RESEARCH ARTICLE OPEN ACCESS

Feature Selection for Anomaly-Based Intrusion Detection using Rough Set Theory

Abhinav S. Raut*, Kavita R. Singh**
*Yeshwantrao Chavan college of engineering, Nagpur, India[1]rautabhinav@yahoo.com
**Yeshwantrao Chavan college of engineering, Nagpur, India[2]singhkavita19@yahoo.co.in

ABSTRACT
In network based intrusion detection system, the large number of irrelative, redundant characteristics of features increases the processing and saving time of data. For improving the anomaly detection accuracy, we are implementing important rough set based feature selection techniques, in which original data set is reduced to some essential feature subset based on certain defined criterion. First, we discuss the Entropy-Based feature reduction technique, in which it determines only those attributes that provides more gain in information. Secondly Open-loop and Closed-loop based feature selection technique. Open-loop based feature selection is centered on selection of features based on between-class separability criterion and Closed-loop based feature selection based on feature selection criterion based on predictor performance to select the feature subset. These algorithms are implemented on KDD CUP 99 data set to obtain the low dimensional feature subset.

Keywords- Intrusion detection; anomaly; feature selection; rough set

I. INTRODUCTION

In Network Intrusion Detection (NID), the system needs to handle massive amount of network data in real-time. Network data comprises a variety of features [1], where there exist many irrelevant and redundant features that will drops the intrusion detection accuracy. For developing Intrusion Detection System, namely IDS, the primary aim is to concentrate on the classification rate. To achieve high accuracy most of the detection system not efficient to address the existence of intrusion. All the features of data set are used to compare with the known intrusive patterns which are a very overlong detection technique. In general, there are mainly two approaches for detecting intrusion into computer networks: misuse detection and anomaly detection.Anomaly detection system controls the abnormality by measuring the distance between the anomalous activity and the normal activity based on a chosen threshold.

The primary goal of feature selection is to find the subset which maximizes performance *e.g.* accuracy of classification. By selecting the important features leads to simplification of problem, faster and accurate intrusion detection rate. Feature selection is a quarantined process which includes various techniques such as machine learning and statistical approaches. Some examples are Artificial Neural Network [2, 4], Support Vector Machine (SVM) [2, 5, 6]*etc*. All of these require IDS must be able to determine the proper subset of the most important characteristics to improve the detection accuracy and efficiency.

This paper defines an initial work in finding the optimal subset of features using Rough Set Theory (RST) described section IV. RST can be used as a tool to find the data dependencies and to reduce the number of features contained in the data set.In this paper, we are discussing the three main RST based feature selection techniques, namely Entropy-based, open loop and Closed loop. Entropy based feature selection algorithm select only those features that provides most Information Gain (IG). Open loop feature selection works on interclass separability criterion. The algorithm selects those features which have small within-class scatterand a large between-class scatter. Next in the closed loop feature selection, the algorithm obtains feature subset by using a predictor classification result depends on certain criteria.

This paper is organized as follows: Section II presents the theoretical foundation on basic process intrusion detection. Sections III describe the information about the feature selection operation. Basic concept of Rough set theory which is used for feature reduction approach is discussed in Section IV. Next section presents the important technique of feature selection which includes Entropy-based, open-loop and Closed-loop based feature selection. Section VI presents the detailed information about simulation result, experimental configuration andnext subsection contains the feature selection result along with conclusion and future work.

II. THEORETICAL FOUNDATION

Feature selection procedure used in this paper isshown in Figure 1:

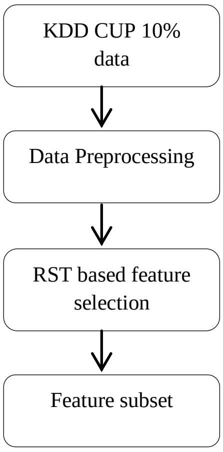


Figure 1. Process of feature selection method

The architecture of feature selection as shown in Figure 1 describes that how feature subset are extracted for processing data (here KDD CUP 99 data set). We use 10% of the original data set as input for our simulation process. Next step is preprocessing, in which feature values are preprocess to put them in single format as there are many feature values represented in different format. Now toeliminate irrelevant or redundant feature present in the data set we applied the RST based feature selection techniques to obtain a feature subset discussed in this paper.

