Hybrid Feature Selection for Modelling Intrusion Detection System and Cyber Attack Detection System

S. Vijayasankari
Assistant Professor
Department of Master of Computer Applications
E.M.G. Yadava Women's College
Madurai – 625 015, India

K. Ramar
Professor of Computer Science Engineering
Principal, Einstein College of Engineering
Sir.C.V.Raman Nagar
Tirunelveli – 627 012, India

ABSTRACT

Intrusion Detection Systems (IDS) and Cyber Attack Detection System (CADS) have to be provided in a Generalised Discriminant Analysis Algorithm. It is an important approach to nonlinear features and extensively used tool for ensuring network security. Complex relationships exist between the features, which are difficult for humans to discover. The conventional Linear Discriminant Analysis feature reduction technique is not suitable for nonlinear data set. Artificial Neural Network and C4.5 classifiers to result in supervisory algorithm are used. If real-time detection is desired IDS must reduce the amount of data to be processed. Currently IDS examine all data features to detect intrusion or misuse patterns. Some of the features may be redundant or contribute little to the detection process. The purpose of this research work is to identify important input features in building IDS that is computationally efficient and effective. The performance of two feature selection algorithms involving Bayesian Networks (BN) and Classification and Regression Trees (CART) and an ensemble of BN and CART were investigated. Empirical results indicate that significant input feature selection is important to design IDS with efficient, effective and lightweight for real world detection systems. Finally, hybrid architecture for combining different feature selection algorithms for real world intrusion detection was proposed.

Keywords

Cyber attack, Data mining, Hybrid feature selection, Intrusion detection, Classification.

1. INTRODUCTION

An Intrusion can be defined as any set of actions that threaten the integrity, confidentiality or availability of a network resource. An Intrusion Detection System (IDS) is device or software applications that monitor network and system activities for malicious activities or policy violations and produces reports to a management station. In IDS large number of data is to be examined for a small network. The analysis become difficult due to the superfluous features and makes it difficult to sense the suspicious behaviour patterns. Hence IDS becomes a significant and extensively used tool for ensuring network security and also this system is essential for protecting network and its resources from illegal penetration. The internet has brought great benefits of the modern civilization. According to the American Computer Emergency Response Team / Coordination Center (CERT) [1] statistics, network cases annually showed index growth in recent years and this attack became a new weapon to a world war [7].

Among the available data set the entire dataset cannot be used for IDS and thus reduction was done by using data filtering, data clustering and feature selection. Thus some of the irrelevant data which may not be useful to the IDS are identified and removed before processing. The amount of audit data that an IDS needs to examine is very large even for a small network and analysis is also a difficult even with computer assistance because extraneous features can make it harder to detect suspicious behaviour patterns [5], [10]. Multifaceted associations exist between the features, which are not easy for humans to discover. Thus IDS must reduce the amount of data to be processed. It is incredibly significant if real-time detection is desired. Clustering can be executed to discover the hidden patterns in data and significant features for use in detection. It can also be used as a reduction system by storing the characteristics of the clusters as an alternative of the actual data. In multifaceted classification domains, features may have false correlations, which delay the process of detecting intrusions. In addition some features may be redundant as the information they add is contained in other features. Additional features can increase computation time, and also have a crash on the accuracy of IDS.

Two data mining techniques such as Markov blanket feature selection and classification and regression trees are applied for feature selection and classification in IDS [12]. Feature selection is an important step in building intrusion detection models [13], [14], [15]. Feature selection get better classification by searching the subset of features, which best classifies the training data. IDS were developed by using many machine learning paradigms, expert systems and fuzzy inference systems. The irrelevant features which are unimportant may be eliminated, without significantly lowering the performance of the IDS. Very little scientific efforts are diverted to model efficient IDS feature selection and are modelled as a classification problem in a machine learning context.

The anomaly detection method was applied for mobile ad hoc networks to detect the intrusions and they created a normal profile under the absence of attacks [11]. The attack profile is created by simulating attacks such as black hole and flooding. After collecting the audit data, it was converted into an appropriate form for the detection process. The size of the audit data is reduced by means of feature selection technique and the genetic algorithm is used for detection. Feature extraction includes feature construction, space dimensionality reduction, sparse representations, and feature selection [4]. All these techniques are used as pre-processing to machine learning and statistics tasks of prediction, pattern recognition and regression.

