Intrusion Detection System Using Improved Ensemble Technique with DAG Approach

*Deepak Kumar Singh, Beerendra Kumar

Abstract- The Internet, computer networks and information are vital resources of current information trend and their protection has increased in importance in current existence. The intrusion detection system (IDS) plays a vital role to monitors vulnerabilities in network and generates alerts when found attacks. Today the educational network services increasing day today so that IDS becomes essential for security on internet. This paper proposes efficient intrusion detection architecture which named "IDS using improved ensemble techniques with DAG approach" (IDSIETD). The IDSIET contains a new improved algorithm of attribute reduction which combines rough set theory and a method of establishing multiple rough classifications and a process of identifying intrusion data. The concept of pair wise support vector machine (PSVM) has been improved by Direct Acyclic Graph (DAG) approach. The experimental results illustrate the effectiveness of proposed architecture.

Our proposed work is implemented in MATLAB. For implementation purpose write various function and script file for implementation of our proposed architecture. For the test of our hybrid method, we used DARPA KDDCUP99 dataset. This data set is basically set of network intrusion and host intrusion data. This data provided by UCI machine learning website.

Proposed method compare with exiting ensemble techniques like simple ensemble technique and Hybrid ensemble Techniques and generate the improved ensemble technique to getting better result such as detection rate, precision and recall Rate.

Keywords- IDS, IDSIETD, Neural Network, rough set theory, Network Security, MATALAB, KDDCUP99 Dataset, PSVM, DAG.

1. INTRODUCTION

Intrusion data classification and detection process is very complex process in network security. In current network security scenario various types of Intrusion attack are available some are known attack and some are unknown attack. The attack of know Intrusion detection used some well know technique such as signature based technique and rule based technique. In case of unknown Intrusion attack of attack detection is various challenging task. In current trend of Intrusion detection used some data mining technique such as classification and clustering. The process of classification improves the process of detection of Intrusion. In this dissertation used graph based technique for Intrusion classification and detection. The continuity of chapter discusses feature extraction process of Intrusion data, directed acyclic graph technique, support vector machine and proposed methodology.

1.2 FEATURES EXTRACTION:

Intrusion classification can either have single variable approach or a multi-variable approach to detect Intrusion depending on the algorithm used. In the single variable approach a single variable of the system is analyzed. This can be, for example, port number, CPU usage of a local machine etc. In multi-variable approach a combination of several features and their inter-correlations are analyzed. In addition based on the method the way in which features are chosen for the IDS can be divided into two groups; into feature selection and feature reduction.

1.3 FEATURE SELECTION:

In the feature selection method the features are either picked manually from the data monitored or by using a specific feature selection tool. The most suitable features are selected by handpicking from the feature spectrum based on the prior knowledge about the environment that the IDS are monitoring. For example features that can distinguish certain type of traffic from the traffic flows are picked for the network traffic model training. The idea behind the feature selection tools is to reduce the amount of features into a feasible subset of features that do not correlate with each other. Examples of feature selection tools are Bayesian networks (BN) and classification and regression tree (CART). Feature selection process is illustrated in Fig. on the left there are the features (F0…FN) that are available from the data monitored, which is, for example, from network traffic. On the right side is the output (F0...FM) of the selection tool. The number of features in the output varies based on the selection tool used and the inter-correlation of features in the input. Following the basic principles of feature analysis the number of features in the

- *Deepak Kumar Singh is currently pursuing masters degree program in Computer Science Engineering, CSE College,Jhansi, India, E-mail: deepakk005@gmail.com*
- *Beerendra Kumar is Assistant Professor in CSE College, Jhansi, India*

output (M in Fig.) is in most of the cases less than the number of features in the input (N in Fig.). However, it is possible that the output is equal to the input.

