TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 2
ISSN: 2054 - 7420

Enhancing the Capability of IDS using Fuzzy Rough Classifier with Genetic Search Feature Reduction

[1]Ashalata Panigrahi and [2]Manas Ranjan Patra
Department of Computer Science & Engineering, SMIT, Berhampur
India
[1]ashalata.panigrahi@yahoo.com; [2]mrpatra12@gmail.com

ABSTRACT

Rapid expansion of computer network throughout the world has made security a crucial issue in a computing environment. In the recent past several cyber-attacks have corrupted data of many organizations and creating serious problems. Intrusion Detection System which is increasingly a key part of system defense is used to identify abnormal activities in a computer system. The success of an intrusion detection system depends on the selection of the appropriate features in detecting the intrusion activity. Experiments have been conducted using four classifier techniques, viz, Fuzzy NN, Fuzzy Rough NN, VQNN, Fuzzy Rough Ownership NN. We have studied the accuracy, recall, precision, false alarm rate, error rate of all the classifier techniques

Keywords : Genetic search, Fuzzy rough theory, Hybrid Intrusion Detection System, False alarm rate.

1. INTRODUCTION

Network Intrusion Detection Systems are increasingly demand today due to the continuous increase in number of network attacks in networks. The primary aim of Intrusion Detection System (IDS) is to protect the availability, confidentiality and integrity of critical networked information system. IDS may perform either misuse detection or anomaly detection and may be deployed as either a network-based system or a host-based system. Misuse detection systems are most widely used and they detect intruders with known patterns as: network packet, like source address, destination address, source and destination ports or even some key words of the payload of a packet. Anomaly detection systems identify deviations from normal behavior and alert to potential unknown or novel attacks without having any prior knowledge of them. They have the ability of detecting unknown attacks [1]. The performance of the current intrusion detection system can be improved by utilizing hybrid intrusion detection

DOI: 10.14738/tnc.22.97
Publication Date: 3rd April 2014
URL: http://dx.doi.org/10.14738/tnc.22.97

techniques. The performance of hybrid classification techniques is better than distinct classification methods**.** Yang et al. [2] proposed a wrapper-based feature selection algorithm to find most important features from the training dataset by using random mutation hill climbing method, and then employs linear support vector machines (SVM) to evaluate the selected subset-features. M.Govindarajan et al. [3] have proposed the hybrid architecture. It has proved that, the performance is better for distinct classification methods**.** Horng et al. [4] have proposed an IDS that combines a hierarchical clustering algorithm, a simple feature selection procedure, and the SVM technique. At first, the hierarchical clustering algorithm is used to generate training instances. Then, the simple feature selection procedure was applied to eliminate unimportant features from the training set. Finally, the obtained SVM model classifies the network traffic data. Tsang et al. [5] in their work defines attribute reduction with fuzzy rough sets and analyzes its structures in details and they have developed a formal definition of reduction with fuzzy rough sets.

2. PROPOSED HYBRID MODEL

Hybrid classifiers enhance the accuracy of classification. The objective of hybrid classifier is to merge few machine learning techniques. All the attributes may not contribute in the analysis process for identifying intrusive behavior. In this work we have used genetic search as feature reduction method. After attribute reduction the data are classified by four classification techniques namely, Fuzzy NN, Fuzzy Rough NN, Fuzzy Rough Ownership NN, and Vaguely Quantified NN. The performance of these classifiers is evaluated in terms of their detection accuracy, precision, recall, fitness value, false alarm rate, error rate.

3. METHODOLOGY

3.1 Hybridization of Rough Sets and Fuzzy sets

Fuzzy Set

A fuzzy set in X is an X → [0, 1] mapping, while a fuzzy relation in X is a fuzzy set in X × X. For all y in X, the R-forest of y is the fuzzy set R_y is defined by

$$R_y(x) = R(x,y) \tag{1}$$

For all x in X, if R is reflexive and symmetric fuzzy relation, that is

$$R(x,x) = 1 \tag{2}$$

$$R(x,y) = R(y,x) \tag{3}$$

holds for all x and y in X, then R is called a "fuzzy tolerance ratio.

Rough Set

Rough Set Theory is a mathematical tool to deal with imprecise and insufficient knowledge. In rough set theory, membership is not the primary concept unlike fuzzy sets. It deals with

inconsistency, uncertainty, and incompleteness by imposing an upper and a lower approximation to set membership.