2. FEATURE SELECTION AND CLASSIFICATION

2.1 Bayesian Learning Modelling of Input Features

The Bayesian Network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a Directed Acyclic Graph (DAG). In DAG whose nodes represents random variable in the Bayesian sense; they may be observable quantities, latent variables, unknown parameters or hypothesis. The prime two tasks of Bayesian Methods are

i. Provides practical learning algorithms
ii. Needs prior probabilities

2.2 Markov Blanket Modelling of Input Features

Markov Blanket of the target variable A, is denoted as MB(A) and it is the key concept for solving the variable selection problem. MB(A) is defined as the set of variable condition on which all other variables are probabilistically independent of ‘A’. Thus the knowledge of the values of the Markov Blanket variable should render all other variables superfluous for classifying the independent variable ‘A’

2.3 Concept learning relation of Bayes theorem

Bayes classifier offers a powerful supervised classification technique. This model assumes all input attributes to be of equal importance and independent of one another. Bayesian Network can be used to compute the conditional probability of one node, given values assigned to the other nodes. A common Bayesian Network classifier learning is that we can get a set of features that are on the Markov blanket of the class node. The Markov blanket of a node ‘k’ is the union of k’s parents, k’s children and the parents of k’s children. This subset of nodes shields k from being affected by any node outside the blanket. While using a Bayesian Network classifier on whole data, the Markov blanket of the class node form feature selection and all features outside the Markov blanket are deleted from the Bayesian Network.

2.4 Input Feature Modelling for CART

Classification And Regression Tree (CART) is one of the popular methods of building decision trees in the machine learning community. CART builds a binary decisions tree by splitting the records at each node, according to a function of a single attribute. It uses the Gini index for determining the best split. The classification and regression tree methodology is technically called as binary recursive partitioning [9]. This process is binary because parent nodes are always split into exactly two child nodes and recursive because the process is repeated by treating each child node as a parent. The key elements of CART analysis are a set of rules for splitting each node in a tree, deciding when tree is complete and assigning a class outcome to each terminal node. Defence American Research Project Agency (DARPA) intrusion data set with 5092 cases and 41 variables, CART considers up to 5092 times 41 splits for a total of 208772 possible splits. In initial splitting it produces two nodes each of which were split in the same manner as the root node. Once again all the input fields are examined to find the candidates splitters. If no split was found it significantly decreases the diversity of the given node and is labelled as a leaf node.

Instead of attempting to decide whether a given node is terminal or not, the algorithm proceeds by growing trees until it is not possible to grow them any further. In CART complex tree is build and then pruning it back to the optimally general tree on the basis of the results of cross-validation or test set validation. The tree is pruned back according to the performance of the various pruned versions of the tree on the test set data. By using cross validation, the tree that is most likely to do well on new unseen data can be chosen. The CART algorithm is relatively robust with respect to missing data. If the value is missing for a particular predictor in a particular record, that record will not be used in determination of the optimal spit when the tree is being built. In effect, CART will utilize as much information as it has on hand in order to make the decision for picking the best possible split. When CART is being used to predict on new data, missing values can be handled via surrogates. Surrogates are split values and predictors that mimic the actual split in the tree and can be used when data for prepared predictor is missing.

3. EXPERIMENT SETUP

3.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a class specific method represents data to make it useful for classification [2], [3], [16]. It locates the optimal transformation matrix so as to preserve most of the information that can be used to discriminate between the different classes. Therefore this analysis requires the data to have appropriate class labels.

3.2 Generalized Discriminant Analysis

The Generalized Discriminant Analysis (GDA) is used for multi-class classification problems. The large variations in the attack patterns of various attack classes, there is usually a considerable overlap between some of these classes in the feature space. In this situation, a feature transformation mechanism that can minimize the between-class scatter is used. GDA is a method designed for nonlinear classification based on a kernel function which transforms the original space to a new high - dimensional feature space [17], [18]. Generalized Discriminant Analysis algorithm is better than the Liner Discriminant Analysis for the case of large scale dataset where the number of training samples is large [20].