III. FEATURE SELECTION

. Features may be irrelevant (having no effect on the processing performance) or relevant (having an impact on the processing performance). Because there is an abundance of noise, irrelevant or misleading feature in a data set, it is necessary to select appropriate features.Feature selection is an operation which obtains a subset from the feature set, is a one of the fundamental steps in the invention of the classifier. Next, it is utilized for the optimization process to obtain an optimal subset of features. In addition, the data gathered from the large network traffic contains a huge amount of data and it causes a prohibitively high overhead and often becomes a serious problem in IDS.

The effectiveness of features or feature subset is determined by both its relevancy and redundancy. A particular feature is considered to be relevant if it is predicated as decision feature, otherwise irrelevant also features considered redundant if it is highly associated with additional features. Hence, the exploration for a good feature subset implicates finding those features that are highly correlated with the decision features, but are uncorrelated with each other [7].

IV. ROUGH SET THEORY

As compared to classical set theory mentioned in [8], RST is based on the assumption that we have additional information about an element of a set. For example, data about patients suffering from a certain disease may contain information, *e.g.* body temperature, blood pressure, age, *etc*. As all patients defined by the same data, it is indiscernible i.e. similar in view of the data. This data constitutes a separate group called as elementary sets and called as basic knowledge about the patient. Any union of elementary sets is known as a crisp set, and other sets are stated to as rough (vague, imprecise). It is a tool to find out data dependency and to reduce the number of attributes contained in a dataset, requiring no additional information. Over long past year, RST has become a topic of interest for researchers and has been applied to large domain. Given a dataset with discretized attribute values, it is promising to find a subset (called as reduct) of the original attributes using RST that are the most enlightening; By considering the minimum information loss all other attributes can be removed from the dataset. From the dimensionality reduction standpoint, informative features are those that are most predictive of the class attribute.Each rough set has a boundary line; a pair of crisp set is associated called lower and upper approximation. Lower approximation contains all elements that certainly belong to set and upper approximation contains all elements which possibly belong. The boundary region is difference of upper and lower approximation.

IV.1. Basic Concepts

The basic concept of RST is very well explained by Pavel J. in [9],

Let,
U- Set of object called universe
$R \subseteq U * U$ - Indiscernibility relation representing our lack of knowledge about the elements of U.

For simplicity, we assume that R is an equivalence relation. Say that X is a subset of U. We want to characterize the set X according to R. We discuss some basic concept of RST:

Lower approximation - contain the items which certainly belong to X. It is defined as equation 1:

$$R_*(X) = \cup_{x \in U}\{R(x): R(x) \subseteq X\} \quad (1)$$

Upper approximation – contain the items which possibly belong to X and defined as equation 2:

$$R^*(X) = \cup_{x \in U}\{R(x): R(x)\ (X \neq \emptyset)\} \quad (2)$$

Boundary region – contain the set of item which can be classed as items belong to X or not and defined as equation 3:

$$BN_R(X) = R^*(X) - R_*(X) \quad (3)$$

V. OUGH SET BASED FEATURE SELECTIONTECHNIQUES

V.1. ntropy-Based Feature Reduction

A better technique for rough set feature selection is Entropy-Based reduction (EBR), developed from work carried out in [12]. This approach is based on entropy employed by machine learning techniques [11]. EBR is concerned with observing a dataset and determining those attributes that provides the most gain in information. The entropy of attribute A (which can take values a1,…am) regarding the conclusion *C* (of possible values c1,...cn) is defined as:

$$H(C|D) = -\sum_{j=1}^{m} p(a_j) \sum_{i=1}^{n} p(c_i|a_j) \log_2 p\,(c_i|a_j) \quad (4)$$

```
EBR(C,D)
C, the set of all conditional features;
D, the set of decision features.

(1) $R \leftarrow \{\}$
(2) do
(3)     $T\ R\leftarrow$
(4)     $\forall x \in (C - R)$
(5)         if $H(R \cup \{x\}) < H(T)$
(6)             $T \leftarrow R \cup \{x\}$
(7)     $R\ T\leftarrow$
(8) until $H(D\mid R) = H(D\mid C)$
(9) return $R$
```

