	0.99
Average Recall	0.99
Average F-Measure	0.99

BayesNet was further evaluated on the dataset by taking into account feature reduction using the Information Gain measure. The results of this test are summarized in the following table.

Table 6: Result of BayesNet with selected 11 attribute

Parameter	Value
Accuracy	99.1073 %
Learning Time	3.4 sec
Error Rate	0.8927%
Average True Positive Rate	0.991
Average False Positive Rate	0.001
Average Precision	0.99
Average Recall	0.99
Average F-Measure	0.99

6.2.2 Result of NB

Similar to the BayesNet tests, NB was also evaluated twice; once by taking all attributes into consideration and then by using a reduced attribute subset obtained by the Information Gain measure. The results of these experiments are listed in the following tables

Table 7: Result of NB with all attribute

Parameter	Value
Accuracy	92.73 %
Learning Time	3.21 sec
Error Rate	7.26 %
Average True Positive Rate	0.927
Average False Positive Rate	0.052
Average Precision	0.954
Average Recall	0.927
Average F-Measure	0.939

Table 8: Result of NB with selected 11 attribute

Parameter	Value
Accuracy	92.697 %
Learning Time	0.59 sec
Error Rate	7.303%
Average True Positive Rate	0.927
Average False Positive Rate	0.052
Average Precision	0.953
Average Recall	0.927
Average F-Measure	0.939

6.3 Interpretation of Results:-

R C Staudemeyer [1] have also done the feature reduction to 11 attribute and gave the result. He did an extended series of experiments with the aim to extract a reduced features set with only. He had done a series of experiment after every run reducing and/or adding individual and groups of features.

Now we compare the result of the above mentioned research paper and result obtained from our experiment. Table given below show the result from the R C. Staudemeyer paper.

Table 9: Result from the R C Staudemeyer research paper

Feature Used	Classifier	Accuracy	normal		dos		probe		r 2 l		u 2 r	
			TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
11	BayesNet	91.32 %	0.988	0.089	0.957	0.003	0.804	0.01	0.053	0.002	0.471	0.002
11	NB	77.39 %	0.895	0.076	0.792	0.116	0.72	0.133	0.085	0.003	0.1	0.001

The tables given below show the result from our experiment.

Table 10: Result from the our experiment

Feature Used	Classifier	Accuracy	normal		dos		probe		r 2 l		u 2 r	
			TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
11	BayesNet	99.1073 %	0.991	0.001	0.996	0.001	0.951	0.006	0.824	0.001	0.6	0.001
11	NB	92.697 %	0.936	0.077	0.917	0.015	0.811	0.007	0.824	0.019	0.6	0.007

Now the figure given below show the comparison of the accuracy of NB from above mentioned result

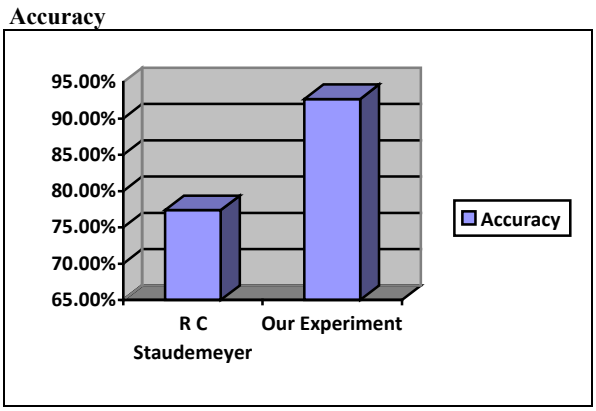

Figure 1: Accuracy comparison of R C Staudemeyer result and our experiment for NB Algorithm

Now the figure given below show the comparison of the accuracy of BayesNet from above mentioned result.

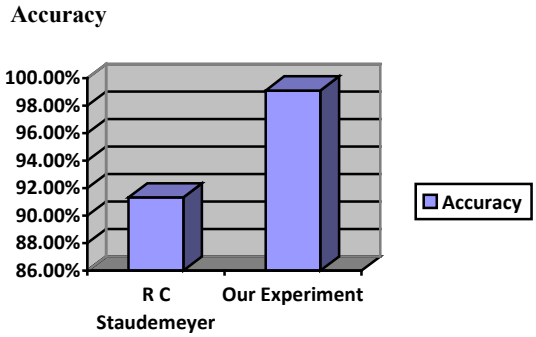

Figure 2: Accuracy comparison of R C Staudemeyer result and Our Experiment for BayesNet

From above figure it is clear that information gain feature reduction method gives the better accuracy which is desirable for good IDS. Especially in the case of BayesNet accuracy is 99.10%.