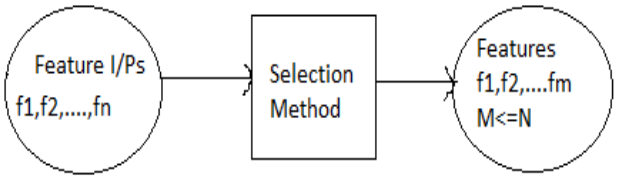

Fig. Feature selection process in feature variable

1.4 SUPPORT VECTOR MACHINE:

Support Vector Machine (SVM) is a novel machine learning method based on statistical learning theory developed by V.N.Vapnik, and it has been successfully applied to numerous classification and pattern recognition problems such as text categorization, image recognition and bioinformatics. It is still in the development stage now. SVM can be used for pattern recognition, regression analysis and principal component analysis. The achievements of SVM in training have Platt's the sequential minimal optimization method, Osuna's the method of Chunking, Joachims' SVM light method and so on. These methods are directed at the training process, and not related to classification process. In the process of SVM training, all the samples are used. So it has no effect on the speed of the classification. Lee and others propose a method of reduction SVM training time and adding the speed of training, reduced support vector machines.

At the same time, because of the reduction of the support vector quantity, the speed of classification is improved to some degree. However, due to the loss of some support vector classification, precision has declined, especially when the number of support vector is so many that the accuracy of its classification will decline. The concept of SVM is to transform the input vectors to a higher dimensional space Z by a nonlinear transform, and then an optical hyper-plane which separates the data can be found. This hyper-plane should have the best generalization capability. As shown in Fig. 2, the black dots and the white dots are the training dataset which belong to two classes. The Plane H series are the hyper-planes to separate the two classes. The optical plane H is found by maximizing the margin value $2/\|w\|$. Hyper-planes H_1 and H_2 are the planes on the border of each class and also parallel to the optical hyper-plane H. The data located on H_1 and H_2 are called support vectors.

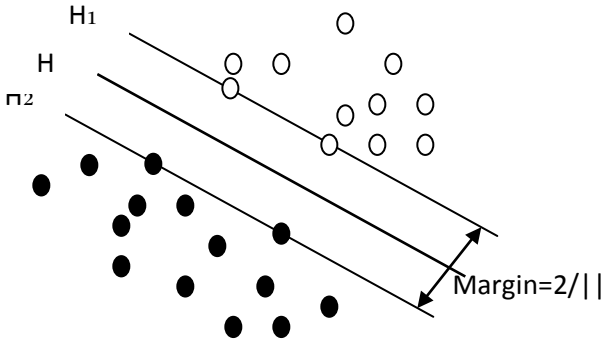

Fig. 2 The SVM Binary Classifications

For training data set $(x_1, y_1),...,(x_l, y_l), y_i \in \{-1,1\}$, to find the optical hyper-plane H, a nonlinear transform, $Z = \Phi(x)$, is applied to x, to make x become linearly dividable. A weight w and offset b satisfying the following criteria will be found:

$$\begin{cases} w^T z_i + b \geq 1, & y_i = 1 \\ w^T z_i + b \leq -1, & y_i = -1 \end{cases} \quad (1)$$

i.e.

$$y_i(w^T z_i + b) \geq 1, \quad i = 1,2,...,l \quad (2)$$

Assume that the equation of the optical hyper plane H (Fig.4.1) is $w_0^T z + b_0 = 0$, then the distance of the data point in any of the two classes to the hyper plane is:

$$\rho(w,b) = \min_{x|y=1} \frac{z^T w}{\|w\|} - \max_{x|y=-1} \frac{z^T w}{\|w\|} \quad (3)$$

A w_0 is to be found to maximize

$$\rho(w_0, b_0) = 2/\|w_0\| = 2/\sqrt{w_0^T w_0} \quad (4)$$

Then the search of the optimal plane H turns to a problem of a second order planning problem.

$$\min_{w,b} \Phi(w) = \frac{1}{2}(w^T w) \quad (5)$$

Subject to $y_i(w^T z_i + b) \geq 1, \quad i = 1,2,...,l$ (6)

