Anomaly Detection using Feature Selection and SVM Kernel Trick

R. Ravinder Reddy
Asst.professor
CSE, CBIT
Hyderabad,India

Y. Ramadevi, PhD
Professor & Head
CSE, CBIT
Hyderabad,India

K.V.N Sunitha, PhD
Principal
BVRIT
Hyderabad,India

ABSTRACT
Analysis of system security becomes a major task for researchers. Intrusion detection plays a vital role in the security domain in these days, Internet usage has been increased enormously and with this, the threat to system resources has also increased. Anomaly based intrusion changes its behaviour dynamically, to detect these types of intrusions need to adopt the novel approaches are required. Detection of intrusion is very important at the same time both accuracy and speed are imperative factors in the real environment. Analyzing intrusive behaviour of the network data is crucial because it contains huge amounts of data as well as the dimensions of the data are also a problem to researchers in detecting intrusive behaviour. In this paper rough set theory is used for the dimensional reduction and the feature selection. Once feature selection is done, Support Vector Machines (SVM) is used to classify the reduct data by using kernel trick. SVM works based on the structural risk minimization principle. It classifying the data in the faster manner with more accuracy to detect the intruder, here we achieved better results than existing techniques.

Keywords
kernel trick, anomaly detection, support vector machine, features selection.

1. INTRODUCTION
Usage of the internet has increased enormously with this everyone is using and benefitting the services of the internet. Along this, the threat to the information also increased gradually. Exchanging the information via the internet is common these days than the other modes of communication. The intruders are trying to get this information for doing an unauthorized transaction in the web. There are so many techniques available to find the intruder's actions in the system. Most of the people are applying machine learning, data mining [1] and soft computing techniques to detect the intruder's attitude.

From the inception of the IDS model by Dorothy E. Denning [2] many investigators working on this model and applying distinguished methods to design the intrusion detection system, even though there are mystifying complications are there in the present intrusion detection systems which comprise:

1. False positives
2. False Negatives
3. Efficiency
4. IDS Security

Now the biggest challenge for the researchers is to reduce the false alarm and detect the novel attacks with accuracy. To meet the intrusion measures like the high accuracy and the low false alarm rate requires efficient methods. In this paper we are using the rough set based feature selection and support vector machine classification technique. In the research area of intrusion detection desperate Improvement is required in the two aspects one is accuracy of the classifier and the detection speed, these two issues are an open research challenge for contemporary researchers. In this paper we are trying to improve these factors by using the contrasting data aid approach instead of changing classifier's techniques, it gives substantial progress in the accuracy and speed.

The rest of the paper is organized as the following manner, section 2 outlines the concepts which are used in this paper. In section 3 process of implementation, in section 4 experiment and results, in section 5 conclusions and future work are discussed.

2. RELATED WORK
In this section briefly outline of the concepts used in this paper are presented.

2.1. Intrusion Detection
Intrusion is an attempt to access the system resources in an unauthorized way to modify or destroy the resources from outsiders or may be the insiders. So intrusion detection system is the second wall of the protection of the system. The firewall will only filter packets. Basically based on the behavior of intruders it divides two aspects

- **Misuse detection**

In this approach user as to define the predefined patterns or signatures to detect the malicious behavior. These type intrusion detection systems can we called as statistical systems like snort, Bro etc,. These systems will detect the known attacks accurately but the problem with this is it won't detect the unknown attacks.

- **Anomaly Detection**

Intruder's behavior is dynamic in nature to break the security firewalls they will come up with new patterns of attacks in disguised manner. In these situations we need a dynamic model to detect the attacks. Anomaly detection is used to detect these types' of attacks in dynamic nature. It will detect novel attacks but the false rate is high.

2.2 Support Vector Machines
A support vector machine is a machine learning technique introduced by vapnik, further it is developed and used by many researchers in the era of artificial learning, its widespread dominance in machine learning applications, heightened acceptance by researchers, it is being enforce in both supervised and unsupervised techniques. Theme of the Support Vector Machine (SVM) is to derive a hyper plane

that maximizes the separating margin between the classes [3]. The promising property of the SVM is that it is an approximate implementation of structural risk minimization principle based on statistical learning theory rather than the empirical risk minimization method. In which the classification function derived by minimizing the mean square error over the training data set.