Now we compare the TPR (Normal) for Algorithm BayesNet and NB

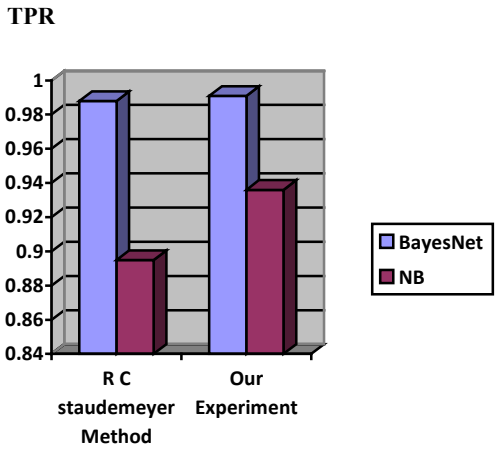

Figure 3: TPR comparison of R C Staudemeyer result and our experiment for BayesNet and NB

For a good IDS TPR should be high. Above figure shows that TPR of the NB and BayesNet algorithm is higher when we reduce the feature of the data set using information gain. Especially in the case of BayesNet TPR is 0.99

Now we compare FPR (Normal) for algorithms BayesNet and NB.

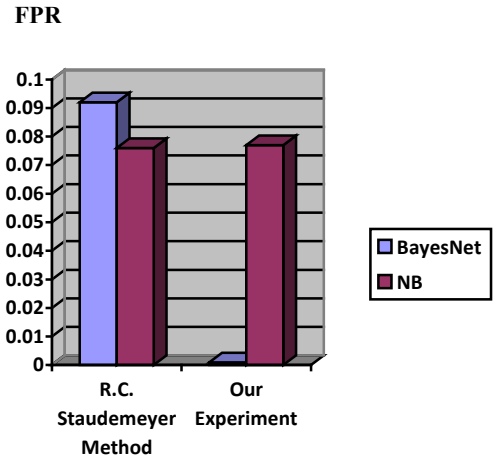

Figure 4: FPR comparison of R C Staudemeyer result and our exp for BayesNet and NB

For a good IDS FPR should be low. Above figure shows that FPR of the BayesNet algorithm is lower when we reduce the feature of the data set using information gain. Especially in the case of BayesNet FPR is 0.001 .In the case of NB algorithm FPR of the algorithm is almost equal in both cases.

From above figures it is clear that Accuracy, TPR and FPR is better in our experiment. So we can say that reduction of the feature using information gain is better technique.

Now we compare the result of the NB and the BayesNet algorithms. Firstly we compare the result after run the algorithm with all attribute. Secondly we compare the result after run the algorithm with reduced 11 attribute than only we conclude that which one algorithm is good best for the intrusion detection.

Comparison of NB and BayesNet

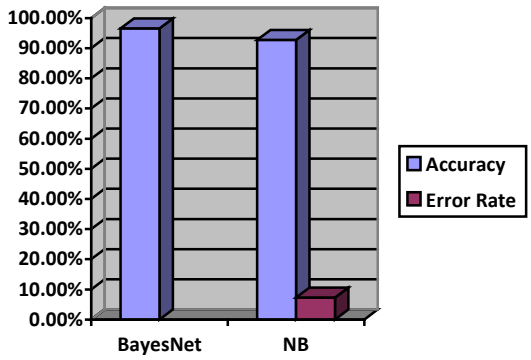

Figure 5: Comparison of accuracy and error rate (all attribute) for NB and BayesNet

Above figure 5 show that BayesNet algorithm has the high accuracy rate i.e. 99.10 % as compare to NB .Error rate is also very low in case of BayesNet algorithm. It shows that BayesNet algorithm is good for the intrusion detection purpose than NB

Now we compare the result of BayesNet and NB after reduction of the features of the data set. The figure given below show the accuracy and error rate of the NB and BayesNet algorithm when features of the data set reduced to 11 using information gain.

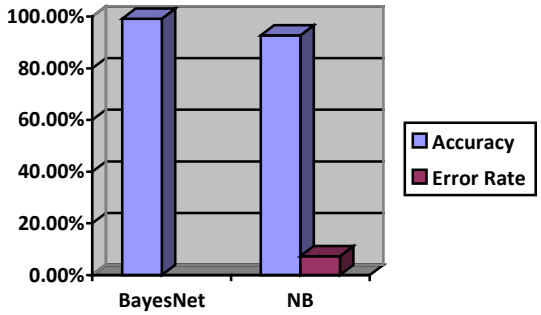

Figure 6: Comparison of accuracy and error rate (selected 11 features) for NB and BayesNet

Above figure shows that accuracy of the BayesNet is better than NB when we run the algorithm after reduction of the features. Error rate is also low in case of BayesNet almost zero than NB

Now we compare the TPR and FPR of the BayesNet and NB algorithm with all attribute and with selected 11 attributes.