If the sample data is not linearly dividable, find the minimum value of

$$\Phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \quad (7)$$

Whereas ξ can be understood as the error of the classification and C is the penalty parameter for this term. By using Lagrange method, the decision function of

$$w_0 = \sum_{i=1}^{l} \lambda_i y_i z_i \quad (8)$$

will be

$$f = \text{sgn}[\sum_{i=0}^{l} \lambda_i y_i (z^T z_i) + b] \quad (9)$$

From the functional theory, a non-negative symmetrical function $K(u,v)$ uniquely define a Hilbert space H, K is the rebuild kernel in the space H:

$$K(u,v) = \sum_i \alpha \varphi_i(u) \varphi_i(v) \quad (10)$$

This stands for an internal product of a characteristic space:

$$z_i{}^T z = \Phi(x_i)^T \Phi(x) = K(x_i, x) \quad (11)$$

Then the decision function can be written as:

$$f = \text{sgn}[\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b] \quad (12)$$

The development of a SVM image classification model depends on the selection of kernel function K. There are many kernels which can be used in modeling of Support Vector Machines. These models consists linear, polynomial, radial basis function (RBF) and sigmoid function:

$$K(x_i, x_j) = \begin{cases} x_i{}^T x_j & Linear \\ (\gamma x_i{}^T x_j + coefficient)^{\deg ree} & Polynomial \\ \exp(-\gamma \mid x_i - x_j \mid^2 & RBF \\ \tanh(\gamma x_i{}^T x_j + coefficient) & Sigmoid \end{cases}$$

Currently the RBF is the most popular used kernel in Support Vector Machines because of their localities and finiteness responses over the whole range of the real x-axis. Improper kernel function might generate poor performance. Currently there is no effective "learning" method to choose a proper kernel function for a specific problem. The selection is decided by the experiment result at this time. In our proposed system, two kernel functions are tested: Radial Basis Function-RBF and Polynomial Function.

$$K_{poly}(x_1, x_2) = (x_1 * x_2 + 1)^p \quad (13)$$

$$K_{RBF}(x_1, x_2) = \exp(-p \parallel x_1 - x_2 \parallel^2)$$

Due to its better performance, RBF was chosen as the kernel function in the model.

2 TRAINING OF DAG:

A DAG is a graph based multi-classification technique in this technique pair-wise SVMs used, let the decision function for class i against class j, with the maximal margin, be:

$$Dij(x) = wijT\emptyset(x) + bij \quad (14)$$

Where Wij is the d-dimensinal vector, $\emptyset(x)$ is a mapping function that maps x into the d-dimensional feature space bij is the bias term and Dij(x)= --Dji(x) The regions Ri are shown in Fig. 4.4 .1 with labels of class I, II and III.

$$Ri = \{x|Dij(x) > 0, j = 1,2, \ldots\ldots n, j \neq i\} \quad (15)$$

If x is in Ri, we classify x into class i. if x is not in Ri(i=1,2,……,n), x is classified by voting. Namely ,for the input vector x,Di(x) is calculate at follow:

$$Di(x) = \sum_{i \neq j, j=1}^{n} \text{sign}(Dij(x))$$

Where $\text{sign}(x) = \begin{cases} 1 \text{ for } x \geq 0, \\ -1 \text{ for } x < 0, \end{cases}$

And x is classified into class

$$\arg \max_{i = 1,2, \ldots, n} Di(x)$$

If x∈ Ri.Di(x) =n -1 and Dk(x)<n-1 for k≠I. thus x is classified into I, but if any of Di(x) is not n-1,may be satisfied for plural is. In this case x is unclassified.

In the shaded region in Fig. 3, Di(x) =0 (i=1, 2 and 3). Therefore, this region is unclassified, although the unclassified region is much smaller than that for the one-against-all support vector machine.

In pair wise SVMs, classification reduces the unclassifiable regions that occur for one-against-all support vector machines but it still exists. To resolve this problem, Vapnik proposed to use continuous decision functions. Namely, we classify a data into the class with maximum value of the decision functions. Inoue and Abe proposed fuzzy support vector machines, in which Membership functions are defined using the decision functions. Another popular solution is DAG SVM that uses a decision tree in the testing stage. Training of a DAG is the same as conventional pair wise SVMs.