The experimental data was prepared by the 1998 DARPA Intrusion Detection Evaluation program by MIT Lincoln Labs [6]. The data was developed from Knowledge Discovery and Data mining (KDD) competition by DARPA and is measured as a standard yardstick for intrusion detection evaluation program [8]. The data set which was developed is having 24 attack types and that could be classified into five categories such as Normal, Probing, Denial of Service (DOS), Remote to User (R2L) and User to Root (U2R). The latest data set having 4,940,000 records with 744 MB and the data set has 41 attributes for each connection record along with one class label. Some derived features which are in the form of nominal or numeric and useful in distinguishing normal connection from attacks. A few host features examine the connections in the past two seconds that have the same destination host as the current connection, and calculate statistics related to protocol behaviour, service, etc. But some service features examine the connections in the past two seconds that have the same service as the current connection. Additional related records were furthermore sorted by destination host, and features were constructed using a window of 100 connections to the same host in place of a time window and these are called as host-based traffic features. R2L and U2R attacks don't have any sequential patterns like DOS and Probing, since the former attacks have the attacks embedded in the data packets whereas the later attacks have many connections in a short amount of time. A few features that look for doubtful performance in the data envelope and these are called content features.

Data reduction, training phase and testing phase are the three phases of the experiment. The important variables for real-time intrusion detection are selected by feature selection in the data reduction phase. In the training phase, Bayesian neural network and classification and regression trees constructs a model using the training data to give maximum generalization accuracy on the unseen data. Then the test data is passed through the saved trained model to detect intrusions in the testing phase. In our experiments 11982 records were generated randomly through 41 features. The features are labelled in the sequence of A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, AA, AB, AC, AD, AF, AG, AH, AI, AJ, AK, AL, AM, AN, AO and the class label is named as AP. Data set has five different classes namely Normal, Probing, Denial of Service (DOS), Remote to User (R2L) and User to Root (U2R). The training and test record includes 5092 and 6890 respectively. All IDS models are trained and tested with the same set of data. As the data set has five different classes and it is classified into 5-class binary classification. The Normal data belongs to class 1, Probe belongs to class 2, DOS belongs to class 3, R2L belongs to class 4 and U2R belongs to class 5. All experiments were performed using an Intel Pentium 4 processor D915, 2.8GHZ, 4M, SL9DA with 512MB RAM.

4. RESULTS AND DISCUSSION

4.1. C4.5 Classifier

Algorithms for constructing decision trees are with the most well known and widely used of all machine learning methods. C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. C4.5 is one such system that learns decision-tree classifiers. Several authors have recently noted that C4.5's performance is weaker in domains with a preponderance of continuous attributes, numeric class label than for learning tasks that have mainly discrete attributes [19]. Classification tree is a prediction mode in machine learning and it is also called as decision tree. Tree pattern graph is similar to flow chart structure and any internal nodes of leaves represent distributed situation of various types. Top down tree construction and bottom up pruning are the two methods for tree construction. In which C4.5 is used for top down tree construction. The detection and identification of attack and non attack behaviours can be generalized as follows. True Positive (TP): a legitimate attack which triggers an IDS to produce an alarm when it is actually normal, True Negative (TN): when no attack has taken place and no alarm is raised, False Positive (FP): an event signalling an IDS to produce an alarm when no attack has taken place, False Negative (FN): a failure of an IDS to detect an actual attack

Confusion matrix contains information of actual and predicted classifications done by classifier. In the performance of such a system is commonly evaluated using the data in a matrix. The following table 1 shows the details of a confusion matrix.

Table 1. Confusion Matrix

Forecasted Actual	Normal	Attack
Normal	True Negative (TN)	False Positive (FP)
Attack	False Negative (FN)	True Positive (TP)

In the confusion matrix, above rows correspond to predicted categories, while columns correspond to actual categories. Comparison of detection rate: Detection Rate (DR) is given by DR= TP/ (TP+FN) * 100%.

Comparison of false alarm rate: False Alarm Rate (FAR) refers to the proportion that normal data is falsely detected as attack behaviour FAR= FP/ (FP+TN) * 100%.

The reported results in terms of detection rate, false alarm rate, training time and testing time of Artificial Neural Network (ANN) and C4.5 decision tree classifiers are summarized for Linear Discriminant Analysis technique and for Generalized Discriminant Analysis technique in the following tables 2 and 3 respectively.