In the linear SVM it searches for a hyper plane with the largest margin, But in the nonlinear SVM it transforms its data from its original coordinate space into a new space, so that a linear decision boundary can be used to separate the instances in transformed space. But it leads a severe problem of dimensionality. To solve this problem kernel trick has been used. The kernel trick is a method for computing the similarity in the transformation space using the original attribute set. But selecting an appropriate kernel is a big problem, In the SVM the commonly used kernels are polynomial, sigmoid and RBF. The prime reasons of using SVM for intrusion detection is the data used in the IDS model is huge to handle such data is difficult, but in the SVM the number of operations is not necessarily proportional to the number of features. The kernel defines a similarity measure between two data points and thus allows one to incorporate prior knowledge of the problem domain [17].

2.3 Rough Set Theory

Rough set theory introduced by Z Pawlak [4, 5, 6], it is a mathematical tool used for representing un précised and vague data. Rough set theory mainly used for feature selection and knowledge discovery. It helps us to find out minimal attribute sets called reducts to classify objects without deterioration of classification quality. Intrusion detection data contains huge number of training records. They store the huge quantity of data which is hard to manage from a computational point of view. The aim of feature subset selection is to find out a minimum set of relevant attributes that describe the data set as well as the original all attributes do. Finding reducts in large information systems is still an np-hard problem [7].

2.4 Dataset

To evaluate any system we need a benchmark input and compare the results. Fortunately for evaluation of the intrusion detection system we have The KDD Cup'99 dataset. It is a public repository to promote the research works in the field of intrusion detection. Under the sponsorship of the Defence Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL), MIT Lincoln laboratory has collected and distributed the datasets for the evaluation of IDS. Further it is evaluated and transformed to NSL KDD dataset for the better evaluation of the intrusion systems. The NSL KDDCup99 [8, 9] dataset is subsets of the DARPA benchmark dataset. It contains 41conditional attributes and one class label. It is a standard dataset being used for intrusion detection. In this the attacks are distributed in a probabilistic manner.

3. DESIGN AND IMPLEMENTATION

Anomaly detection is challenging task in the security environment, Because of its uncertain presence in the data and many new types of attacks being imposed into the system regularly [10]. To weigh such behaviour in system need a fast and accurate technique is required. In this we suggested a technique which is depicted in Figure1, which incorporates the following steps:

1. Pre processing
2. RST based feature selection
3. SVM Train
4. SVM Test
5. Results analysis

These steps are explained in the following sub sections.

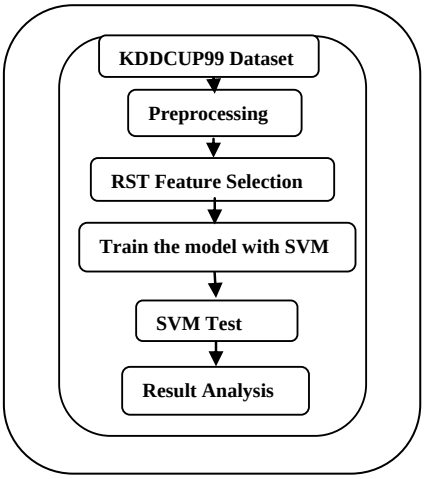

Fig1: Process flow of Proposed RST SVM Approach for Intrusion Detection.

3.1 Data Pre-Processing

The pre-processing step makes the data ready for the experimentation, it contains several approaches, in this phase redundant and useless data be filtered and modified. In this paper we used the NSL-KDD cup'99 dataset which derived from the DARPA is used to test the system performance. The Discretization technique is used to convert quantitative data into the qualitative data format. For example, continuous attributes to be converted into discrete intervals. KDD data set contains heterogeneous data, SVM can't handle such data, and for this purpose descretization is applied. Each record of the NSL KDD dataset contains 41 conditional attribute and one class label which is labelled as either normal or an attack, with exactly one specific attack type [8, 9]. To handle such a huge data is very difficult in intrusion domains because it requires both speed and accuracy are prime features. To handle the dimensionality problems feature selection is a solution.

3.2 Feature Selection using Rough set Theory

Feature selection is a technique to select a subset of features without losing the decision boundary. There are many feature selection techniques are there like PCA, feature ranking and other approaches, in this paper we used rough set theory for feature reduction. The main contribution of rough set learning is the concept of the reducts. A reduct is a minimal set of conditional attributes with the same capability of objects classification as the whole set of attributes [6].

A dataset may contain more than one reduct. Computing a reduct is still hot research. Core is the intersection of all possible reducts in the given dataset. Reduct core gives the better performance. Bazan et al [11] given different methods

of computing the reduct, different reduct computation methods using the rough set are mentioned below.