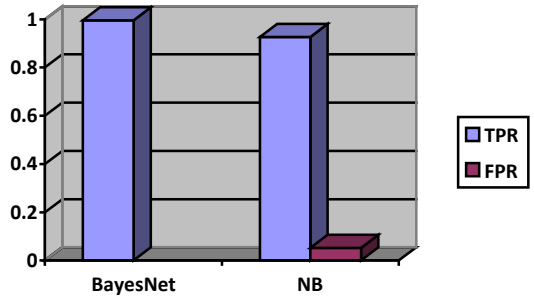

Figure 7: Comparison of TPR and FPR (all attributes) for NB and BayesNet

Figure above shows the TPR and FPR of the BayesNet and NB algorithm when run with the all attributes of the data set. Figure shows that TPR of the BayesNet is higher than NB algorithm which is desirable. Figure also shows that FPR of the BayesNet is almost zero which is desirable for a good intrusion detection algorithm.

Now we compare the same comparison of TPR and FPR when we run the algorithm with selected 11 features.

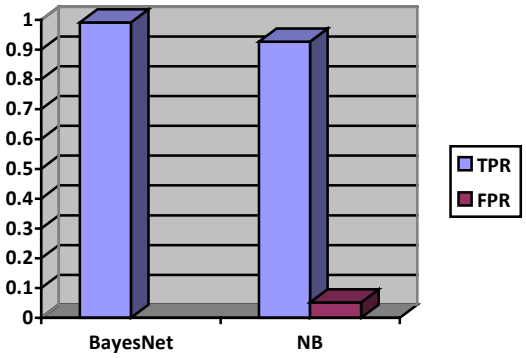

Figure 8: Comparison of TPR and FPR (selected 11 attributes) for NB and BayesNet

Figure above shows the TPR and FPR of the BayesNet and NB algorithm when run with the all attributes of the data set. Figure shows that TPR of the BayesNet is higher than NB algorithm which is desirable. Figures also show that FPR of the BayesNet is almost zero which is desirable for a good intrusion detection algorithm.
Now it is clear that BayesNet gives better result in both cases when we run with all features and with selected features. So we can say that BayesNet is better than NB for the intrusion detection.

7. CONCLUSION

In this paper, we described the applications of machine learning to intrusion detection. We reduced the features of the data set using information gain of the attributes. After reducing the feature of the data set we pass the data set to the algorithm. Result shows that drastically decreased in learning time of the algorithm and increase in accuracy and TPR.We also compare the result with previous research work done by others. Comparison shows that reduction of the feature using information gain technique is suitable for the feature reduction. Using Weka, we analysed two algorithms towards their suitability for detecting intrusions from KDD99 dataset. We showed that machine learning can be effectively applied to detect novel intrusions and focused on anomaly detection. The two learning algorithms, NB and BayesNet were compared at the task of detecting intrusions. BayesNet with an accuracy rate of approximately 99% was found to perform much better at detecting intrusions than NB. Based on the experiments done in the paper and their corresponding results, we can state the following: Machine learning is an effective methodology which can be used in the field of intrusion detection.

1. The inherent nature of machine learning algorithms makes them more suited to the intrusion detection field of information security.
2. It is possible to analyses huge quantities of audit data by using machine learning techniques, which is otherwise an extremely difficult task.
3. Information gain is the suitable technique for the feature reduction
4. BayesNet algorithm is suitable for the intrusion detection with high accuracy rate, high TPR and low FPR which is desirable for good IDS.

8. FUTURE WORK

Future work to this paper can be summarized as follows:
1. In this paper, two learning algorithms were tested and compared. The Weka offers a collection of many other learning schemes, which can be tested and evaluated.
2. The parameters of the machine learning schemes used in this paper were default. It may be possible to further improve the performance of these schemes towards intrusion detection by optimizing these parameters.
3. Owing to the limited processing power, memory available for the experiments conducted and the scope of the paper, a reduced subset of the actual dataset was used. These experiments can be repeated by taking the entire dataset which may further improve the performance of the learner.
4. The attribute selection mechanism used in this paper was based on the Information Gain concept. Also, the top 11 attributes with maximum information gain (plus the class attribute) were used by the algorithms. It is possible to conduct experiments in which a different attribute selection mechanism is used and also, different numbers of attributes are selected to be given as inputs to the algorithms.