Table 2. Detection Rate, False Alarm Rate, Training Time and Testing Time of ANN and C4.5 Classifier with LDA Technique

Class	ANN				C.4.5			
	DR	FAR	TR	TE	DR	FAR	TR	TE
Normal	95.97	23.13	43s	30s	97.16	33.83	40s	29s
Prob	95.91	55.72	16s	15s	92.14	71.19	16s	16s
DOS	95.91	0.71	55s	27s	87.21	0.14	50s	26s
R2L	10.14	0.16	17s	15s	10.66	1.27	15s	12s
U2R	16.99	11.11	10s	10s	22.33	47.55	10s	9s

The detection rate of Linear Discriminant Analysis technique and Generalized Discriminant Analysis technique for Artificial Neural Network and C4.5 are compared in the figure 1 and figure 2. The false alarm rates of LDA and GDA for ANN and C4.5 decision tree algorithm are compared in the following figure 3 and figure 4 respectively. The training time of LDA and GDA for ANN and C4.5 decision tree algorithm are compared in the figure 5 and figure 6 respectively. Similarly the Testing Time of LDA and GDA for ANN and C4.5 decision tree algorithm are compared in the figures 7 and 8 respectively.

Table 3. Detection Rate, False Alarm Rate, Training Time and Testing Time of ANN and C4.5 Classifier with GDA Technique

Class	ANN				C4.5			
	DR	FAR	TR	TE	DR	FAR	TR	TE
Normal	99.8	30.01	39s	25s	99.78	9.88	32s	23s
Prob	97.19	75.9	15s	13s	99.61	24.14	13s	11s
DOS	98.11	0.29	48s	24s	98.01	0.46	45s	22s
R2L	12.11	0.38	14s	11s	67.02	0.05	12s	9s
U2R	24.09	22.99	10s	8s	57.01	5.7	7s	6s

