A Hybrid Model for Intrusion Detection Based on Genetic Clustering and PSO Algorithm

Rakesh Singh Thakur
M.Tech Scholar, Department of IT
SVITS, Indore India
E-mail-rkt2583@gmail.com

Gaurav Shrivastava
AP, Department of IT
SVITS, Indore India
E-mail-gaurav2086@gmail.com

ABSTRACT

Reduction and selection of intruder attribute in intrusion detection system play an important role in process of detection. The huge number of attribute in intruder induces a problem in detection process and increase more time in decision making process. In this paper we tried to propose a very simple and fast clustering method for intrusion detection. A hybrid scheme based on coupling two different algorithms one is particle of swarm optimization and other is k-means algorithm. The main originality of proposed approach relies on associating two techniques: extracting more information bits via specific linguistic techniques, space reduction mechanisms, and moreover arcing cluster to aggregate the best clustering result. For the validation and performance evaluation of proposed algorithm used MATLAB software and KDDCUP99 dataset 10%. This dataset contains approx 5 lacks number of instance. The process of result shows that better detection ratio in compare of k-means and k-means-GA technique of intrusion detection.

Keyword: - Feature selection, Intrusion detection system, Genetic Algorithm, Clustering.

INTRODUCTION

An Intrusion Detection System (IDS) inspects the activities in a system for suspicious behavior or patterns that may indicate system attack or misuse. There are two main categories of intrusion detection techniques; Anomaly detection and Misuse detection [1]. The former analyses the information gathered and compares it to a defined baseline of what is seen as “normal” service behavior, so it has the ability to learn how to detect network attacks that are currently unknown. Misuse Detection is based on signatures for known attacks, so it is only as good as the database of attack signatures that it uses for comparison. Misuse detection has low false positive rate, but cannot detect novel attacks. However, anomaly detection can detect unknown attacks, but has high false positive rate.

An intrusion detection system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches. In other words, intrusion detection is the act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a system/network. Traditionally, intrusion detection systems have been classified as a signature detection system, an anomaly detection system or a hybrid/compound detection system [13]. A signature detection system identifies patterns of traffic or application data presumed to be malicious while anomaly detection systems compare activities against a “normal” baseline. On the other hand, a hybrid intrusion detection system combines the techniques of the two approaches. Both signature detection and anomaly detection systems have their share of advantages and drawbacks. The primary advantage of signature detection is that known attacks can be detected fairly reliably with a low false positive rate. The major drawback of the signature detection approach is that such systems typically require a signature to be defined for all of the possible attacks that an attacker may launch against a network. Anomaly detection systems have two major advantages over signature based intrusion detection systems. The first advantage that differentiates anomaly detection systems from signature detection systems is their ability to detect unknown attacks as well as “zero days” attacks [2]. This advantage is because of the ability of anomaly detection systems to model the normal operation of a system/network and detect deviations from them. A second advantage of anomaly detection systems is that the aforementioned profiles of normal activity are customized for every system, application and/or network, and therefore making it very difficult for an attacker to know with certainty what activities it can carry out without getting detected. However, the anomaly detection approach has its share of drawbacks as well. For example, the intrinsic complexity of the system, the high percentage of false alarms and the associated difficulty of determining which specific event triggered those alarms are some of the many technical challenges that need to be addressed before anomaly detection systems can be widely adopted.

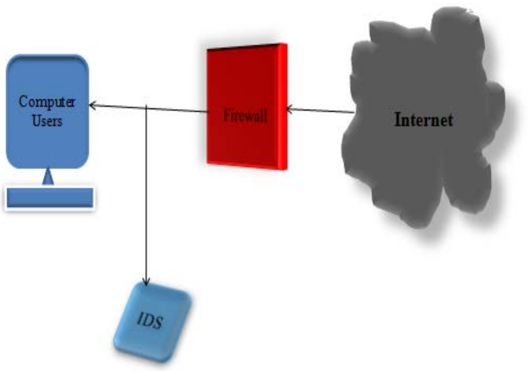


Figure 1: Intrusion detection systems in the network environment.

Types of intrusion detection systems there are two types of intrusion detection systems that employ one or both of the intrusion detection methods outlined above. Host-based systems base their decisions on information obtained from a single host (usually audit trails), while network-based intrusion detection systems obtain data by monitoring the traffic in the network to which the hosts are connected [4]. An intrusion detection system dynamically monitors the events taking place in a monitored system, and decides whether these events are symptomatic of an attack or constitute a legitimate use of the system.