9. REFERENCES

[1] R.C. Staudemeyer, Prof. C.W. Omlin, "Feature Set Reduction for Automatic Network Intrusion Detection with Machine Learning", Max-Born-Institute for Nonlinear Optics and Short Pulse Spectroscopy. 2009.

[2] Knowledge Discovery in Databases DARPA archive. Task Description.KDDCUP 1999 DataSet http://www.kdd.ics.uci.edu/databases/kddcup99/task.htm

[3] Pingchuan Ma," Log Analysis-Based Intrusion Detection via Unsupervised Learning" Master of Science, School of Informatics, University of Edinburgh,2003.

[4] James P. Anderson, "Computer security threat monitoring and surveillance," Technical Report 98-17, James P.Anderson Co., Fort Washington, Pennsylvania, USA, April 1980

[5] Dorothy E. Denning, "An intrusion detection Model," IEEE Transaction on Software Engineering", SE-13(2), 1987, pp. 222-232.

[6] M. Bahrololum, E. Salahi and M. Khaleghi, "Machine Learning Techniques for feature Reduction in Intrusion Detection Systems: A Comparison" 2009 Fourth International Conference on Computer Science.

[7] Meng Jianliang, Shang Haikun, "The application on intrusion detection based on K Means cluster algorithm" International Forum on Information Technology and Application, 2009.

[8] Li Tian, Wang Jianwen, "Research on Network Intrusion Detection System Based on Improved K-means Clustering Algorithm" International Forum on Computer Science Technology and Applications, 2009.

[9] Gary Stein, Bing Chen," Decision Tree Classifier for network intrusion detection with GA based feature selection", University of Central Florida. Proceedings of 43rdannual Southeast regional Conference. Volume-2 ,2005.

[10] Rung Ching Chen, Kai Fan Cheng and Chia – Fen Hsieh,"Using rough set and support vector machine for network intrusion detection" International Journal of Network Security and Its Application (IJNSA),Vol 1, No 1, April 2009.

[11] Lin Ni , Hong Ying Zheng " An Unsupervised Intrusion Detection Method Combined Clustering with Chaos Simulated Annealing" Proceeding of the Sixth International on Machine Learning and Cybernetics, Hong Kong, 19-22, August 2007.

[12] Jiong Zhang and Mohhammad Zulkernine," Anomaly based Network Intrusion detection with Unsupervised outlier detection"School of Computing Queen's University, Kingston,Ontario,Canada.IEEE

International Conference ICC 06,Volume-9, 11-15 June 2006.pp 2388-2393.

[13] Cuixiao Zhang, Guobing Zhang, Shanshan Sen, "A mixed unsupervised clustering based Intrusion detection model" Third International Conference on Genetic and Evolutionary Computing, 2009.

[14] Juan Wang, Quren Yang and Dasen Ren , " An intrusion detection algorithm based on decision tree technology" Asia Pacific Conference based on Information Processing,2009.

[15] Yan Luo and Jeffrey J.P. Tsai, "A Framework for Extrusion Detection Using Machine Learning"11th IEEE Symposium on Object Oriented Real-Time Distributed computing (ISORC),2008.

[16] Reza Sadoddin and Ali A. Ghorbani, "A Comparative Study of Unsupervised Machine Learning and Data Mining Techniques for Intrusion Detection" Springer-Verlag Berlin Heidelberg, 2007.

[17] Anshu Veda "Intrusion Detection Using Data mining Techniques" Report IIT Bombay2006.

[18] Khalid Alsubhi,Nizar Bouabdallah,Raouf Boutaba "Performance Analysis is Intrusion Detection and Prevention System"12th IFIP/ IEEE International Symposium on Intergrated Network Management 2011.

[19] Dewan Md. Farid, Nouria Harbi, and Mohammad Zahidur Rahman "Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection," International Journal of Network Security & Its Applications, Vol. 2, No. 2, April2010, pp. 12-25.

[20] Hongwei Gao,Dingju Zhu, Xiaomin Wang "A Parallel Clustering Ensemble algorithm for Intrusion Detection System "2010. Ninth IEEE International Symposium on Distributed Computing and Applications to Business. Engineering and Science.

[21] Shaohua Teng, Hongle Du Wei Zhang, Xiufen Fu "A Cooperative Network Intrusion Detection Based on Heterogeneous Distance Function Clustering"2010 ,14th IEEE International Conference on Computer Supported Cooperative Work in Design.