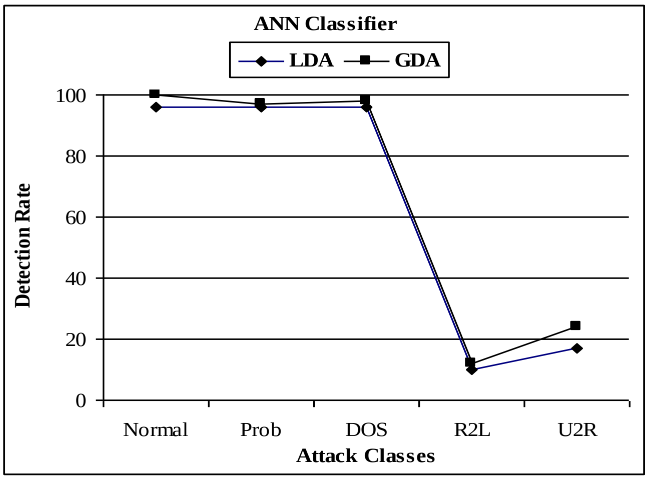

Fig 1: Comparison of detection of LDA and GDA for ANN

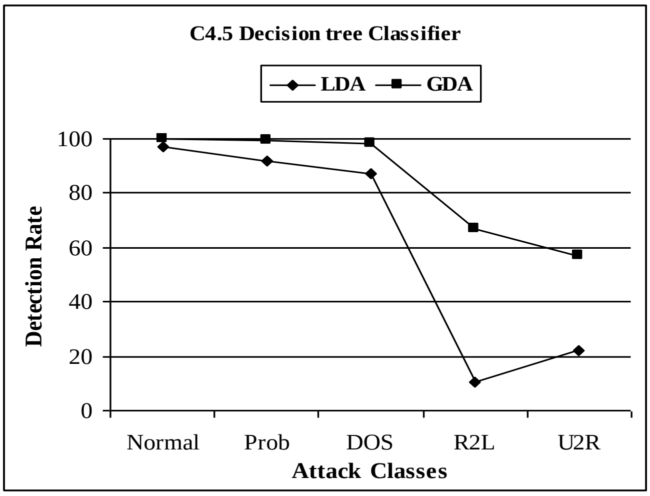

Fig 2: Comparison of detection of LDA and GDA for C4.5

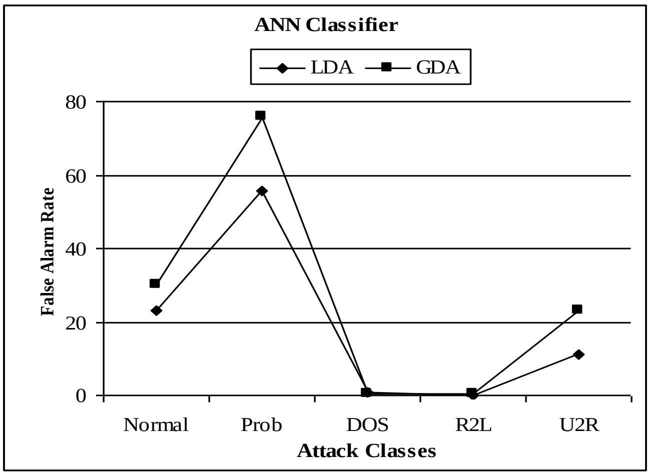

Fig 3: Comparison of False Alarm Rate of LDA and GDA for ANN

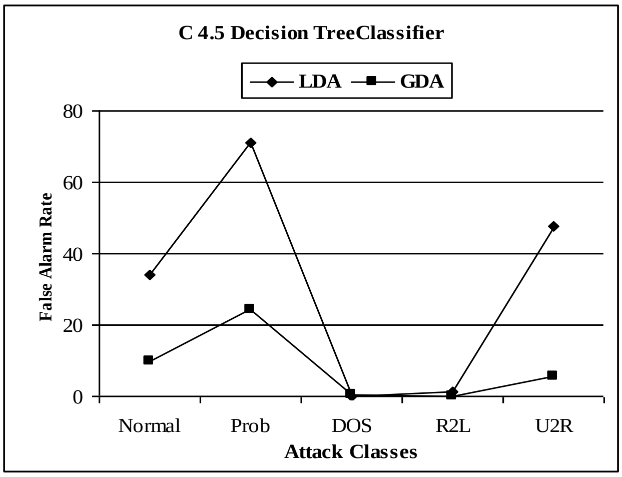

Fig 4: Comparison of False Alarm Rate of LDA and GDA for C4.5

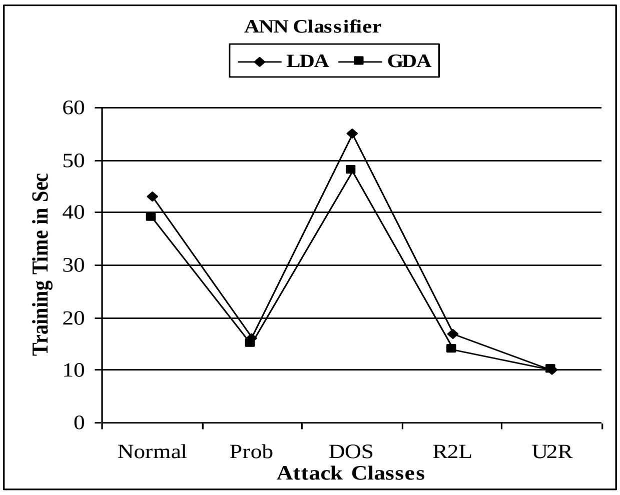

Fig 5: Comparison of Training Time of LDA and GDA for ANN

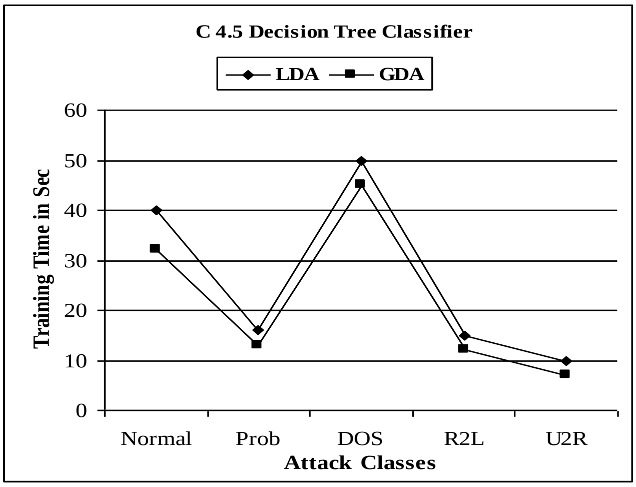

Fig 6: Comparison of Training Time of LDA and GDA for C4.5

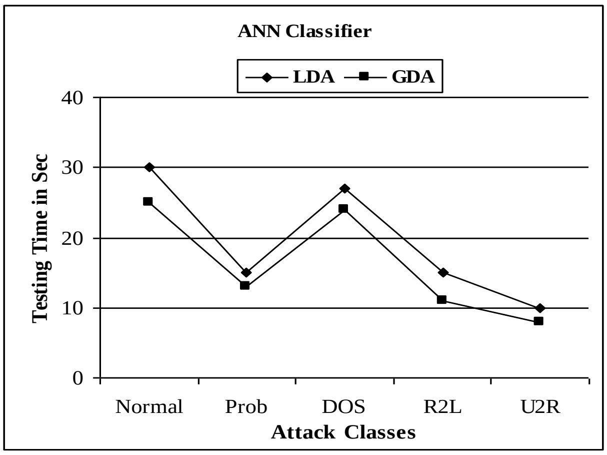

Fig 7: Comparison of Testing Time of LDA and GDA for ANN

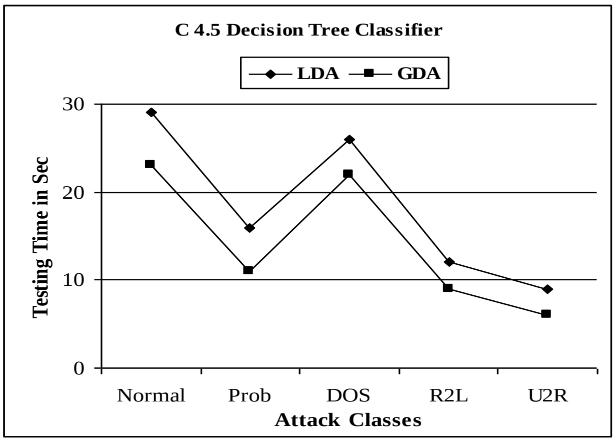

Fig 8: Comparison of Testing Time of LDA and GDA for C4.5

4.2. IDS Modelling via Bayesian Network

Markov blanket modelling is used to retrieve the most significant features and found out those 17 variables of the data set of the class node. These 17 variables are A, B, C, E, F, H, K, M, N, Q, V, W, X, Y, Z, AC and AE. In addition Bayesian network classifier is created using the training data and then the classifier is used on the test data set to classify the data as an attack or normal. The following table 4 illustrates the performance of Bayesian belief network by using the original 41 variable data set and the 17 variables reduced data set. The training and testing times for each classifier are decreased when 17 variable data set is used. By using this 17 variable data set there is a slight increase in the performance accuracy for Normal class compared with the 41 variable data set.

Table 4. Performance of Bayesian Belief Network 41 Variables Attack with 17 Variables

Class	Train (sec)	Test (sec)	Accuracy (%)	Train (sec)	Test (sec)	Accuracy (%)
Normal	42.14	19.02	99.57	23.29	11.16	99.64
Probe	49.15	21.04	99.43	25.07	13.04	98.57