Let (X, A) be an information system where X is the universe of discourse and A is a non-empty finite set of attributes such that a : X → V_a for every a ∈ A. The set V_a is called the "value set of a". Given B ⊆ A there is an associated equivalence relation R_B:

$$R_B = \{ (x,y) \in X^2 \mid \forall a \in B, a(x) = a(y) \} \quad (4)$$

If (x,y) ∈ R_B, then x and y are indiscernible by attributes from B. The equivalence classes of the B- indiscernibility relation are denoted by $[x]_B$.

Let A be a subset X. A can be approximated using the information contained within B by constructing the B-lower and B-upper approximations of A.

$$R_B{\downarrow}A = \{ x \in X \mid [x]_B \text{ subset } A \} \quad (5)$$

$$R_B{\uparrow}A = \{ x \in X \mid [x]_B \cap A \neq \emptyset \} \quad (6)$$

The tuple ($R_B{\downarrow}A$, $R_B{\uparrow}A$) is called a rough set.

3.2 Fuzzy Nearest Neighbor (FNN) Classification

The Fuzzy K-Nearest Neighbor (FNN) algorithm [6] was introduced to classify test objects based on their similarity to a given number K of neighbors, and these neighbors' membership degree to (crisp or fuzzy) class labels. For the purpose of (FNN), the extent C(y) to which an unclassified object y belongs to a class C is computed as:

$$C(y) = \sum_{x \in N} R(x,y)\, C(x) \quad (7)$$

where N is the set of object y's K nearest neighbors, and R(x,y) is the [0,1]-valued similarity of x and y.

```
FNN ( X, C, y, K )
   X, the training data set;  C,  the set of decision classes;
   y, the objects to be classified;
   K, the number of nearest neighbors.
   Begin
     N ← get Nearest Neighbors ( y, K )
     For each  C ∈  C  do
        C' ( y)  =  ∑x∈N R(x,y) C(x) C(x)
        Output  arg max ( C' (y) )
   End
```

Figure 1: The Fuzzy K-Nearest Neighbor (FNN) Algorithm

3.3 Fuzzy-Rough nearest Neighbor Algorithm (FRNN)

In FRNN algorithm the nearest neighbors are used to construct the fuzzy lower and upper approximations of decision classes, and test instances are classified based on their membership to these approximations. FRNN algorithm combines fuzzy-rough approximation with the

classical FNN approach [7]. The rationale behind the algorithm is that the lower and upper approximation of a decision class, calculated by means of the nearest neighbors of a test object y, provides good clues to predict the membership of the test object to that class. The algorithm is dependent on the choice of a fuzzy tolerance relation R. Given the set of conditional attributes A, the fuzzy tolerance relation R is defined by

$$R(x,y) = \min_{a \in A} R_a(x,y) \quad (8)$$

in which $R_a(x,y)$ is the degree to which objects x and y are similar for attribute a. Here we choose

$$R_a(x,y) = 1 - \frac{|a(x) - a(y)|}{|amax - amin|} \quad (8)$$

If (R↓C) (y) is high , it reflects that all of y's neighbours belong to C. A high value of (R↑C) means that at least one neighbor belongs to that class.

```
FRNN ( X, C, y )
    X, the training data set;  C, the set of decision classes;
    y, the objects to be classified;
Begin
      N ← get Nearest Neighbors ( y, K )
       τ←   0   ,  Class ←  Ø
       for each  C ∈  C  do
              if (( R↓C )(y) +   ( R↑C)(y)) / 2  ≥ τ then
              τ  ←   (( R↓C )(y) +   ( R↑C)(y)) / 2
              end
         end
  output Class
End
```

The Fuzzy Rough nearest Neighbor Algorithm

3.4 Vaguely Quantified Nearest Neighbors (VQNN)

VQNN depends only on the summation of the similarities of each class. It uses the linguistic quantifiers "most" and "some". Given a couple (Q_u , Q_l) of fuzzy quantifiers that represent "most" and "some" respectively, the lower and upper approximation of C. VQNN assigns a class to a target instance y as follows:

Determine NN , the K nearest neighbors of y.

Assign y to the class C for which ($R\downarrow^{Qu} C$)(y) + ($R\uparrow^{Ql} C$)(y) is maximal.