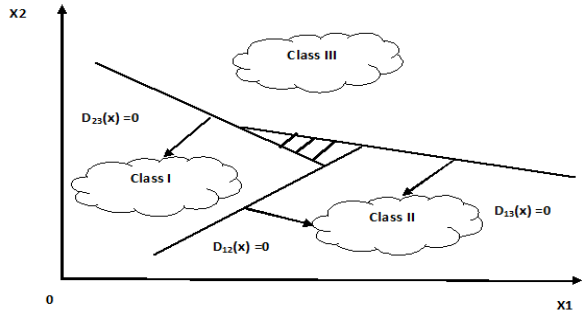

Fig. 3: Unclassifiable Regions By The Pair-Wise Formulation.

Classification by DAGs is faster than by conventional pair-wise SVMs or pair-wise fuzzy SVMs. Fig. 4 shows the decision tree for the three classes shown in Fig. 5. In the Fig 4 show that x does not belong to class i.

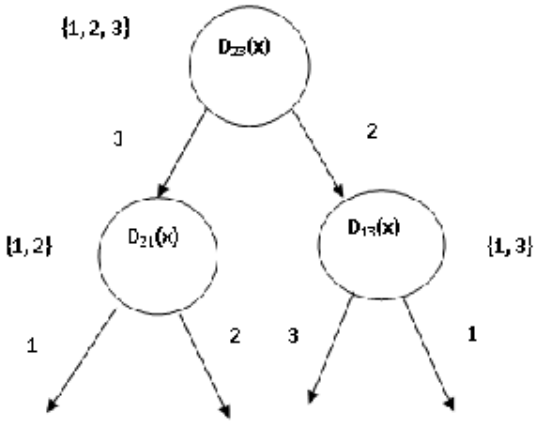

Fig. 4: DAG Classification.

As the top-level classification, we can choose any pair of classes. And except for the leaf node if Dij(x)>0, we consider that x does not belong to class j, and if Dij(x)<0 not class i.

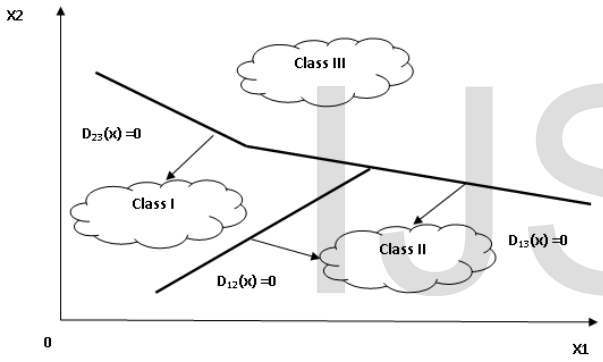

Fig. 5: Generalization Region By DAG Classification.

Thus, if D12(x)> 0, x does not belong to class II. Therefore, it belongs to either class I or class III, and the next classification pair is classes I and III. The generalization regions become as shown in Fig. 5. Unclassifiable regions are resolved, but clearly the generalization regions depend on the tree formation.

3 PROPOSED METHOD:

In this section discuss the graph based ensemble based intrusion detection technique. The graph based technique basically work with collection of features of attribute of network data. The network traffic data passes through the graph set, the set collect the similar type of attribute and discard the dissimilar attribute. The discarded attribute used as feature of collection attribute for ensemble process. The ensemble point fist select the all feature attribute for selection and pass through the classification. The process of algorithm discuss in two phase first phase discuss the collection of attribute and second phase discuss the ensemble of attribute for classification.

Phase-I:

Step1: Initially input Intrusion data passes through preprocessing function and extracted feature part of Intrusion data in form of traffic type.

Step2: the extracted traffic feature data converted into feature vector.

Step 3: In phase of feature mapping in feature space of DAG create a fixed class according to the group of data.

Step 4: steps of processing of DAG.

Initialize Gaussian hyper plane margin.

Choose a random vector from training data and present it to the DAG.

The weight of the plane support vector is estimated. The size of the vector decreases with each iteration.

Each vector in the SV's neighborhood has its weights adjusted to become more like the SV. Vector closest to the SV are altered more than the vector furthest away in the neighborhood.