DOS	54.52	23.02	99.69	28.49	14.14	98.16
U2R	30.02	15.23	64.00	14.13	7.49	60.00
R2L	47.28	12.11	99.11	21.13	13.57	98.93

4.3. IDS Modelling via CART

Depending upon the contribution for the construction of the decision tree the important variables are identified and ranking to the variables were generated in terms of percentages. The variables that have 0.00% rankings are eliminated and considered only the primary splitters or surrogates. This resulted in a reduced 12 variable data set with C, E, F, L, W, X, Y, AB, AE, AF, AG and AI as variables. Further the classifier is constructed using the training data and then the test data is passed through the saved trained model. Table 5 compares the performance of CART using the 41 variable original data set and the 12 variable reduced data set. Normal class is classified 100 percent correctly. Furthermore, the accuracies of classes U2R and R2L have increased by using the 12 variable reduced data set. It is also found that CART could classify accurately on smaller data sets. Further, we used the Bayesian reduced 17 variable data set to train CART and the CART reduced 12 variable dataset to train Bayesian network. As illustrated in Table 6 except R2L all other classes were classified well by the CART algorithm. Moreover, training and testing time for each class are greater for Bayesian network classifier compared to CART algorithm.

Table 5. Performance of Classification And Regression Trees Using 41 Variable Dataset With 12 Variable Dataset Attack

Class	Train (sec)	Test (sec)	Accuracy (%)	Train (sec)	Test (sec)	Accuracy (%)
Normal	1.15	0.18	99.64	0.80	0.02	**100.00**
Probe	1.25	0.03	97.85	0.85	0.05	**97.71**
DOS	2.32	0.05	99.47	0.97	0.07	**85.34**
U2R	1.10	0.02	48.00	0.45	0.03	**64.00**
R2L	1.56	0.03	90.58	0.79	0.02	**95.56**

Table 6. Performance of Bayesian and Cart with Reduced Dataset Using Bayesian with 12 Variables Cart Attack With 17 Variables