Figure 2. Entropy based reduction

This technique can be extended to dealing with subsets of attributes instead of separate attribute only. By means of this entropy measure, the rough set based attribute reduction algorithm [12] can be modified to that presented in the figure 2. The significant advantages of this technique that it does not call for any threshold, the search for best feature subsets stopped when subsequent subset entropy is equal to that of the complete feature set. The resultant entropy of subset is zero for consistent data. It is important to note that any subset with entropy of 0 will also receive a corresponding rough set dependency of 1. Rough set reducts can be found by this technique, if the data are considered consistent. Taking to the exampRe, EBR first estimates the entropy of each individual attribute:

TABLE I: ENTROPY OF INDIVIDUAL ATTRIBUTES

Subset E	Entropy
{a}	1.1887219
{b}	0.75
{c}	0.9387219
{d}	0.75

As seen from Table 1, the {b} & {d} subsets have lowest entropy, so the algorithm selects attribute b because it is evaluated first and then it adds to current feature subset. Next step is to evaluate the entropy of all subsets containing b and one other attribute.

TABLE II: ENTROPY OF SUBSET

Subset	Entropy
{a, b}	0.5
{b, c}	0.59436095
{b, d}	0.0

Again from Table 2, Subset {b, d} is selected because it has lower entropy. Furthermore, stopping criterion has been met as the {b, d} subset has entropy value equal to the entire feature set $(H(D\langle b,d\rangle) = 0 = H(D|C))$. Hence the algorithm terminates and gives this feature subset. The whole dataset can now be reduced to these features only. The returned subset is rough set reduct because the resulting entropy is zero.

V.2. Open-Loop Feature Selection

Świniarski [13] discusses the two important feature selection techniques, open-loop based feature selection and closed-loop based feature selection. Detailed overview of thesetechniques, discusses in following section.

Before presenting these techniques we understand the basic concept of feature selection criteria. There are certain feature selection criteria used by a feature selection algorithm to select the appropriate features. Among such criteria some might satisfy the monotonicity property,

$$J_{feature}\left(X^{+}_{feature}\right) \geq J_{feature}\left(X_{feature}\right) \quad (5)$$

Here $X_{feature}$ and $X^{+}_{feature}$ denotes a feature subset and a larger feature subset containing $X_{feature}$ as a subset respectively. This implies that adding a feature to a given feature set will cause the value of the criterion stay the same or increase:

$$J_{feature}(\{x1\}) \leq J_{feature}(\{x1,x2\}) \leq J_{feature}(\{x1,x2,x3\}) \leq J_{feature}(\{x1,x2,\ldots,xn\})$$

(6)

Open-loop feature selection criteria [13] are built on information (similar to interclass separability) contained only in the data set alone. The main benefit of this is that they do not utilize a feedback from the predictor quality in the feature selection process.

Some feature selection criteria which are based on interclass separability have the origins in the idea of Fisher's linear transformation. Concentrating on this idea, a good feature (with a high discernibility power) will cause a small within-class scatter and a large between-class scatter.