1. Johnson's reduct
2. Genetic algorithm based reducts
3. Hotte's Reduct
4. Manual Reducer

Jhonson's algorithm uses a simple greedy approach to compute single reduct. Genetic approach is used to select the minimum hitting sets. Hottes algorithm gives singleton reduct. Manual reducer depends on the features selected by humans. Using the rough set theory approach we computed the reducts using the Genetic algorithm. We conducted experiments on RSES tool to compute the reduct set [11].

Algorithm: Reduct feature set computation for KDD Dataset

Input: NSL-KDD dataset

Output: reduct feature set

1. Apply the descretisation for the quantitative attributes
2. Creation of a reduct of conditional attributes using genetic algorithm. The arrangement of attributes according to its importance.
3. Reduct sets.

Here we used the rough set exploration system tool is used to calculate the reducts. To calculate the reduct for KDD dataset we used the Genetic algorithm. KDD dataset is a heterogeneous type of data in size it is a huge data set, for these types of dataset searching of reduct in optimally is a critical task; Genetic algorithm works efficiently such types of data, it uses the optimization technique for generating reducts. After the RST experimentation technique, the feature table has to be reduced to 15 features, the subset of feature set is follows

{duration, src_bytes, dst_bytes, hot, num_root, count,srv_count, dst_host_count, dst_host_srv_count, dst_host_same_srv_count, dst_host_diff_srv_count , dst_host_same_src_port_rate , dst_host_srv-diff_host_rate , dst_host_serror_rate , dst_host_rerror_rate }

3.3 Intrusion Detection using SVM Classification

The Support Vector Machine [12] is one of the most successful classification algorithms in the data mining area. SVM is applied in different ways for finding the intrusion detection [13] in supervised as well as in the unsupervised manner. The Support Vector Machine classification task involves SVM Train and SVM Predict.

3.3.1 SVM Train

NSLKDD Train dataset is used to train the SVM. Using the training data it finds support vectors, after obtaining the support vectors, it will construct SVM model. According to this model, the SVM will classify the test dataset into target classes. Theoretically, SVMs [12, 14] are learning machines which plot all the training vectors in high dimensional feature space. All the vectors are labeled according to their class. In SVMs the data is classified based on the support vectors which are members of the training input set outlining a hyper plane in feature space. Computing the hyper plane to separate the data points leads to a quadratic optimization problem [15]. To solve these problems SVM uses the kernel trick.

Implementation of the SVM contains three kernel functions; here we worked on these kernel functions.

1. Gaussian Kernel (Radial Basis Function)
2. Polynomial kernel
3. Sigmoid kernel

The svmlib [12] provides the implementation details of these kernels in binary as well as multi class classification approaches. Here we mention the basic usage of these kernel functionalities in the following section.

(i) Gaussian Kernel Function: The Gaussian kernel is an example of the radial basis kernel function.

$$K(X_i, X_j) = \exp\{-(\|X_i - X_j\|)\}/2\sigma \text{ ---- (1)}$$

(ii) Polynomial Kernel Function: This Polynomial kernel is a non-stationary kernel. Polynomial kernels are well suited for problems where all the training data is normalized. Adjustable parameters are the constant term 'c' and the polynomial degree 'd' .

$$K(X_i, X_j)=(X_i, X_j)^d + c \quad \text{------- (2)}$$

(iii) Sigmoid Kernel Function: Sigmoid Kernel is also called as Hyperbolic Tangent Kernel and the Multilayer Perceptron (MLP) kernel. The Sigmoid Kernel comes from the Neural Networks field, where the bipolar sigmoid function is often used as an activation function for artificial neurons.

$$K(X_i, X_j) = \tanh(k(X_i, X_j) + r \quad \text{----- (3)}$$

In training phase, the SVM training model is built and SVM kernel function is selected to generate classification results, we run the experimentation for these three kernel functions.

3.3.2 SVM Test

After building the SVM model using the training data, it has to be tested with test data. We used NSL-KDD test dataset to predict the target value of data instance. In the experimentation we used the multi class SVM. The KDD training dataset contains four main types of attacks, all the threats grouped into these four categories, it predicts appropriate class based on the model. The SVM Predict calculates the accuracy rate and False alarm rate. The following fundamental formulas are used to estimate the performance of the system: accuracy rate (AR) and false positive rate (FPR).

$$\text{Accuracy Rate} = \frac{\text{Total number of correct classified process}}{\text{Total number of process}} \times 100\% \text{ --- (4)}$$

$$\text{False Alarm Rate} = \frac{\text{Total number of misclassified process}}{\text{Total number of process}} \times 100\% \text{ ---- (5)}$$

4. EXPERIMENTS AND RESULTS

Conducting the experimentation NSL-KDD [8, 9] dataset is used. Using the feature selection reduct feature set is computed. By using the reduct deduce the new dataset with the reduct. After preparing the new dataset we run the SVM on both the data sets using kernel trick. Here we applied the three kernel functions for both the datasets and obtain the results.