Class	Train (sec)	Test (sec)	Accuracy (%)	Train (sec)	Test (sec)	Accuracy (%)
Normal	20.10	10.13	98.78	1.03	0.04	**99.64**
Probe	23.15	11.17	99.57	1.15	0.13	**100.00**
DOS	25.19	12.10	98.95	0.96	0.11	**99.97**
U2R	11.03	5.01	48.00	0.59	0.02	**72.00**
R2L	19.05	12.13	98.93	0.93	0.10	**96.62**

4.4 Feature Ranking via Support Vector Machines

A novel approach has been done to evaluate the performance of CART and Bayesian network using the reduced dataset. Table 7 shows the performance comparisons of CART and Bayesian network using 19 variables. Except U2R the 17 and 12 variable dataset performs well for all the other classes.

Table 7. Performance of Cart and Bayesian Network Using 19 Variables

Class	Bayesian	CART
Normal	99.57	95.50
Probe	96.71	96.85
DOS	99.02	94.31
U2R	56.00	84.00
R2L	97.87	97.69

4.5. Ensemble Approach via Reduced Data Set

The Bayesian network classifier and CART models were constructed individually to obtain a high - quality generalization performance. The ensemble approach is used for 12, 17 and 41 variable dataset and the final outputs were decided as follows:

- Depending on the generalization accuracy each classifier's output is given a weight in the scale in the range of 0 to 1.
- If both classifiers agree then the output is determined accordingly.
- If there is a difference then the decision given by the classifier with the maximum weight age is accounted.

Ensemble approach for IDS consists of reduction, ensemble based intrusion, detection system, Bayesian network and classification, and regression trees. The developed IDS model for the different attack classes are shown in the table 8.

Table 8. Developed IDS Model for Different Attack Classes

Sl. No	Class	Selection
1	Normal	CART
2	R2L	Ensemble
3	DOS	CART
4	Probe	Ensemble
5	U2R	CART

After summarizing all the empirical results the hybrid IDS model was developed. By using this hybrid model Normal, Probe and DOS could be detected with 100% accuracy and U2R and R2L with 80% and 99.47% accuracies respectively. Performances of ensemble approach using different data sets are shown in table 9. From the results it can be concluded that ensemble approach gives better performance than the two individual separately used models. The ensemble approach basically exploits the differences in misclassification (by individual models) and improves the overall performance.

Table 9. Performance of Ensemble Approach using Different Datasets Class 12 Variables Vs 17 Variables Vs 41 Variables

Class	12 Variables	17 Variables	41 Variables
Normal	**100.00**	99.64	99.71
Probe	99.86	**100.00**	99.85
DOS	99.98	**100.00**	99.93
U2R	**80.00**	72.00	72.00
R2L	**99.47**	99.29	99.47

5. CONCLUSION

Data reduction is analysed in Intrusion Detection Systems and Cyber Attack Detection System with Generalised Discriminant Analysis technique to overcome the limitations of Linear Discriminant Analysis technique. In this research the new techniques for intrusion detection and Cyber attack performed data reduction was investigated and evaluated their performance on the benchmark intrusion data. The initial experiments compress data was not successful. The feature selection method using Markov blanket model and decision tree analysis defeat this. Following this, we explored general Bayesian Network classifier and Classification and Regression Trees as intrusion detection models. The performance comparisons using different reduced data sets were also demonstrated. The proposed ensemble of BN and CART combines the complementary features of the base classifiers. Finally, a hybrid architecture involving ensemble and base classifiers for intrusion detection was proposed. From the empirical results, it is evident by using the hybrid model Normal, Probe and DOS could be detected with 100% accuracy and U2R and R2L with 80% and 99.47% accuracies respectively. Our future research will be directed towards developing more accurate base classifiers particularly for the detection of U2R type of attacks. Thus an architecture for real world IDS and CADS is analysed.

6. REFERENCES

[1] American computer Emergency response Team / Coordination centre (CERT), http://www.cert.org., January, 2012.

[2] Kemal Polat, Salih Güneş, and Ahmet Arslan 2008, 'A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine', Expert Systems with Applications, Vol. 34, pp.482-487.