Let us consider the original (total) data set T_{all} containing N_{all} cases (xi, c^{i}_{target}) with patterns x constituted using n-features and labeled by one target class c^{i}_{target} from all l possible classes. For a data set T_{all} we will represent the number of cases in each class c_i (i = 1, 2….l) by N_i ($\sum Ni = N_{total}$). One needs to specify a function which provides a larger value when a within-class scatter is smaller or a between-class scatter is larger, to term the feature selection criteria. The following criterion, based on interclass separability, may be defined:

$$J_{feature} = \frac{|S_b|}{|S_w|} = \frac{\det(S_b)}{\det(S_w)} \quad (7)$$

Where,

$$S_w = \sum_{i=1}^{l} \sum_{j=1, x^j \in c_i}^{N_i} (x^j - \mu_i)(x^j - \mu_i)^T$$

$$S_b = \sum_{i=1}^{l} N_i (\mu - \mu_i)(\mu - \mu_i)^T$$

Whereμdenotes the total data mean and the determinant | S_b |denotes a scalar representation of the between-class scatter matrix, and similarly, the determinant | S_w |denotes a scalar representation of the within-class scatter matrix.

V.3. Closed-Loop Feature Selection

In this method, we discuss the problem of specifying a feature selection criterion for a prediction task based on the original data set T_{all} containing N_{all} cases (x; target) formed by n-dimensional input patterns x (whose elements represent all features X) and targets of the turnout. Consider that the m-feature subset $X_{feature}$ belongs to X should to be evaluated based on the closed-loop type criterion. After reduction process, reduced data set $T_{feature}$ with patterns containing only m features from the subset $X_{feature}$ should be built. After that a predictor $PR_{feature}$ (we take here k-nearest neighbors) used for good quality feature evaluation. In the simplest form of feature selection a less expensive predictor can be used. After selecting a reduced feature set $X_{feature}$ for a feature set X, a predictor $PR_{feature}$ is decided for feature selection. For a good feature evaluation. It is necessary to define criterion $J_{PRfeature}$ for choosen predictor and an error counting method which will direct how to estimate performance by averaging of results. To evaluate the performance of predictor $PR_{feature,}$ an extracted feature data set is splits into training set N_{tra} and testing set $N_{test.}$ Each case (x^{i}_{f} ; $target^{i}$) of both sets contains a feature pattern x_f labeled by a target. The evaluation criteria can be set independently for prediction classification and prediction-regression.

We describe a feature selection criterion for classification task of predictor, a feature subset $T_{feature}$ case contains pairs (x_f ; c_{target}) of a feature input pattern x_f and a categorical-type target c_{target} taking a value of one of possible l classes c_i. The quality of classifier $PR_{feature,}$ Computed on test set $T_{feature}$;test with the help of N_{test} Patterns, can be measured using following criterion $J_{PRfeature}$:

$$J_{PR_{feature}} = J_{all\,miscl} = \frac{n_{all\,miscl}}{N_{test}}\, 100\% \quad (8)$$

where $n_{all\ miscl}$is the number of all misclassified patterns, and Ntest is the number of all the tested patterns. This basic criterion estimates the probability of an error (expressed in percent) by the relative frequency of an error.

VI. SIMULATION RESULT

Following section gives the detailed information about KDD CUP 99 data set used here for simulation process. It also describes the different types of attacks and shows the number of samples each attack type contains in 10% of original data set.

VI.1. Experimental Data

KDD CUP 1999 [14] dataset consumes large concentration in the evaluation of anomaly based intrusion detection method. This dataset is prescribed by Stolfo *et al.* [15] and put together based on data captured in DARPA'98 IDS evaluation program [16]. DARPA'98 is tcpdump data of 7 weeks of network traffic, which can be managed into 5 million connection records. The test data for around two weeks have 2 million connection records and training data set be made of approximately 4,900,000 single connection vectors; each one contains 41 and it is labeled as either normal or an attack along with particular attack type. The attacks are falling into one of the following four categories discuss by Srinivas M. et al [17].

VI.1.1. Denial of Service Attacks*:* A denial of service attack is a class of attacks in which an attacker denies legitimate user access to the system by making a particular computing or memory resource too busy. Examples are Apache2, Back, Land, Mailbomb, SYN Flood, and Ping of death, Process table, Smurf, Syslogd, Teardrop and Udpstorm.

