INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

A NOVEL METHOD FOR INTRUSION DETECTION BASED ON ECC AND RADIAL BIAS FEED FORWORD NETWORK

Deepak Rathore*[1], Anurag Jain[2]

*[1,2] Department of Computer Science & Engg. , RITS, Bhopal (M.P), India

ABSTRACT

In this paper Neural Network model combined with prototype clustering and classification for fast and accurate detection of intrusion in host based system. Previous RBF suffered from grouping of pattern of intrusion, now this problem are reduced using Distance variable ensemble cluster classification and increase the rate of detection of infected data in host system. Our methodology test in KDD CUP 99 and calculate the rate of detection Accuracy, Precision, Recall, False positive rate, false negative rate, true positive rate, True negative rate.

Keywords— Intrusion detection, Classification, Clustering, ECC, ECC-RBF, KDD CUP 99.

INTRODUCTION

Intrusion detection is a critical process in network security. Traditional methods of network intrusion detection are based on the saved patterns of known attacks. They detect intrusion by comparing the network connection features to the attack patterns that are provided by human experts. The main drawback of the traditional methods is that they cannot detect unknown intrusions. Even if a new pattern of the attacks were discovered, this new pattern would have to be manually updated into the system. intrusion detection technology was divided into two categories:

Corresponding Author*

Email- rathore.rath@gmail.com

I. Misuse Detection

Misuse Detection: Misuse Detection identify illegal invasion by the known attack methods and system vulnerabilities. The main disadvantages of the methods is: because all known intrusion model have been implanted in the system, once any form of unknown intrusion appear, it cannot be detected. But detection efficiency of this method is high.

II Anomaly Detection

Anomaly Detection: Anomaly detection identifies illegal invasion or authority exceeding by checking whether the current user behavior has been deviated from normal behavior profile established. The advantage of this method is no needing on understanding the defect of system, good adaptability. But the possibility of error is high.

Intrusion detection is a software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a Management Station. Intrusion Detection System (IDS) is an important detection that is used as a countermeasure to preserve data integrity and system availability from attacks. The work is implemented in two phases; in first phase clustering by K-means is done and in next step of classification is done with k-nearest neighbors and decision trees. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity [10].

Clustering: The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group and so may be considered as a form of data compression [10]

Classification: classification is an effective means for distinguishing groups or classes of objects; it requires the often costly collection and labeling of a large set of training tuples or patterns which the classifier uses to model each group. It is often more desirable to proceed in the reverse direction Typical classify models have the linear regression model, the decision tree model, the model based on rule and the neural network model [10].

The Intrusion Detection techniques are used to detect the intrusions based on the KDD Cup 1999 dataset. Two datasets, KDD and UNM, are used in experiments to evaluate the performance of the proposed new model. The KDD dataset consists of network connection records generated by a TCP/IP dump. There are 41 features in each record. 10% of the original data are training data with a label which identifies which category the record belongs[7].KDDCup99 training dataset is about four giga bytes of compressed binary TCP dump data from seven weeks of network traffic, processed into about five million connections record each with about 100 bytes. The two weeks of test data have around two million connection records. Each KDDCup'99 training connection record contains 41 features and is labeled as either normal or an attack, with exactly one specific attack type [15].Based on KDD Cup 99, many researchers have been conducted, which falls into two key parts: detection model generation and intrusion feature selection.

We developed an Intrusion Detection System using AMSTAR neural network to learn patterns of normal and intrusive activities and to classify observed system activities. we further investigate the time taken for training and testing, generate Confusion matrix using simulation tool and compare it with five classification techniques (Gaussian Mixture, Radial Basis Function, Binary Tree Classifier, SOM, and ART)[17].

Radial Basis Function classifiers calculate discriminate functions using local Gaussian functions. A total of six simulations were performed using the RBF algorithm .Each simulation used initial clusters created using K-means algorithm, .Weights are trained using least-square matrix inversion to minimize the squared error of the output sums given the basis function outputs for the training patterns. During training and testing variance are increased to provide good coverage of the data .For each simulation using the RBF.[17]

RELATED WORK

The classifications of intrusion detection and methods of data mining applied on them were introduced. Then, intrusion detection system design and implementation of based on data mining were presented. Such a system used..Experiments showed that new type of attack can be detected effectively in the system, and knowledge base can be updated automatically, so the efficiency and accuracy of the intrusion detection were improved, and security of the network was enhanced[4]Network security is becoming an increasingly important issue, since the rapid development of the Internet. Network Intrusion Detection System (IDS), as the main security defending technique, is widely used against such malicious attacks. Data mining and machine learning technology has been extensively applied in network intrusion detection and prevention systems by discovering user behavior patterns from the network traffic data. The training data has been clustered into 2-clusters before feeding the initial population hoping that data will be divided into normal and abnormal clusters.