The upper and lower approximation of Vaguely Quantified rough sets are defined as

$$((R\downarrow^{Qu} C)(y)) = Q_u(\frac{\sum_{x \in X} \min(Rx,y),C(x)}{\sum_{x \in X} R(x,y)}) \quad (9)$$

$$((R\uparrow^{Ql} C)(y)) = Q_l(\frac{\sum_{x \in X} \min(Rx,y),C(x)}{\sum_{x \in X} R(x,y)}) \quad (10)$$

The fuzzy quantifiers Q_u , Q_l are increasing [0,1] → [0,1] mapping such that Q_u (1) = Q_l(1)=1 and Q_u (0) = Q_l(0)=0. This classifier based on rough set theory is capable of handling noise data.

3.5 Fuzzy Rough Ownership Algorithm

A fuzzy-Rough ownership is an attempt to handle both "fuzzy uncertainty" and "rough uncertainty". The fuzzy-rough ownership function τ_c of class C defined for an object y as,

$$\tau_c(y) = \sum_{x \in X} \frac{R(x,y)C(x)}{|X|} \quad (11)$$

The fuzzy relation R is determined by :

$$R(x,y) = \exp(-\sum_{a \in A} Ka(a(y) - a(x))2/(m - 1) \quad (12)$$

where m controls the weighting of the similarity and K_a is a parameter that decides the bandwidth of the membership. K_a is defined as

$$K_a = \frac{|X|}{2 \sum_{x \in X} || a(y) - a(x) ||2/(m-1)} \quad (13)$$

$\tau_c(y)$ is interpreted as the confidence with which y can be classified to class C. The algorithm does not use fuzzy lower or upper approximations to determine class membership.

```
FROWNN( X, A, C, y )
 X the training data set; A the set of conditional features;
 C the set of decision classes; y the object to be classified.
begin
      for each a ∈ A do
            Ka = |X| / (2 Σ_{x∈X} || a(y) − a(x) ||2/(m−1))
      end
N ← | x |
for each A ∈ C do τc(y) = 0
for each x ∈ N do
            d = Σ_{a∈A} Ka(a(y) − a(x))^2
     for each A ∈ C do
       τc(y) + = C(x).exp(− d^{1/(m - 1)} ) / | N |
     end
end
output    arg max τc(y)
              A ∈ C
```

The Fuzzy Rough Ownership Algorithm

4. EXPERIMENTAL SETUP

4.1 NSL-KDD Dataset

NSL- KDD is a dataset proposed by Tavallace et al. [8]. NSL-KDD data set is a reduced version of the original KDD 99 dataset. NSL-KDD consists of same features as KDD 99. The data set consists of 41 feature attributes out of which 38 are numeric and 3 are symbolic. Total number of records in the data set is 125973 out of which 67343 are normal and 58630 are attacks. The dataset contains different attack types that could be classified into four main categories namely, Denial of Service (DOS), Remote to Local (R2L) ,User to Root (U2R), and Probing.

Denial of Service (DOS) : A DOS attack is a type of attack in which an attacker overwhelms the victim host with a huge number of requests, example ping-of-death, smurf, etc.

Remote to Local Attacks (R2L) : A remote to local attack is an attack in which the intruder tries to exploit the system vulnerabilities in order to control the remote machine through the network as a local user, for example guessing password etc.

User to Root Attacks (U2R) : These attacks are exploitations in which the attacker starts off on the system with a normal user account and attempt to abuse vulnerabilities in the system in order to gain super user privileges, for example phf, etc.

Probing : Probing is an attack in which the attacker scans a machine or a networking device in order to determine weakness or vulnerabilities that may later be exploited so as to compromise the system, for example port-scan.

The percentage distribution of data under different categories is depicted in table I and Figure 2.

Table: 1 Data Distribution & Percentage of NSL-KDD Dataset

Class	Number of Records	% of occurrence
Normal	67343	53.48%
DOS	45927	36.45%
R2L	995	0.78%
Probes	11656	9.25%
U2R	52	0.04%
Total	125973	100%