VI.1.2. User to Root Attacks:User to root exploits are a class of attacks in which an attacker attempts to gain root access to the system by gaining access to a normal user account on the system. Examples are Buffer_overflow, Rootkit, loadmodule, Perl.

VI.1.3. Remote to User Attacks:A remote to user attack is a class of attacks in which an attacker sends packets to a machine over a network system, User is not legitimate who does not have an account on the machine; exploits some vulnerability to gain local access as a user of that machine. Examples are Ftp_write, Guess_passwd, Imap, Phf, Sendmail, Xlock and Xsnoop.

VI.1.4. Probing:Probing is a class of attacks in which an attacker scans a network of computers to collect data or find known vulnerabilities. An attacker has a map of machines and services that are available on a network which the attacker uses to look for exploits. Examples are Ipsweep, Mscan, Nmap, Saint, and Satan.

VI.2. Simulation Set

From the standard KDD CUP 99 data set, we have used given train and test set that consists of 10 % of entire data set. The Details about the number of samples belongs to each category of records present in data set is shown below [18].

TABLE III: NUMBER OF ATTACKS IN TRAININGKDDCUP 99 DATA SET

Data Set	Normal	Dos	U2R	R2L	Probe
Original KDD	972780	388337 0	50	1126	41102
10%KD D	97277	391458	52	1126	4107

KDD CUP 99 dataset classified into 5 classes (Normal, DOS, U2R, R2L, Probe) it contains 22 attack types and size of each attack as shown in Table 4.

TABLE IV: ATTACK TYPES AND SAMPLE SIZE IN 10% KDD CUP 99 DATA SET

	Attack Type (Number of Samples)
Normal	Normal (97277)
DOS	Smurf(280790), Neptune (107201),Back (2203), Teardrop (979), Pod (264), Land (21)
U2R	Buffer_overflow (30), Rootkit (10), loadmodule (9), Perl (3)
R2L	Warezclient (1020), Guess_passwd (53), Warezmaster (20), Imap (12), ftp_write (8), Multihop (7), Phf (4), Spy (2)
Probe	Satan (1589), Ipsweep (1247), Portsweep (1040), Nmap (231)

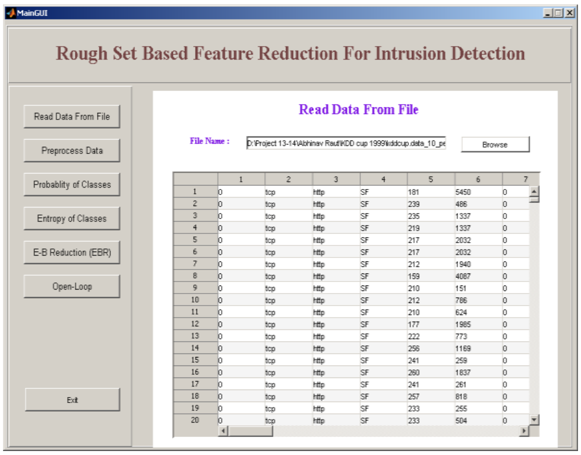
VI.3. Experimental configuration

In this paper, for simulation we have used Intel core i3 processor with 1.68 GHz of speed with 8 GB of RAM and windows XP operating system. The feature selection algorithms have been implemented in MATLAB R2011b.

VI.4. Feature selection results

This section shows the rough set based feature selection simulation results.

In first step of simulation process Figure 3 shows the basic data loading of KDD CUP 99 data set. There are 41 features and one decision feature positioned in last column. Next we preprocess the data set shown in Figure 4 by transforming all feature value in numerical format as many feature values are in characters.