[3] Jing Gao, Haibin Cheng and Pang-ning Tan, 2006 'A novel framework for Incorporating Labeled Example into anomaly detection', Proceedings of the Siam Conference on Data mining

[4] Gopi K, Kuchimanchi, Vir V Phoha, Kiran S Balagani, Shekhar R Gaddam 2004, 'Dimension Reduction Using Feature Extraction Methods for real-time Misuse Detection Systems', Proceedings of the IEEE on Information.

[5] Mukkamala S, Sung A.H. and Abraham A. 2003, Intrusion Detection Using Ensemble of Soft Computing Paradigms, Third International Conference on Intelligent Systems Design and Applications, Springer Verlag Germany, pp. 239-248.

[6] MIT Lincoln Laboratory. http://www.ll.mit.edu/IST/ideval/

[7] Information security report, http://www.isecu-tech.com.tw/.2012.

[8] KDDCup99dataset, August 2003 http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[9] Brieman L., Friedman J., Olshen R. and Stone C. 1984, Classification of Regression Trees. Wadsworth Inc.

[10] Lee W., Stolfo S. and Mok K. 1999, A Data Mining Framework for Building Intrusion Detection Models, In Proceedings of the IEEE Symposium on Security and Privacy.

[11] Nallusamy. R, Jayarajan. K, Duraiswamy. K. 2009, 'Intrusion Detection In Mobile Ad Hoc Networks Using GA Based Feature Selection', Georgian Electronic Scientific Journal: Computer Science and Telecommunications, No.5 (22), pp. 28-35.

[12] Chebrolu, S., A. Abraham, J. Thomas. 2005 , 'Feature deduction and ensemble design of intrusion detection systems', Journal of Computers & Security, Vol-24, pp.295-307.

[13] Khoshgoftaar T.M, Nath S.V, Zhong S, and Seliya. N. 2005, "Intrusion Detection in Wireless Networks Using Clustering Techniques with Expert Analysis," Proc. Fourth Int'l Conf. Machine Learning and Applications.

[14] A. Boukerche, K.R.L. Juc, J.B. Sobral, and M.S.M.A. Notare. 2004, "An Artificial Immune Based Intrusion Detection Model for Computer and Telecommunication Systems," Parallel Computing, Vol. 30, pp. 629-646.

[15] Y.-H. Liu, D.-X. Tian, and D. Wei. 2006, 'A Wireless Intrusion Detection Method Based on Neural Network', Proc. Second IASTED Int'l Conf. Advances in Computer Science and Technology.

[16] W. Zhao, R. Chellappa, and N. Nandhakumar. 1998, "Empirical Performance Analysis of Linear Discriminant Classifiers," Proc. Computer Vision and Pattern Recognition, pp. 164-169.

[17] Baudt. G and Anouar. F. 2000, 'Generalized Discriminant Analysis Using a Kernal Approach,' Neural Computation.

[18] Fukunaga. K. 1990, 'Introduction to Statistical Pattern Classification', Academic Press, San Diego, California, USA.

[19] Quinlan, J. R. 1994, 'C4.5: Programs for Machine Learning', Machine learning, Vol.16, pp. 235-240

[20] Shailendra Singh and Sanjay Silakari. 2009, 'Generalized Discriminant Analysis algorithm for feature reduction in Cyber Attack Detection System', International Journal of Computer Science and Information Security, Vol.6, pp. 173-180.

7. AUTHORS PROFILE

S. Vijayasankari, M.C.A., M.Phil., has completed her Graduation in Computer Science and Post Graduation in Master of Computer Applications from Madurai Kamaraj University, Madurai, and also completed her Master of Philosophy degree in Computer Science from the same University. She is currently working as an Assistant Professor in M.C.A department at E M G Yadava Women's College, Madurai- 625 014, India. Her area of interest is Data Mining.

Dr. K. Ramar, M.E. Ph.D., FIE, has completed his Bachelor of Engineering degree from Government College of Engineering, Tirunelveli, Madurai Kamaraj University, Post Graduation from P.S.G. College of Technology, Bharathiar University, Coimbatore and Doctorate degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli. He is a life member in CSI - Mumbai, SSI - Tiruvananthapuram, ISTE -NewDelhi and also fellow member in Institution of Engineers, Kolkatta. He is having a credit of service as Professor in the Computer Science and Engineering for the past 25 years and guided many Ph.D, M.E. and M.Phil Dissertations. Currently he is working as a Principal at Einstein College of Engineering, Sir.C.V.Raman Nagar, Tirunelveli - 627 012, India, His area of interest are soft computing, Image Processing, Data Mining and Computer networking