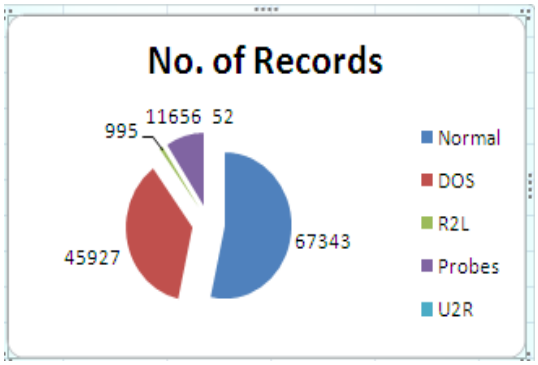

Figure 1: Distribution of Records

4.2 Feature Selection

Redundant and irrelevant attributes of dataset may lead to reduce detection accuracy. Effective input attributes selection from intrusion detection dataset is one of the important challenge for constructing high performance intrusion detection system. Feature selection is the process of finding a subset of features from the original dataset. The basic objective of feature selection method is to remove noise attributes and find important attributes which can represent data as a whole can improve the performance of intrusion detection and decrease the computation time. In this study Genetic search technique applied as feature selection method.

4.3 Genetic Search

Genetic algorithm is used as a search method. GA is based on principles of evolution and natural selection. Genetic search performs a global search. A Genetic algorithm mainly composed of three operators: reproduction, crossover, and mutation. Reproduction selects good string; crossover combines good strings to try to generate better offspring's; mutation alters a string locally to attempt to create a better string. In each generation, the population is evaluated and tested for termination of the algorithm. If the termination criterion is not satisfied, the population is operated upon by the three genetic algorithm operators and then re-evaluated. This procedure is continued until the termination criterion is met. [9]

Table 2: List of Selected Features

Feature Selection Method	No. of features selected	Feature Names
Genetic Search	16	Service, Flag, Src_bytes. Dst_bytes, Land, Urgent,Logged_in, Srv_count, Serror_rate, Srv_serror_rate, Rerror_rate, Same_srv_rate, Diff_srv_rate, Dst_host_count, Dst_host_same_srv_rate, Dst_host_srv_serror_rate.

4.4 Cross-Validation

The 10-fold cross-validation method is used to estimate the performance of different techniques. The entire dataset is divided into two different subsets, namely, training set and testing set. The training set is used to perform the analysis and the test set is used to validate

the analysis. In 10-Fold cross validation given dataset is partitioned into 10 subsets. From these 10 subsets 9 subsets are used to perform a training fold and a single subset is used as the testing data. The process is repeated 10 times such that each subset is used as a test subset once. The estimated accuracy is then the mean of the estimates for each of the classifiers.

4.5 Evaluation Measurement

The performance of IDS is measured and evaluated by the value of precision, accuracy, recall, false alarm rate.

TP (True Positive): The number of malicious records that are correctly identified.

TN (True Negative): The number of legitimate (not attacks) records that are correctly classified.

FP (False Positive): The number of records that are incorrectly identified as attacks though they are actually the legitimate ones.

FN (False Negative): The number of records that are incorrectly classified as legitimate activities though those are actually malicious.

Accuracy measure the probability that the algorithm can correctly predict positive and negative examples which are given by:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{14}$$

Precision is a measure of the accuracy provided that a specific class has been predicted which is given by:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{15}$$

Recall measure the probability that the algorithm can correctly predict positive examples which is given by:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{16}$$

False Alarm Rate is defined as the number of normal instances incorrectly labeled as intrusion divided by the total number of normal instances which is given by:

$$\text{False Alarm Rate} = \frac{FP}{FP+TN} \tag{17}$$

F- Value is the harmonic mean of Precision and Recall which measure the quality of classification which is given by

$$\text{F - Value} = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{18}$$

$$\text{Fitness Value} = \frac{TP}{TP+FP} * \frac{TN}{TN+FP} \tag{19}$$

$$\text{Error Rate} = 1 - \text{Accuracy} \tag{20}$$

Kappa Statistic: Kappa is a chance-corrected measure of agreement between the classification and true classes. Kappa statistic is used to access the accuracy of any particular measuring case. It is used to distinguish between the reliability of the data collected and their validity [10]. The value of kappa is less than or equal to 1. The value of 1 indicates perfect agreement.

Mean Absolute Error (MAE): The Mean Absolute Error measures the average magnitude of the errors.

Root Mean Squared Error (RMSE): Root Mean Squared Error is a quadratic scoring rule which measures the average magnitude of the errors. Measures the difference between forecast and corresponding observed values, are each squared and then averaged over the sample. Finally, the square root of the average is taken.