Repeat from step 2 for enough iteration for convergence.

Calculating the SV is done according to the Euclidean distance among the node's weights (W1, W2, … , Wn) and the input vector's values (V1, V2, … , Vn).

The new weight for a node is the old weight, plus a fraction (L) of the difference between the old weight and the input vector… adjusted (theta) based on distance from the SV.

Phase-II:

Input: N_list: collection of intrusion attributes

Output: N_type: number of classified class

G = (V, E) ←empty //define the feature data in graph mode

1: NP_list←K-means(N_list, K_v) //grouping of data
2: for h ∈ N P_list do
3: h.nn ← Nearest-neighbor (NP_list - {h})
4: h.sc←Compute-SC(h,h.nn) //Reduction of attribute
5: V←V $\cup$ {h} //commutate number of attribute
6: V←V $\cup$ {h.nn}
7: if h.sc < th_{sc} then //check class group
8: E←E $\cup$ {(h,h.nn)} //add this DAG
9: endif
10: end for
11: for each pair of components (g1,g2) ∈ G do
12: μ_1←mean-dist (g1), μ_2←mean-dist (g2)
13: If $\frac{\mu_1+\mu_2}{2*centroid\_dist(g1,g2)}$ > 1 then g1←Merge (g1, g2)
14: end for // Now allot the class labels
15: N_type ← empty
16: for x ∈ N list do
17: h ← PseudopointOf (x)
18: N_type ← N_type $\cup$ {((x) , h.ccomponent)}

19: end for

Step 5: After processing of support vector finally Intrusion data are classified

4 PROPOSED MODEL:

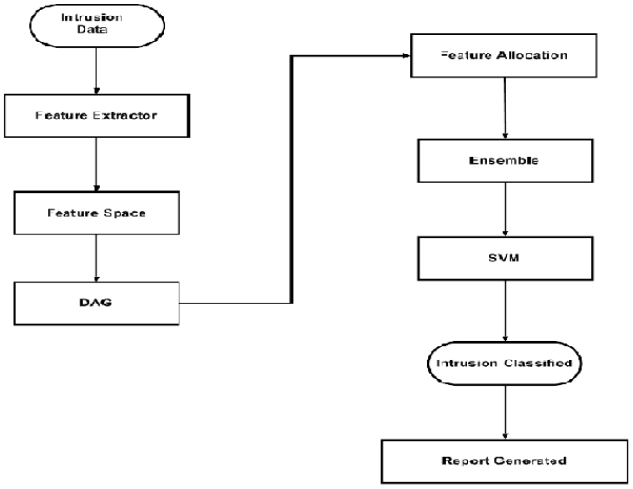

Fig. 7: Flow Diagram of A Proposed Model

5 IMPLEMENTATION & RESULT ANALYSIS:

In this dissertation we perform experimental process of proposed improved ensemble for intrusion detection system. The proposed method implements in matlab and tested with very reputed data set from UCI machine learning research centre. In the research work, I have measured detection accuracy, true positive rate, false positive rate, true negative rate and finally false negative rate error of classification ensemble method. To evaluate these performance parameters I have used KDDCUP99 datasets from UCI machine learning repository namely intrusion detection dataset.

For the purpose of implementation and performance evaluation i implement the simple Ensemble technique and the Hybrid Ensemble technique and we implement a new ensemble algorithm named improved ensemble technique and we compare third algorithm with previous two and the result shown in tables 1.