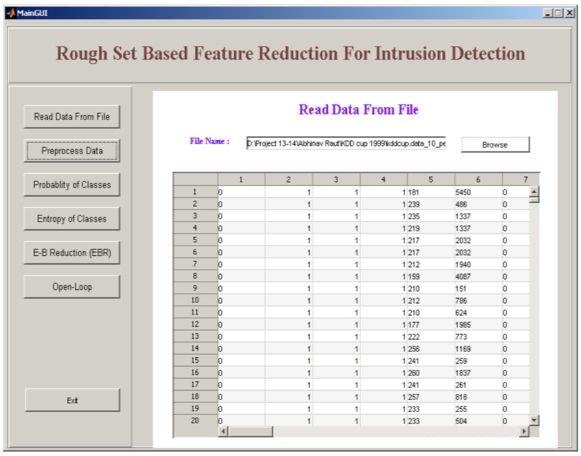


Read Data From File

File Name : D:\Project 13-14\Abhinav Raut\KDD cup 1999\kddcup.data_10_pe

	1	2	3	4	5	6	7
1	0	tcp	http	SF	181	5450	0
2	0	tcp	http	SF	239	486	0
3	0	tcp	http	SF	235	1337	0
4	0	tcp	http	SF	219	1337	0
5	0	tcp	http	SF	217	2032	0
6	0	tcp	http	SF	217	2032	0
7	0	tcp	http	SF	212	1940	0
8	0	tcp	http	SF	159	4087	0
9	0	tcp	http	SF	210	151	0
10	0	tcp	http	SF	212	786	0
11	0	tcp	http	SF	210	624	0
12	0	tcp	http	SF	177	1985	0
13	0	tcp	http	SF	222	773	0
14	0	tcp	http	SF	256	1169	0
15	0	tcp	http	SF	241	259	0
16	0	tcp	http	SF	260	1837	0
17	0	tcp	http	SF	241	261	0
18	0	tcp	http	SF	257	818	0
19	0	tcp	http	SF	233	255	0
20	0	tcp	http	SF	233	504	0

Figure 3. KDD CUP 1999 data set



Read Data From File

File Name : D:\Project 13-14\Abhinav Raut\KDD cup 1999\kddcup.data_10_pe

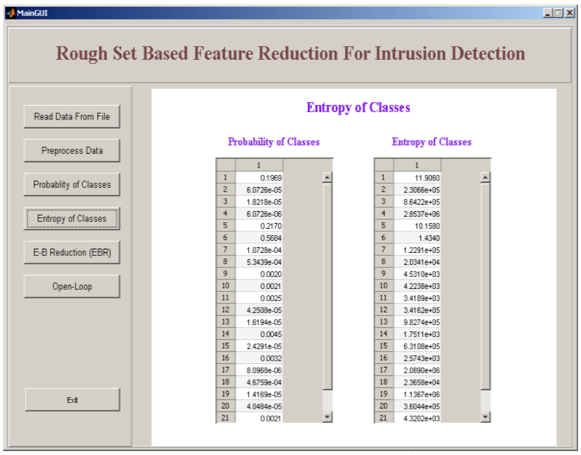
	1	2	3	4	5	6	7
1	0	1	1	1	181	5450	0
2	0	1	1	1	239	486	0
3	0	1	1	1	235	1337	0
4	0	1	1	1	219	1337	0
5	0	1	1	1	217	2032	0
6	0	1	1	1	217	2032	0
7	0	1	1	1	212	1940	0
8	0	1	1	1	159	4087	0
9	0	1	1	1	210	151	0
10	0	1	1	1	212	786	0
11	0	1	1	1	210	624	0
12	0	1	1	1	177	1985	0
13	0	1	1	1	222	773	0
14	0	1	1	1	256	1169	0
15	0	1	1	1	241	259	0
16	0	1	1	1	260	1837	0
17	0	1	1	1	241	261	0
18	0	1	1	1	257	818	0
19	0	1	1	1	233	255	0
20	0	1	1	1	233	504	0

Figure 4. Preprocessing of data

This section shows the simulation result of three rough set based feature selection techniques discussed previously.