5. RESULTS AND DISCUSSIONS

Genetic search technique selected 16 attributes from 41 attributes from the data set. Four classification techniques namely, Fuzzy NN, Fuzzy Rough NN, VQNN, Fuzzy Ownership-NN were used for comparison. All the classification techniques are tested using 10-fod cross-validation. Table 3 depict the performance of four classifier techniques in terms of correctly classified instances and incorrectly classified instances. Fuzzy ownership NN technique identifies highest number of correctly classified instances and less number of incorrectly classified instances. Table 3 shows Fuzzy NN has highest mean absolute error rate.

Table : 3 Comparison of Different Parameters

Feature Reduction Method	Test Mode	Classifier Techniques	Correctly Classified Instances	Incorrectly Classified Instances	Kappa Statistic	MAE	RMSE
Genetic Search	10-Cross Validation	Fuzzy NN	94.606%	5.394%	0.9074	0.0216%	0.1469%
		Fuzzy Rough NN	99.0236%	0.9764%	0.9829	0.0109%	0.0768%
		VQNN	98.772%	1.228%	0.9785	0.0054%	0.0649%
		FROWNN	99.4173%	0.5565%	0.9903	0.0028%	0.0413%

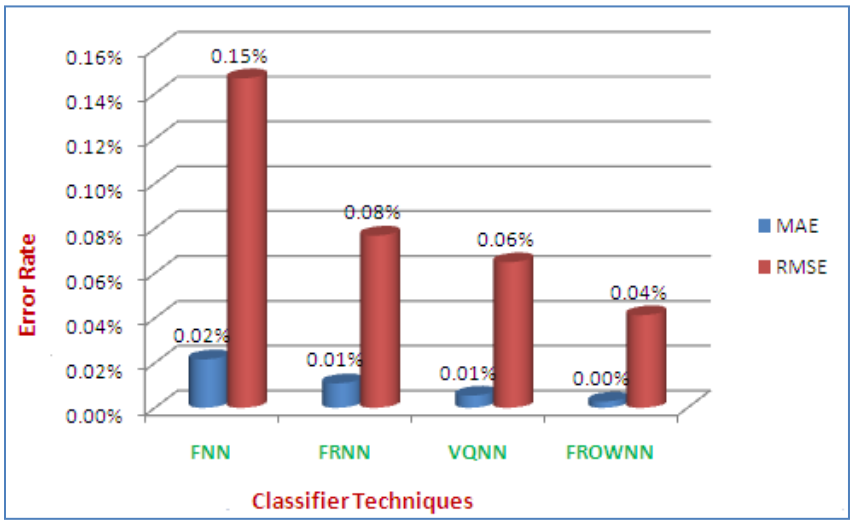

Figure 2: Mean Absolute Error vs Root Mean Squared Error

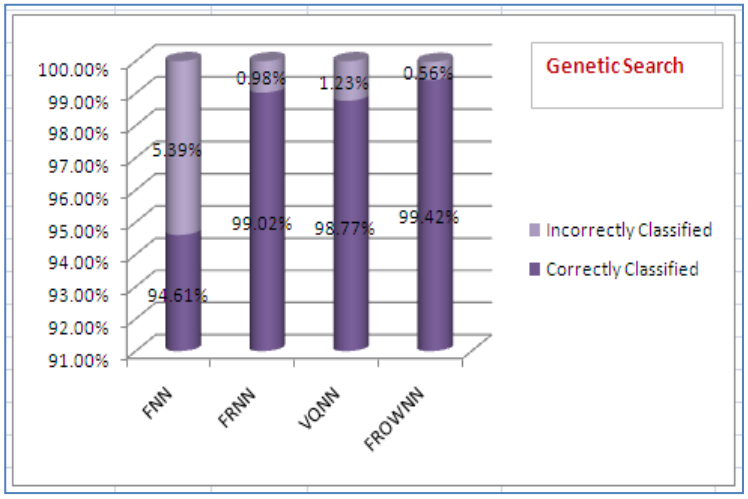

Figure 3 Correctly vs. Incorrectly Classified Instances

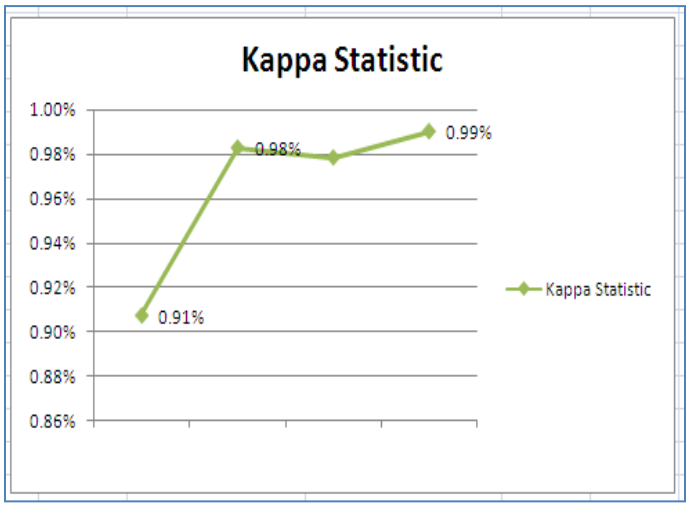

Figure 4: Comparison of Kappa Statistic of Four Classifier Techniques

Table : 4 Comparison of Accuracy, Precision , Recall, F-Value of Four Classifier Techniques

Feature Reduction Method	Test Mode	Classifier Techniques	Accuracy	Precision	Recall	F-Value
Genetic Search	10-Fold Cross-Validation	FNN	98.0845%	97.5875%	98.3149%	97.9499%
		VQNN	98.9069%	99.3382%	98.3063%	98.8198%
		FRNN	99.0736%	99.2458%	98.76002%	99.00234%
		FWNN	**99.5197%**	**99.5054%**	**99.5054%**	**99.5053%**