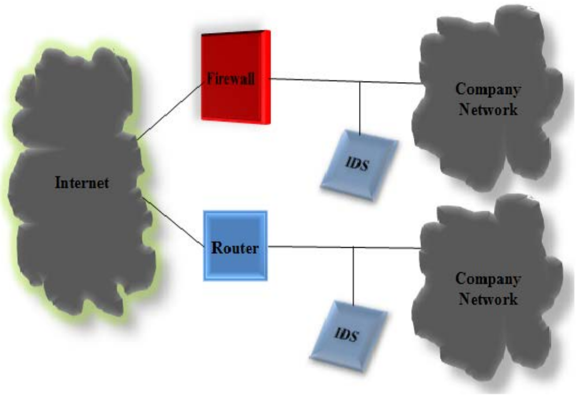


Figure 2: Intrusion detection systems in a network.

Figure depicts the organization of IDS where solid arrows indicate data/control flow while dotted arrows indicate a response to intrusive activities. Section-II gives the information of about feature selection for intrusion detection process. In section III discuss the about clustering. In section IV discuss the proposed method. In section V comparative result finally, in section VI conclusion and future scope.

II FEATURE SELECTION

Feature selection is an essential data processing step prior to applying a machine learning algorithm. It is a process of determining whether a feature is relevant or not for a particular problem. Using effective features to design classifier not only can reduce the data size but also can improve the performance of the classifier and enhances data understanding or visualization [15]. One of the major problems in feature reduction is to select effective attributes that have the best discrimination ability between the classes. There are two common approaches for feature reduction: Wrapper and Filter. A Wrapper method selects feature subset based on the performance of the learning algorithm that is going to be used [3]. Wrapper method is totally dependent on the learning algorithm. On the other hand Filter methods evaluate features according to statistical characteristics of the data only without the involvement of any learning algorithm. The wrapper approach is generally considered to produce better feature subsets but runs much more slowly and requires more computing resource than a filter. Different techniques have been used to tackle the problem of feature selection. feature ranking algorithms to reduce the feature space of the DARPA data set from 41 features to the six most important features. They used three ranking algorithms based on Support Vector Machines (SVMs), Multivariate Adaptive Regression Splines (MARSs), and Linear Genetic Programs (LGPs) to assign a weight to each feature. Experimental results showed that the classifier’s accuracy degraded by less than 1 percent when the classifier was fed with the reduced set of features. Sequential backward search was used in [8], [9] to identify the important set of features: starting with the set of all features, one feature was removed at a time until the accuracy of the classifier was below a certain threshold. Different types of classifiers were used with this approach including Genetic Algorithms, Neural Networks, and Support Vector Machines.

III CLUSTERING

Cluster analysis is the organization of a collection of patterns (usually represented as a vector of measurements, or a point in a multidimensional space) into clusters based on similarity. Intuitively, patterns within a valid cluster are more similar to each other than they are to a pattern belonging to a different cluster [5]. The variety of techniques for representing data, measuring proximity (similarity) between data elements, and grouping data elements has produced a rich and often confusing assortment of clustering methods [8]. It is important to understand the difference between clustering unsupervised classification) and discriminate analysis (supervised classification) [12]. In supervised classification, we are provided with a collection of labeled (pre-classified) patterns; the problem is to label a newly encountered, yet unlabeled, pattern. Typically, the given labeled (training) patterns are used to learn the descriptions of classes which in turn are used to label a new pattern. In the case of clustering, the problem is to group a given collection of unlabeled patterns into meaningful clusters. In a sense, labels are associated with clusters also, but these category labels are data driven; that is, they are obtained solely from the data. Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data mining, document retrieval, image segmentation, and pattern classification. However, in many such problems, there is little prior information (e.g., statistical models) available about the data, and the decision-maker must make as few assumptions about the data as possible.

IV PROPOSED METHODOLOGY

In this section we discuss the proposed algorithm for intrusion detection. The proposed algorithm is combination of k-means algorithm and particle of swarm optimization algorithm. The process of seed selection is done by particle of swarm optimization.

The proposed algorithm of intrusion detection describe as Step 1 initialized the rand function as artificial particle and the range of particle is range of data. The state of particle and velocity of particle select as random fashion. Select random particle as cluster center. Here describe s

$$X_i^{(0)} = \left(p_{i1}^{(0)}, p_{i2}^{(0)}, \dots\dots\dots\dots p_{ik}^{(0)}\right) \dots\dots\dots\dots\dots\dots\dots\dots (4.5.1)$$

Where $p_{i1}^{(0)}$ refers to the jth cluster centroid in solution suggested by the ith particle. Now intelligence of swarm suggests the value of center point.