VI.4.1. Entropy-Based Feature Selection:First, Entropy-based feature selection is based on the criteria of IG. It selects those features that provide most gain in information. For calculating information gain of features first we have to calculate probability and entropy of classes present in data set shown in Figure 5.

Figure 6 shows the information gain of every feature present in the data set. Now based on the entropy-based criteria of comparing the original data set with generating subset, it gives the feature subset when their stopping criteria meet.



Entropy of Classes

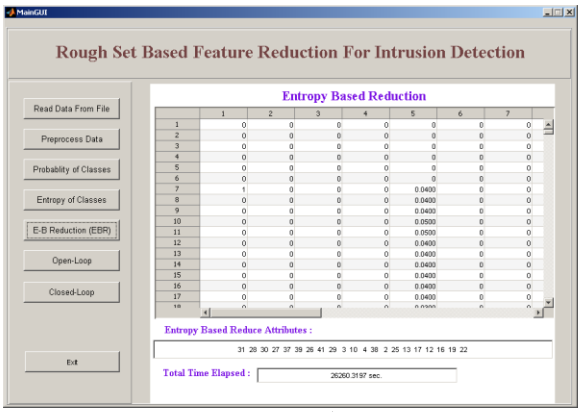
Probability of Classes

	1
1	0.1969
2	6.0726e-05
3	1.8218e-05
4	6.0726e-06
5	0.2170
6	0.5684
7	1.0728e-04
8	5.3439e-04
9	0.0020
10	0.0021
11	0.0025
12	4.2508e-05
13	1.6194e-05
14	0.0045
15	2.4291e-05
16	0.0032
17	8.0968e-06
18	4.6759e-04
19	1.4169e-05
20	4.0484e-05
21	0.0021

Entropy of Classes

	1
1	11.9060
2	2.3066e+05
3	8.6422e+05
4	2.8537e+06
5	10.1580
6	1.4340
7	1.2291e+05
8	2.0341e+04
9	4.5310e+03
10	4.2238e+03
11	3.4189e+03
12	3.4162e+05
13	9.8274e+05
14	1.7511e+03
15	6.3108e+05
16	2.5743e+03
17	2.0890e+06
18	2.3658e+04
19	1.1367e+06
20	3.6044e+05
21	4.3202e+03

Figure 5. Probability and Entropy of classes



Entropy Based Reduction

	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	1	0	0	0	0.0400	0	0
8	0	0	0	0	0.0400	0	0
9	0	0	0	0	0.0400	0	0
10	0	0	0	0	0.0500	0	0
11	0	0	0	0	0.0500	0	0
12	0	0	0	0	0.0400	0	0
13	0	0	0	0	0.0400	0	0
14	0	0	0	0	0.0400	0	0
15	0	0	0	0	0.0400	0	0
16	0	0	0	0	0.0400	0	0
17	0	0	0	0	0.0400	0	0

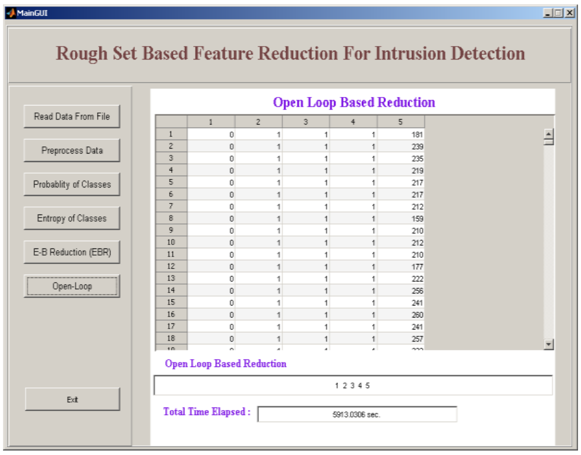
Entropy Based Reduce Attributes : 31 28 30 27 37 39 26 41 29 3 10 4 38 2 25 13 17 12 16 19 22

Total Time Elapsed : 26260.3197 sec.