In this study, we extend a scalable, incremental, and supervised clustering and classification algorithm CCAS into ECCAS that has the capacity of handling data with both numeric and nominal variables. Two different methods of handling mixed data types are developed. The two methods of ECCAS are tested and compared on a data set with mixed Variable types for intrusion detection. Both methods produce Comparable performance to that of the winning algorithm in a

Data mining contest on the same data set. The performance on different data sets shows the reliability of ECCAS. The testing results for one data set also show that the five phases of ECCAS reduces the impact of the data presentation order on the prediction accuracy. The number of Grid intervals shows the impact on the prediction accuracy of ECCAS. The ECCAS algorithm and the distance Measure could be used in common data mining applications. We are developing methods to adaptively and dynamically adjust the parameters during training, including the grid-interval configuration and the threshold-controlling outlier removal [3]. A Dependable Network Intrusion Detection System (DNIDS) based on the Combined Strangeness and Isolation measure K-Nearest Neighbour (CSIKNN) algorithm. The intrusion detection algorithm analyzes different characteristics of network data by employing two measures: strangeness and isolation. But in general the K-NN still needs intensive computations. The Unsupervised Anomaly Detection Using an Optimized K-Nearest Neighbors Algorithm can work without the need for massive sets of pre-labeled training data. a k-nearest neighbors algorithm to detect anomalies in network connections, as well as the optimization necessary to make the algorithm feasible for a real-world system. [9].

The development of anomaly based intrusion detection systems during the recent years. As several supervised and unsupervised clustering techniques were optimized resulting in more elegant techniques that provided more detection accuracy and lower false alarm rate. Moreover, the newly proposed techniques tend to avoid the creation of unnecessary neurons in the training process to faithfully represent data inputs as applied in hierarchical clustering. Furthermore, this restriction in creating neurons significantly contributes in reducing the complexity of the training process and producing more accurate topologies. Since, our main concern in our research is to increase the quality of clustering and attacks classification for larger scope of attacks. Additionally, increasing the identification rate of novel patterns in the training process as well.

Intrusion Detection System (IDS) is an important detection that is used as a countermeasure to preserve data integrity and system availability from attacks. The work is implemented in two phases; in first phase clustering by K-means is done and in next step of classification is done with k-nearest neighbors and decision trees. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity. This paper proposes an approach which makes the clusters of similar attacks and in next step of classification with K nearest neighbors it detect the attack types. This method is advantageous over single classifier as it detect better class than single classifier system [10].

The work is done through which the novel class is detected the system uses K-Means clustering algorithm which produces different clusters of similar type of attacks and total nodes per clusters in the input dataset. Also it shows the updated cancroids of each parameter in the input dataset. The Classification stage gives details about detection of different types of attacks and number of nodes in dataset. A false negative occurs when an intrusion action has occurred but the system considers it as a non-intrusive behavior. A false positive occurs when the system classifies an action as an intrusion whiles it is a legitimate action. A good intrusion detection system should perform with a high precision and a high recall, as well as a lower false positive rate and a lower false negative rate. To consider both the precision and false negative rate is very important as the normal data usually significantly outnumbers the intrusion data in practice. To only measure the precision of a system is misleading in such a situation.

False alarm rate and detection accuracy are still challenging issues that are not completely solved yet in the field of Anomaly based Intrusion Detection System (AIDS). The reasons behind these issues vary according to the algorithm and the dataset used to train the IDS. Consequently, dealing with high dimensional data requires an efficient data reduction technique that considerably reduces the dimensionality without any substantial loss in the important features. However, the excessive reduction of features will lead to model some intrusive patterns similarly as normal ones. Indeed, this will result in misclassifications that will increase false negative rate, which degrades the accuracy of detection.[14].

METHODOLOGY

ECC Algorithm:

This paper presents an extended version of CCAS clustering and classification algorithm—supervised (CCAS), (ECCAS) that enables the handling of mixed data types. The application of ECCAS to computer intrusion detection, using network traffic data with mixed data type. We apply ECCAS to intrusion detection using the Knowledge Discovery and Data Mining (KDD) Cup 1999 data.