The experiments are conducted separately for both the rough set based SVM(RSTSVM) and SVM. These experiments conducted for three kernel functions. RSTSVM approach gives good performance measures like accuracy and less false alarm rate. This approach is efficient in the case memory usage as well as computational time aspects. This approach enhances the system performance in detecting intrusion behavior.

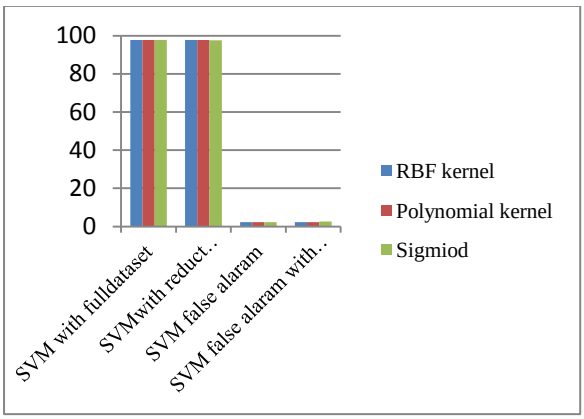

Fig 3: comparison of accuracy and false alarm values of svm and rough sets with svm(RSTSVM)

Here the SVM and the rough set based SVM(RSTSVM) results are compared in the table 1 for three kernel functions. Among these RBF kernel gives the better results than the others. Rough set based method shows good accuracy with less computation. It produces the results in less amount of time so it can be used to detect the real time intrusions in time.

Table 1: Comparison of Accuracy and False alarm rate

Kernel	Accuracy		False alarm rate	
	SVM	Rough set with SVM	SVM	Rough set with SVM
Gaussian RBF	97.73	97.76	2.318	2.283
Sigmoid	97.71	97.73	2.336	2.319
Polynomial	97.69	97.50	2.355	2.55

Table 2: Comparison of SVM classification Training and Testing time taken for 41 features and 15 features

Method / Time	Before Feature selection (With all 41 features)	After Feature selection (With 15 features)
Training time taken	3751 ms	2022 ms
Testing time taken	42,275 ms	34,718 ms

In the Table 2, it shows that the SVM classification Training time and Testing time taken for all 41 features of SVM (Before feature selection) and 15 features of Roughest based SVM (After feature selection). The training and testing time taken for the model with the reduct dataset is less than that of original one. The number of support vectors for both the models are different, double the size of the support vectors are required in the first case.

Table 3: Comparison of feature selection methods Rough Set Theory with existing Principal Component Analysis (PCA)

Feature selection method	Number of selected features	Accuracy
Roughset Theory (RST)	15	97.76
Principal Component Analysis (PCA)	12	86.66

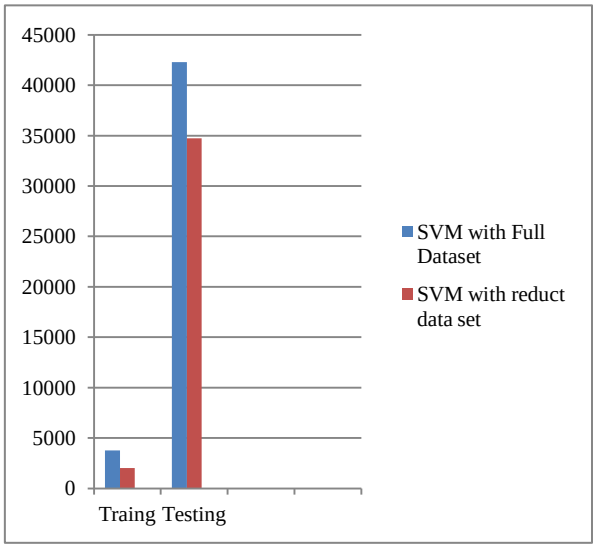

Fig 4: Comparison of SVM Training and testing time in milliseconds for full dataset and reduct dataset

In the Fig 4, it shows that SVM classification Training and testing time taken for 41 features and 15 features, the reduct dataset reduces the training and testing of the SVM without deteriorating the classifier accuracy.