Step 2 estimate the fitness constraints of every particle on given clustering condition the fitness constraints define as

$$F(i) = \frac{\sum_{x=1}^{k} \sum_{cij}^{p} (yp - pij)^2}{T_p} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4.5.2)$$

Where T_p is total number of data point proceeding for the clustering?

Step 3: the total number of iteration of clustering technique is maximized go to Step 7, if it is minimum go to next step

Step 4: The value of Pbest and Gbest stored in swarm search space and otherwise estimate with equation 4.5.1 and equation 4.5.2

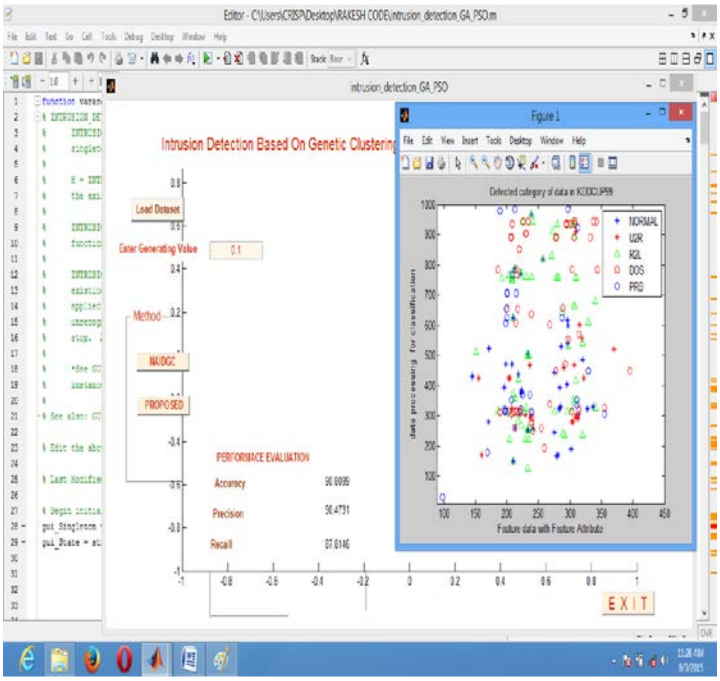
Step 5 minimized the eight value of particle W.

Step 6: If the value of Gbest is constant, unchanged for a number of iterations go to Step 7 otherwise go to Step 3.

Step 7: Use the k-means algorithm to finish clustering task. The clustering terminates when one of conditions meet according to fitness function.

V EXPERIMENTAL RESULT AND PERFORMANCE EVALUATION

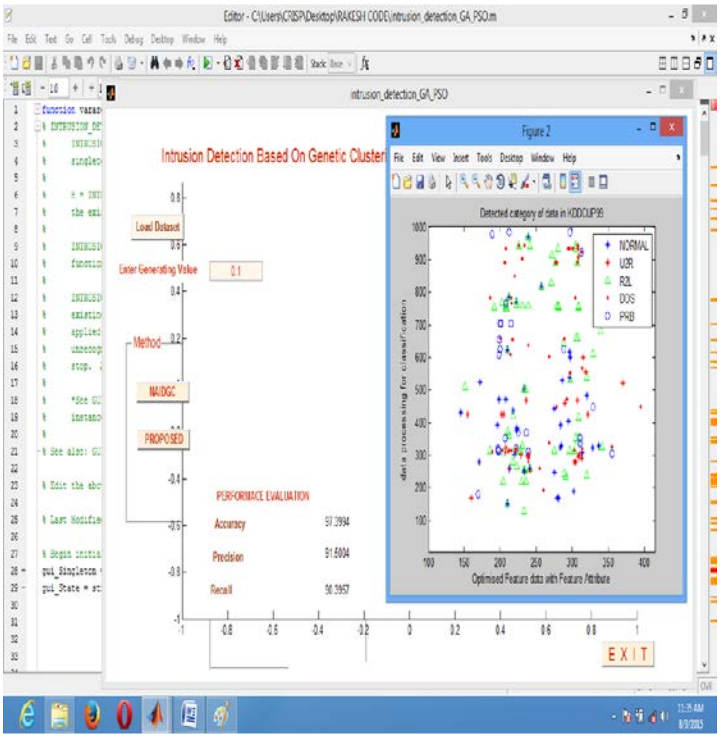
In this section we discuss the about our experimental result and comparative result analysis based on Genetic algorithm with k-means algorithm and our proposed method. On experimenting with different dataset, the number of normal/abnormal packets is being monitor. We have examined five different dataset in our experiment, with each having corresponding number of rejected or normal packets. In our conducted test the packets could either fall under normal packet type or in the category of attack (DOS, R2L.U2R.PROB). The rate of intrusion detection finds on given method of author and used KDDCUP99 dataset for analysis. We have supervised on data set with each 7000 instances of data under .the result of predicted normal and abnormal data is form of confusion matrix.