(1)*Divide dataset into chunk D1, D2, D3, _____Dn+1.*
(2) *Generate discrete random number of seed for generating of cluster.*
(3)*Initialized distance weight factor.*
(4) *Calculate min of data chunk and standard deviation.*
(5) *Compare value at min with seed value.*
(6) *Then generate cluster.*
(7) *Set label of class C1, C2, C3.*
(8) *Assigned training at data.*
(9) *Generate classifier-Merge set of cluster & classifier with label.*
(10) *Calculate standard deviation (error).*
(11) *Ensemble class.*

PROPOSED WORK

ECC-RBF:

In this section of method RBF network implied on clustering classification ensembles technique for better training purpose of data for improvement of classification rate of ensembles technique. A radial basis function (RBF) is a real-valued function whose value depends only on the distance from the origin. If a function ‘h’ satisfies the property h(x)=h(||x||), then it is a radial function. Their characteristic feature is that their response decreases (or increases) monotonically with distance from a central point.

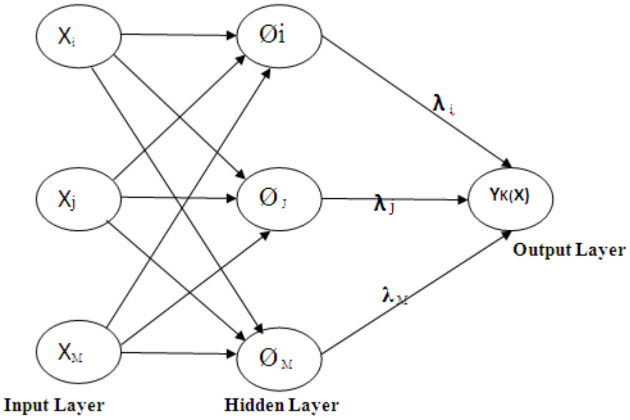

Figure 1: RBF Network

A Gaussian function is specified by its centre and width. The simplest and most general method to decide the middle layer neurons is to create a neuron for each training pattern. However the method is usually not practical since in most applications there are a large number of training patterns and the dimension of the input space is fairly large. Therefore it is usual and practical to first cluster the training patterns to a reasonable number of groups by using a clustering algorithm such as K-means or SOFM and then to assign a neuron to each cluster. A simple way, though not always effective, is to choose a relatively small number of patterns randomly among the training patterns and create only that many neurons. A clustering algorithm is a kind of an unsupervised learning algorithm and is used when the class of each training pattern is not known. But an RBFN is a supervised learning network. And we know at least the class of each training pattern. So we’d better take advantage of the information of these class memberships when we cluster the training patterns.

```
Input: training patterns X== {x1; x2,…………,xP}
Output: centers of clusters
Variable
C: number of clusters
cj : center of the j-th cluster
nj : number of patterns in the j-th cluster
di j : distance between xi and the j-th cluster
begin
  C =1; c1  x1;n1 :=1;
  for i :=2 to P do /* for each pattern */
    for j :=1 to C do /* for each cluster */
      compute di j;
      if di j _R0 then
        /* include xi into the j-th cluster */
        cj  (cjnj +xi)=(ni+1);
        ni :=ni+1;
        exit from the loop;
      end if
    end for
    if xi is not included in any clusters then
      /* create a new cluster */
      C :=C+1;
      cC  xi;
      nC :=1;
    end if
  end for
end
```

APC-III is quite efficient to construct the middle layer of an RBF since we can finish clustering by going through the entire training patterns only once. This is not true with K-means and SOFM clustering algorithms.