In the Table 3, we compare our rough set based Feature selection technique with the PCA feature selection. Our approach has better accuracy as compared to the Principal Component Analysis (PCA) [16] Feature selection method. PCA is a common statistical method for use in multivariate optimization problems to reduce data dimensionality while retaining a large fraction of the characteristics of data. First, PCA projects the training set into eigenspace vectors representing data mean. These eigenspace vectors then predict malicious connections in a workload with normal and attack behavior [19]. Principal component analysis is used for feature subset selection which is based on highest eigenvalues, but the features corresponding to the highest eigenvalues may not have the optimal sensitivity for the classifier due to ignoring many sensitive features.

5. CONCLUSION AND FUTURE WORK

In this work, the Intrusion Detection approach uses the rough set based feature selection and the SVM kernel trick. The kernel divides the data points and increases classification accuracy. Rough set reduct reduces the dimensionality of the input space, thus increases the system performance and decreases the memory usage, it saves lot computation resources. The proposed Intrusion detection system using Rough-set theory and Support Vector Machine (RSTSVM) reduces the number of features from 41 to 15 and compared the performance of the SVM. The RSTSVM method result has a higher accuracy and less false alarm rate as compared to SVM original dataset.

In future genetic algorithm and rough fuzzy techniques can be combinely applied to SVM for enhancing the performance and accuracy of real time intrusion detection.

6. ACKNOWLEDGEMENTS

The first author expreses his gratiude to Dr. C.R Rao, Profesor, Dept. of CIS, HCU for his suport and also take the oportunity to express thanks to Management of CBIT, and Dr.B. Chenna Kesava. Rao, Principal, CBIT for their encouragement and cooperation.

7. REFERENCES

[1] Lee W and Stolfo S., "Data Mining techniques for intrusion detection", In: Proc. of the 7th USENIX security symposium, San Antonio, TX, 1998.

[2] Denning D. (1987) "An Intrusion-Detection Model," IEEE Transactions on Software Engineering, Vol. SE-13, No. 2, pp.222-232.

[3] K.P Lin and M.S Chen, "Efficient kernel approximation for large-scale support vector machine classification," in Proceedings of the Eleventh SIAM International Conference on Data Mining, 2011, pp. 211–222.

[4] Pawlak Z: Rough sets Present state and the future. Foundations of computing and Decision sciences 18,157-163 (1993).

[5] Pawlak Z: Rough Sets and Intelligent Data Analysis, Information Sciences,2002, 147:1-12.

[6] Pawlak Z, Rough Sets, International Journal of Computer and Information Sciences, vol. 11, pp. 341-256, 1982.

[7] Boussouf M (1998) A Hybrid Approach to Feature Selection. Lecture Notes in Artificial Intelligence 1510:231–238.

[8] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[9] Tavallaee M, Bagheri E, Wei Lu, Ghorbani A., "A detailed analysis of the KDD CUP 99 data set," Computational Intelligence for Security and Defence Applications, 2009. CISDA 2009. IEEE Symposium on , vol., no., pp.1,6, 8-10 July 2009

[10] Ravinder Reddy R, et al, "Real time anomaly detection using Ensemblers" Proceedings of 5ht ICISA IEEE conference.

[11] Jan G. Bazan, Marcin Szczuka, "The rough set exploration system (2005)" TRANSACTIONS ON ROUGH SETS III, springer.

[12] LIBSVM -- A Library for Support Vector Machines:www.csie.ntu.edu.tw/~cjlin/libsvm/

[13] R Ravinder Reddy, B.Kavya, Y Ramadevi. "A Survey on SVM Classifiers for Intrusion Detection" International Journal of Computer Applications (0975-8887) July 2014, pp: 38- 44.

[14] V.N.Vapnik, The nature of statistical learning theory. Springer-Verlag, New York. NY, 1995.

[15] C. Cortes and V. Vapnik, "Support-vector network," Machine Learning, vol. 20, pp. 273–297, 1995

[16] Xu P and Chan A., An efficient algorithm on multi-class support vector machine model selection. Proceedings of the International Joint Conference on Neural Networks, 4:3229–3232, 2003.

[17] http://svms.org/kernels/

[18] en.wikipedia.org

[19] Reconfigurable Architecture for Network Intrusion Detection Using Principal Component Analysis, David Nguyen, Abhishek Das, Gokhan Memik, Alok Choudhary — 2006 — In Proc. ACM/SIGDA 14th international.