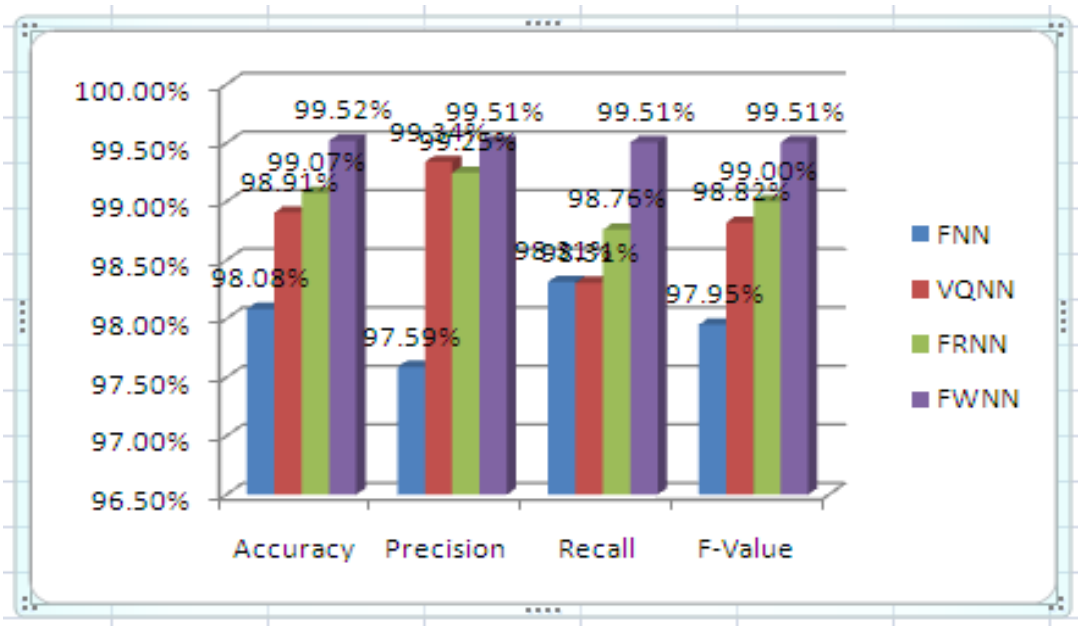

Figure 5 : Comparison of Accuracy, Precision , Recall, F-value

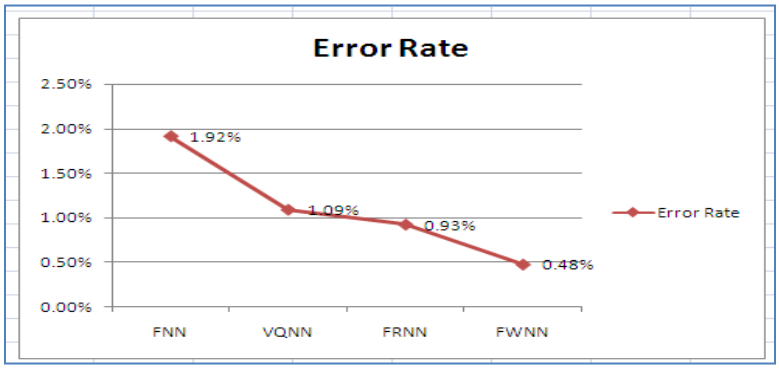

Fig. 6 Error Rate of Four classifier Techniques

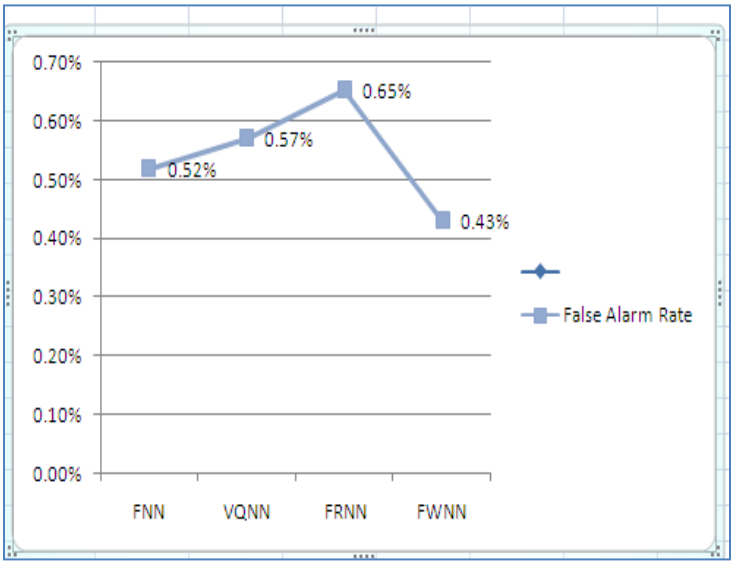

Fig. 7 Comparison of False alarm Rate of classifier techniques

6. CONCLUSION

Intrusion detection in large networks has been a challenging task. In this paper we have proposed hybrid model for intrusion detection. Experiments in our present work show that Fuzzy Ownership NN classifier technique provides best performance compared to other three classifier techniques. In future, we will explore other hybrid techniques with different feature selection methods and study their performance using different parameters.

REFERENCES

[1]. Zorana Bankovic, Dus an Stepanovic, Slobodan Bojanic, Octavio Nieto-Taladriz, " Improving network security using genetic algorithm approach" . Computers and Electrical Engineering, pp. 438-451, 2007.

[2]. Yang Li, J. L. Yang, Z. H. Tian, T. B. Lu, and C. Young, " Building lightweight intrusion detection system using wrapper-based feature selection mechanisms", Computer and Security, Vol. 28, pp. 466-475, September 2009.