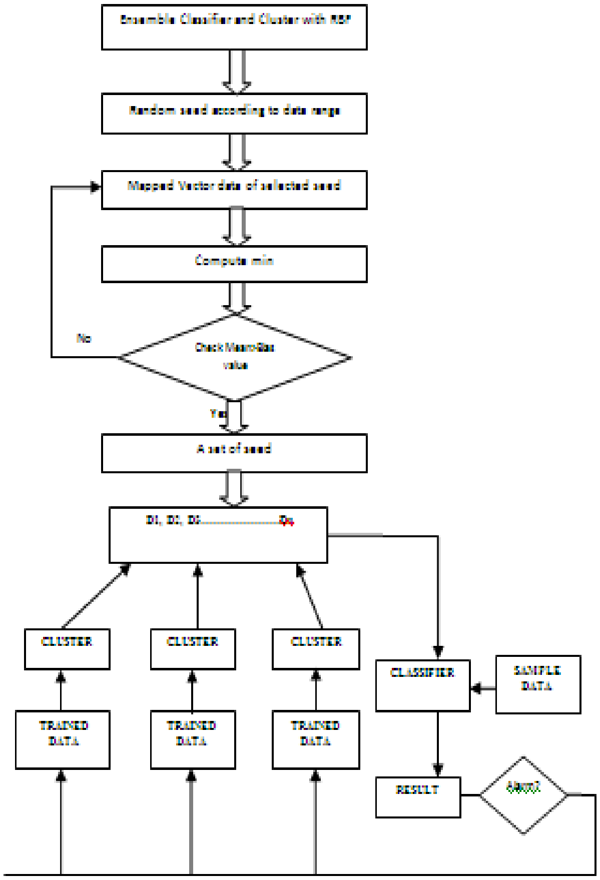
Ensemble Classifier and Cluster with RBF
Random seed according to data range
Mapped Vector data of selected seed
Compute min
No
Check Mean-Bias value
Yes
A set of seed
D1, D2, D3.....................Dn
CLUSTER
CLUSTER
CLUSTER
TRAINED DATA
TRAINED DATA
TRAINED DATA
CLASSIFIER
SAMPLE DATA
RESULT

Figure 2: ECC-RBF MODEL

RESULT

The figure shown below shows the classification:

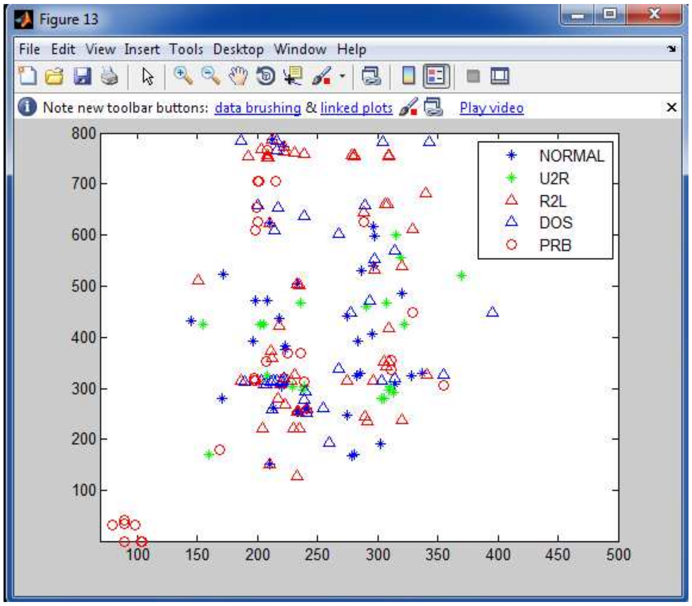

Figure 3: ECC Result

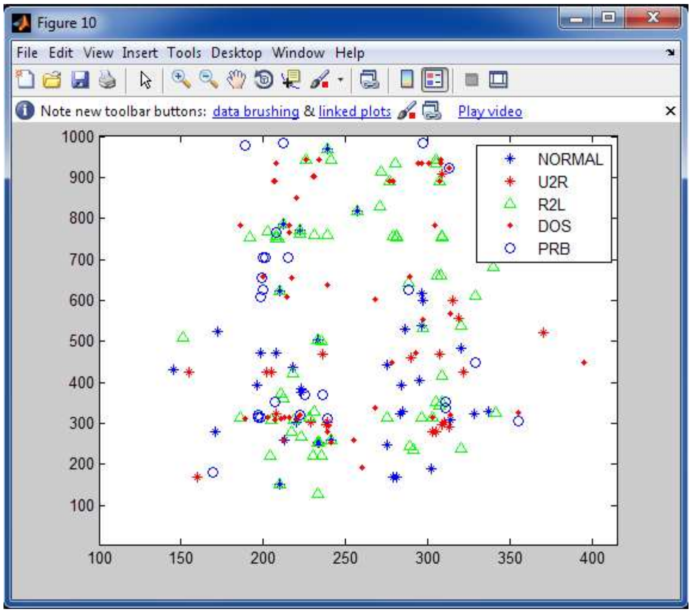

Figure 4: ECC-RBF Result

Metric		Accuracy (%)	Precision (%)	Recall (%)
Data-Set 1	ECC	92.14	87.24	84.43
	ECC-RBF	96.34	91.02	90.01
Data-Set 2	ECC	89.9	84.32	83.23
	ECC-RBF	97.34	90.02	90.12
Data-Set 3	ECC	91.34	86.14	85.11
	ECC-RBF	96.8	90.21	90
Data-Set 4	ECC	92.22	88.21	87.66
	ECC-RBF	98.54	95.76	95.4