Method	GV	Parameters Attacks Type	TPR	TNR	FPR	FNR	DR	PR	RR
ENSEMBLE	0.5	NORMAL	6.032	2.462	3.327	2.450	91.55	83.69	82.69
		DOS	6.123	1.564	2.354	1.450	90.55	82.69	80.69
		PROB	6.243	3.462	1.567	2.110	88.55	81.87	83.67
		U2R	7.032	1.351	0.327	1.450	87.55	84.47	83.69
		R2L	5.232	3.351	2.327	1.590	88.55	86.19	81.68
HYBRID ENSEMBLE	0.5	NORMAL	5.459	2.502	3.467	2.460	97.06	85.78	84.73
		DOS	5.272	3.502	3.467	2.390	94.59	81.73	79.73
		PROB	4.072	3.502	2.367	1.890	95.57	85.74	82.73
		U2R	5.272	2.612	2.607	3.490	95.61	86.73	85.67
		R2L	4.852	2.457	2.167	3.610	93.59	87.76	83.73
IMPROVED ENSEMBLE		NORMAL	5.259	2.602	3.507	2.510	97.56	86.78	85.73
		DOS	5.272	3.603	3.497	2.500	95.58	83.74	82.73
		PROB	4.072	3.612	2.497	2.890	96.57	86.81	83.74
		U2R	5.267	2.643	2.612	3.610	96.64	87.73	86.73
		R2L	4.852	2.457	2.167	3.547	94.58	88.56	84.67

Table 1: Comparison analysis of three algorithms on the basis of detection rates, Precision Rates, and Recall Rates.

On the basis of result outcomes we can say that the improved ensemble algorithm performs better than Simple Ensemble and Hybrid Ensemble algorithms and Improved Ensemble reaches to the detection rate up to 98% better than previous two algorithms.

6 CONCLUSION:

In this thesis proposed a feature based intrusion data classification technique. The reduce feature improved the classification of intrusion data. The reduction process of feature attribute performs by RBF function along with feature correlation factor. The proposed method work as feature reducers and classification technique, from the reduction of feature attribute also decrease the execution time of classification. The decrease time increase the performance of intrusion detection system. Our experimental process gets some standard attribute set of intrusion file such as port type, service, sa_srv_rate, dst_host_count, dst_host_sa_srv_rate. These feature attribute are most important attribute in domain of traffic area. The classification rate in these attribute achieved 98 %.

In this thesis reduction computational time of feature selection process is main objective. Because of this consumed time of each algorithm with different reject threshold measured. As evaluation result shows, although FFR cannot defeat other methodologies in accuracy of

classification and accuracy didn't changed very much, but in speed FFR outperformed all other feature selection method with great differences. We used ID3 classifier for developing efficient and effective IDS. For improving the detection rate of the minority classes in imbalanced training dataset we used standard sampling and we picked up all of the important features of the minority class using the minority classes attack mode.

7 SUGGESTION FOR FUTURE WORK:

The proposed algorithm is a combination of feature selection and feature reduction for intrusion detection system. The feature selection and reduction both improved the performance of classification algorithm, but it not achieved the classification ratio 100%. The process of data sampling improved the reduction process and improved the classification ratio up to 100%. The sampling process design as mixed sampler corresponding to the nature of network traffic data, the network traffic data is mixed data type some are continuous and discrete.

8 REFERENCES:

[1] Abebe Tesfahun, D. Lalitha Bhaskari "Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction" International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, 2013.

[2] Deepak Rathore and Anurag Jain "a novel method for intrusion detection based on ecc and radial bias feed forword network" in Int. J. of Engg. Sci. & Mgmt. (IJESM), Vol. 2, Issue 3: July-Sep.: 2012.

[3] Anshul Chaturvedi and Prof. Vineet Richharia "A Novel Method for Intrusion Detection Based on SARSA and Radial Bias Feed Forward Network (RBFFN)" in international journal of computers & technology vol 7, no 3.

[4] Hachmi Fatma, Limam Mohamed "A two-stage technique to improve intrusion detection systems based on data mining algorithms" IEEE, 2013. Pp 1-6.

[5] Terrence P. Fries "A Fuzzy-Genetic Approach to Network Intrusion Detection" in GECCO 08, July12–16, 2008, Atlanta, Georgia, USA.

[6] Zorana Bankovic, Dusan Stepanovic,Slobodan Bojanic and Octavio Nieto-Taladriz "Improving network security using genetic algorithm approach" in Published by Elsevier Ltd 2007.

[7] Ren Hui Gong, Mohammad Zulkernine and Purang Abolmaesumi "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection" in IEEE 2005.