[3]. M.Govindarjan , and R.M. Chandrasekran , " Intrusion Detection Using Neural Based Hybrid Classification Methods", Computer networks . 55(8): 1662-1671, 2011.

[4]. S.Horng, M.Su, Y.Chen, T.Kao, R.Chen, J.Lai and C.D.Perkasa, " A Novel Intrusion Detection System Based on Hierarchical Clustering and Support Vector Machines" , Expert Systems with Applications, vol.38, no.1, pp.306-313, 2011

[5]. C. C. Tsang, Degang Chen, and D. S. Yeung. Attribute Reduction using Fuzzy Rough Sets,. In IEEE Transaction on Fuzzy Systems, vol. 16, pp. 1130-1140, oct. 2008.

[6]. J.M. Keller , M. R. Gray, J. A. Givens : A Fuzzy K-Nearest Neighbour Algorithm, IEEE Trans. Systems Man Cybernet. 15(4), pp.580-585, 1985

[7]. Jesen,R. and Cornelis,C. “A new approach to fuzzy-rough nearest neoghbour classification”, LNAI 5306, Springer-Verlag, pp. 310-319 (2008).

[8]. M. Tavallaee, E Bagheri; Wei Lu; and A. Ghorbani, A detailed analysis of the KDD CUP 99 data set. Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009), 1-6 (2009)

[9]. D. Goldberg, Genetic Algorithm in Search, Optimization, and Machine Learning, Addison Wesley. 1989.

[10]. kappa at http://www.dmi.columbia.edu/homepages/chuangi/kappa