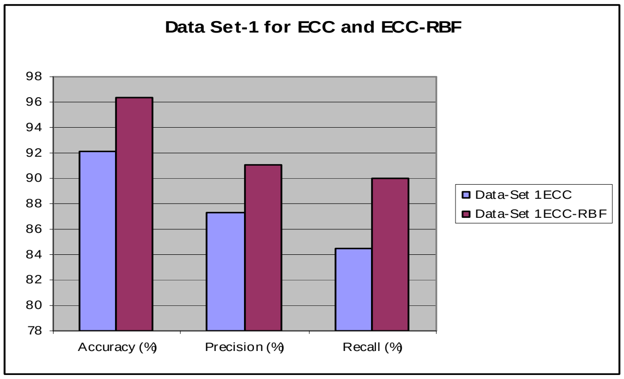

Figure 5: Result of ECC and ECC-RBF for Data Set1

Description:- This figure show the Accuracy, Precision, and Recall on ECC and ECC-RBF for data set 1(KDD CUP 99).

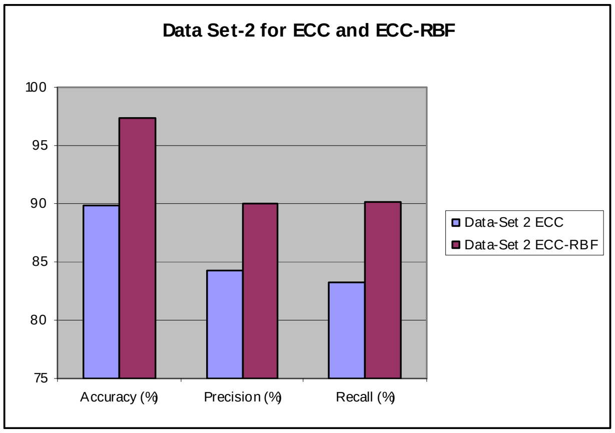

Figure 6: Result of ECC and ECC-RBF for Data Set 2

Description:- This figure show the Accuracy, Precision, and Recall on ECC and ECC-RBF for data set 2(KDD CUP 99).

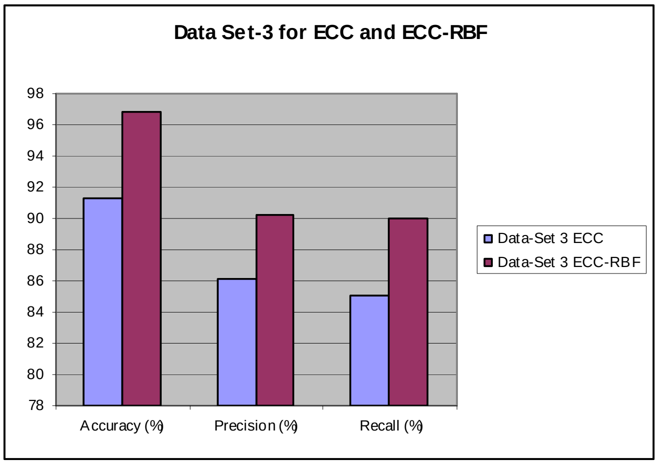

Figure 7: Result of ECC and ECC-RBF for Data Set3

Description:- This figure show the Accuracy, Precision, and Recall on ECC and ECC-RBF for data set 3(KDD CUP 99).

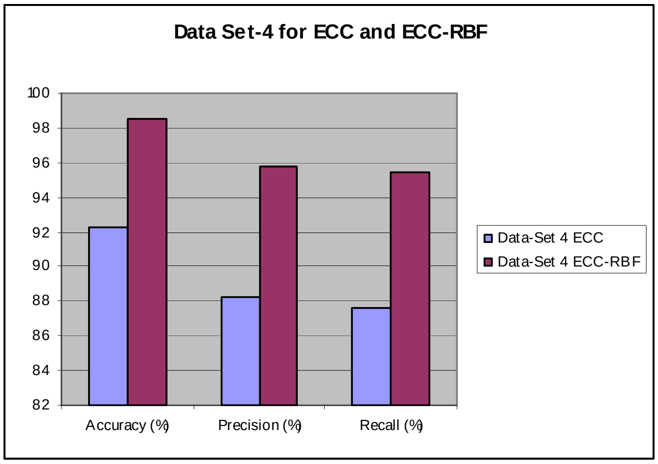

Figure 8: Result of ECC and ECC-RBF for Data Set 4

Description:- This figure show the Accuracy, Precision, and Recall on ECC and ECC-RBF for data set 4(KDD CUP 99).