Editor - C:\Users\CRISP\Desktop\RAKESH CODE\intrusion_detection_GA_PSO.m
intrusion_detection_GA_PSO
Intrusion Detection Based On Genetic Clustering
Load Dataset
Method
NAIDGC
PROPOSED
Figure 1
Detected category of data in KDDCUP99
NORMAL
U2R
R2L
DOS
PRB
data processing for classification
Feature data with Feature Attribute
EXIT

Enter Generating Value: 0.1

PERFORMACE EVALUATION
Accuracy 90.0095
Precision 90.4731
Recall 87.8146

Figure 3: Shows that the intrusion detection system for generating value of 0.1 with using the method NAIDGC.


Editor - C:\Users\CRISP\Desktop\RAKESH CODE\intrusion_detection_GA_PSO.m
intrusion_detection_GA_PSO
Intrusion Detection Based On Genetic Cluster
Load Dataset
Method
NAIDGC
PROPOSED
Figure 2
Detected category of data in KDDCUP99
NORMAL
U2R
R2L
DOS
PRB
data processing for classification
Optimised Feature data with Feature Attribute
EXIT

Enter Generating Value: 0.1

PERFORMACE EVALUATION
Accuracy 97.3994
Precision 91.5004
Recall 90.3957

Figure 4: Shows that the intrusion detection system for generating value of 0.1 with using the PROPOSED method.

For the evaluation and performance measurement of proposed method for number plate recognition, used different set of number plate for both these method correlation and feature selection based method. The result evaluation parameter is recognition time, training time and recognition rate. All result shows in tabular form.

Input value	Method	Accuracy	Precision	Recall
0.1	NAIDGC	90.8099	90.4731	87.8146
	PROPOSED	97.3994	91.5004	90.3957

Table 1: Shows that the comparative results enter of the value 0.1 for the NAIDGC and PROPOSED method and finds the value of accuracy, precision and recall.

Input Value	Method Name	Accuracy	Precision	Recall
0.5	NAIDGC	92.569	92.2322	89.5737
	PROPOSED	99.1585	93.2595	92.1548

Table 2: Shows that the comparative results enter of the value 0.5 for the NAIDGC and PROPOSED method and finds the value of accuracy, precision and recall.

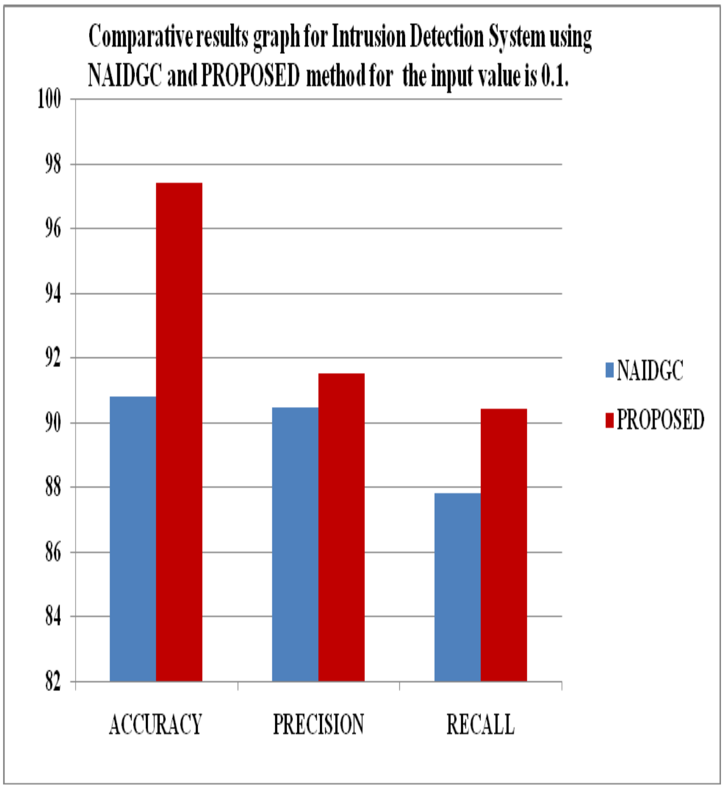


Figure 5: Shows that comparative result analysis of intrusion detection system and NAIDGC and PROPOSED method.