CONCLUSION AND FUTURE WORK

In this paper proposed prototype classifier based on neural network RBF model. Our model test in KDD CUP 99 data set. evaluation process get better result in comparison of ECC,ECC-SOM and other classifier of data mining, in RBF technique the feature selection process suffered a little bit problem during distance adjustment ,now in future we used any optimization technique for the feature selection process in RBF network such as genetic algorithm and Pos.

REFERENCES

[1] S.Devaraju, Dr.S.Ramakrishnan: "analysis of intrusion detection system using various neural network classifiers"1033-1038, ©IEEE 2011.

[2] K kr.Gupta, B Nath,R Kotagiri" Layered Approach Using Conditional Random Fields for Intrusion Detection" 1545-5971/10/$26.00 © 2010 IEEE.

[3] Xiangyang Li,and Nong Ye" A Supervised Clustering and Classification Algorithm for Mining Data With Mixed Variables" 1083-4427/$20.00 © 2006 IEEE.

[4] Jiankun Hu and Xinghuo D. Qiu, Hsiao-Hwa Chen, " A Simple and Efficient Hidden markovmodel Scheme for Host-basedanomaly Intrusion Detection" 0890-8044/09/$25.00 © 2009 IEEE.

[5] Chunyu Miao and W Chen "A Study of Intrusion detection System Based on Data Mining" 978-1-4244-6943-7/10/$26.00 ©2010 IEEE.

[6] Nong Ye, Syed M Emran, Qiang Chen, and Sean Vilbert" Multivariate Statistical Analysis of Audit Trails for Host-Based Intrusion Detection" 0018-9340/02/$17.00 ß 2002 IEEE.

[7] M Tavallaee, E Bagheri, W Lu, and Ali A. Ghorbani" A Detailed Analysis of the KDD CUP 99 Data Set"978-1-4244-3764-1/09/$25.00©2009 IEEE.

[8] Marimuthu, A. And Dr. A.Shanmugam, 2008. "Intelligent Progression for anomaly Intrusion detection",6th International Sympos ium on Applied Machine Intelligence and informatics , SAMI.

[9] ACMSIGKDD, 1(2): 67-75. Liwei (Vivian) Kuang, 2007. DNIDS: A Dependable Network Intrusion Detection System

[10] Ms. P J. Pathak ,S S. Dongre 2012 "Attack Detection By Clustering And Classification Approach" ISSN: 2277 – 9043 IJAR in Computer Science and Electronics Engineering.

[11] Lei Li, De-Zhang Yang, Fang-Cheng Shen "A Novel rule-based Intrusion Detection System by"2010.

[12] John Zhong Lei and Ali Ghorbani" Network Intrusion Detection Using an Improved Competitive Learning Neural Network" 0-7695-2096-0/04 $20.00 © 2004 IEEE.

[13] hai-hua gao, hui-hua yang, xing-yu wang" ant colony optimization based network intrusion feature selection and detection" 0-7803-9091-1/05/$20.00 ©2005 IEEE.

[14] Y I Shakhatreh, K A Bakar "A Review of Clustering Techniques Based on Machine learning Approach in Intrusion Detection Systems" IJCSI Vol. 82011.

[15] wing w. Y. Ng, rocky k. C. Chang, daniel s. Yeung" dimensionality reduction for denial of service detection problems using rbfnn output sensitivity" 0-7803-7865-2/03/$17.00 02003 IEEE.

[16] P. Natesan a;∗, P. Balasubramanie a;∗, G. Gowrison b;∗" Improving Attack Detection Rate in Network Intrusion Detection Using Adaboost Algorithm with Multiple Weak Classifiers" Journal of Information & Computational Science 9: 8 (2012) 2239–2251.

[17] V.Venkatachalam S.Selvan ," Intrusion Detection using an Improved Competitive Learning Lamstar Neural Network" IJCSNS VOL.7 2007.