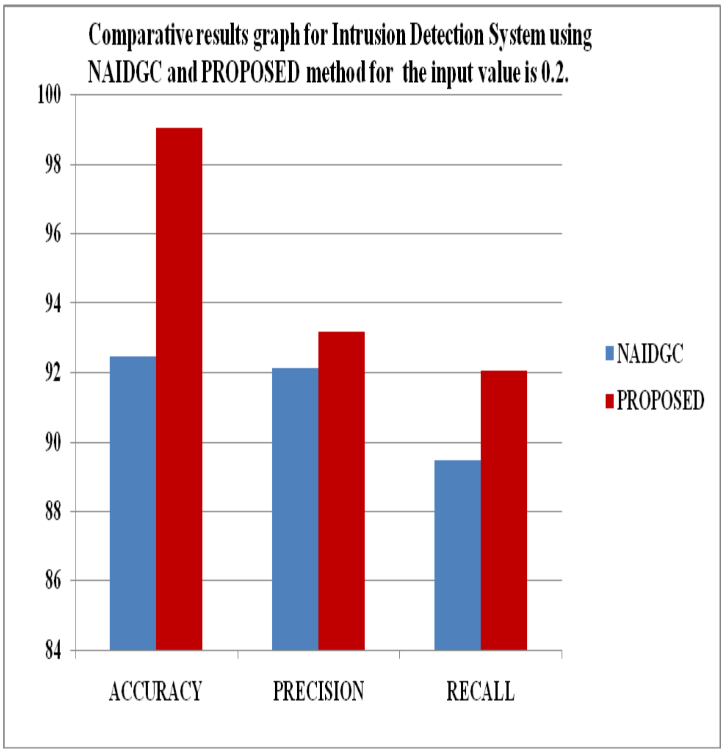


Figure 6: Shows that comparative result analysis of intrusion detection system and NAIDGC and PROPOSED method.

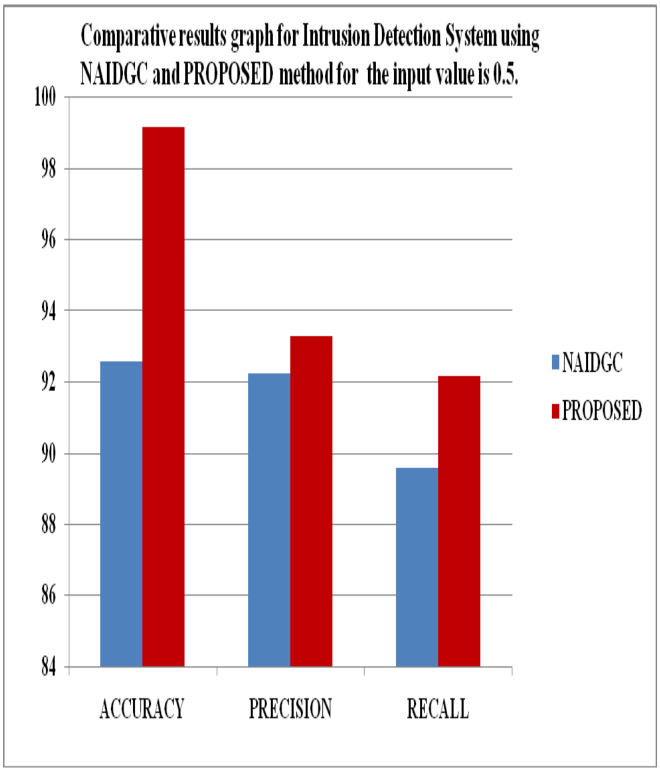


Figure 7: Shows that comparative result analysis of intrusion detection system and NAIDGC and PROPOSED method.

VI CONCLUSION AND FUTURE WORK

In this paper we proposed intrusion detection technique based on particle of swarm optimization along with k-means clustering technique. The modified clustering algorithm perform better clustering task instead of k-means-GA. Particle of swarm optimization control the iteration and also provide the facility of center point selection. The process of particle of swarm optimization reduces the iteration rate and increases the best value of global best value and improves the content of cluster. The proposed technique of k-means implied the process of similarity measure of traffic data of intrusion.

The swarm process overcomes the limitation of genetic algorithm and its control technique of k-means algorithm. The proposed clustering technique not grouped accurately the low capacity of data. Compared the results of the proposed method with some of commonly used clustering technique, the standard k-means-GA methods used as baseline methods are the big-bang and Local cluster group. The improvement of clustering has some limitation discuss here this limitation overcome in future. First the processing of swarm suffred from the data imbalancing problem during cluestering. in future reduces the data imbalancing in partilce of swarm optimization. Second is All the attribute of intruder data are very large so processing of data clustering is slow. And last one is For the improvement of data error rate used sampling technique for balancing of data.

REFERENCES

[1] Huiling Guo, Weichen, Fang Zhang"Research of Intrusion Detection based on genetic clustering algorithm" IEEE, 2012, Pp 1204-1207.

[2] Feng Du "An effective pattern matching algorithm for intrusion detection" ICCSE 2012, Pp 34-38.

[3] Prasanthi S, Sang-Hwa Chung and Won-Suk Kim "An enhanced TCP scheme for distinguishing non-congestion losses from packet reordering over wireless mash networks" IEEE, 2011, Pp 440-447.

[4] Xie Yong He Fubao, Zhang Yilai" A descending suffix tree based pattern matching algorithm for intrusion detection" IEEE, 2012, Pp 21-28.

[5] Li Chen "Using Genetic Algorithm for Network Intrusion Detection" Proc. the United States Department of Energy Cyber Security Group 2004 Training Conference, May 2004.

[6] Jain , Upendra "An Efficient intrusion detection based on Decision Tree Classifier using feature Reduction", International Journal of scientific and research Publications , Vol. 2, Jan. 2012.

[7] Dewan Md. Farid, Jerome Darmont, Nouria Harbi, Nguyen Huu Hoa, Mohammad Zahidur Rahman "Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification" 2008. Pp 1-5.

[8] Gary Stein, Bing Chen, Annie S. Wu, Kien A. Hua "Decision Tree Classifier For Network Intrusion Detection With GA-based Feature Selection" 2556. Pp 1-6.

[9] Ritu Ranjani Singh a, Prof. Neetesh Gupta "To Reduce the False Alarm in Intrusion Detection System using self Organizing Map" in International journal of Computer Science and its Applications.

[10] Z. Xue-qin, G. Chun-hua, L. Jia-jin "Intrusion detection system based on feature selection and support vector machine" Proc. First International Conference on Communications and Networking in China (ChinaCom '06), Oct. 2006.

[11] Zhang , M. Zulkernine "Network Intrusion Detection using Random Forests" School of Computing Queen's University, Kingston Ontario, 2006.

[12] John Zhong Lei and Ali Ghorbani "Network Intrusion Detection Using an Improved Competitive Learning Neural

Network" in Proceedings of the Second Annual Conference on Communication Networks and Services Research IEEE.

[13] P. Jongsuebsuk , N. Wattanapongsakorn and C. Charnsripinyo "Network Intrusion Detection with Fuzzy Genetic Algorithm for Unknown Attacks" in IEEE 2013.

[14] Deepak Rathore and Anurag Jain "a novel method for intrusion detection based on ecc and radial bias feed forword network" in IJACR, Vol. 2, Issue 3: July-Sep.: 2012.

[15] Wing w. Y. Ng, rocky k. C. Chang and daniel s. Yeung "dimensionality reduction for denial of service detection problems using rbfnn output sensitivity" in Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, 2-5 November 2003.

[16] 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.

[17] Mohammad Behdad, Luigi Barone, Mohammed Bennamoun and Tim French "Nature-Inspired Techniques in the Context of Fraud Detection" in ieee transactions on systems, man, and cybernetics part c: applications and reviews, vol. 42, no. 6, november 2012.

[18] Alberto Fernandez, Maria Jose del Jesus and Francisco Herrera "On the influence of an adaptive inference system in fuzzy rule based classification system for imbalanced data-sets" in Elsevier Ltd. All rights reserved 2009.

[19] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E.Vazquez "Anomaly-based network intrusion detection: Techniques, Systems and challenges" in Elsevier Ltd. All rights reserved 2008.

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