Figure 6. Entropy-based feature selection

From the above simulation results we have got feature subset containing 21 features.

VI.4.2. Open-Loop Feature Selection:Open-loop feature selection criteria are basically work on the idea of information (like interclass separability) contained only in the data set. It selection criteria for features is based on the class scatter value. This algorithm generates the feature subset by selecting those features which have within-class scatter is smaller or a between-class scatter is larger.



Rough Set Based Feature Reduction For Intrusion Detection

Open Loop Based Reduction

	1	2	3	4	5
1	0	1	1	1	181
2	0	1	1	1	239
3	0	1	1	1	235
4	0	1	1	1	219
5	0	1	1	1	217
6	0	1	1	1	217
7	0	1	1	1	212
8	0	1	1	1	159
9	0	1	1	1	210
10	0	1	1	1	212
11	0	1	1	1	210
12	0	1	1	1	177
13	0	1	1	1	222
14	0	1	1	1	256
15	0	1	1	1	241
16	0	1	1	1	260
17	0	1	1	1	241
18	0	1	1	1	257

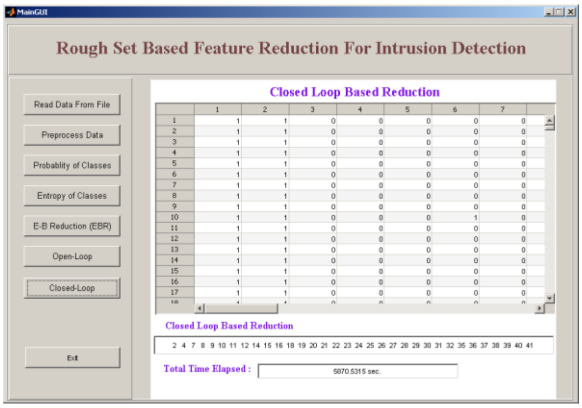
Open Loop Based Reduction: 1 2 3 4 5

Total Time Elapsed : 5913.0306 sec.

Figure 7. Open-loop feature selection

From simulation result obtained in figure 7 shows the subset which contains only five features which are in sequence. First five features are selected.

VI.4.3. Closed-Loop Feature Selection:A better simulation result obtained from closed loop feature selection technique shown in figure 8. A subset containing 33 features have been evaluated from original data set. A predictor (KNN Classifier) has been selected for calculating the quality of features. Estimation of feature goodness can be provided by defining certain criteria and an error counting method which will show how to estimate the performance through averaging of results.



Rough Set Based Feature Reduction For Intrusion Detection

Closed Loop Based Reduction

	1	2	3	4	5	6	7
1	1	1	0	0	0	0	0
2	1	1	0	0	0	0	0
3	1	1	0	0	0	0	0
4	1	1	0	0	0	0	0
5	1	1	0	0	0	0	0
6	1	1	0	0	0	0	0
7	1	1	0	0	0	0	0
8	1	1	0	0	0	0	0
9	1	1	0	0	0	0	0
10	1	1	0	0	0	1	0
11	1	1	0	0	0	0	0
12	1	1	0	0	0	0	0
13	1	1	0	0	0	0	0
14	1	1	0	0	0	0	0
15	1	1	0	0	0	0	0
16	1	1	0	0	0	0	0
17	1	1	0	0	0	0	0

Closed Loop Based Reduction: 2 4 7 8 9 10 11 12 14 15 16 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 35 36 37 38 39 40 41

Total Time Elapsed : 5870.5315 sec.

Figure 8. Closed loop feature selection

CONCLUSION

In this paper we presented the basics of rough set theory and also present the use of rough set theory for feature selection for simulation purpose. We have used freely available standard KDD CUP 99 dataset. We obtained the feature subset from the three important rough set based feature selection techniques: Entropy-based, Open loop and Closed loop. From simulation result it is very much clear that our proposed method have shown efficient results.

FUTURE WORK

We get the reduced feature subset by three important feature selection techniques, discuss previously in this paper. In future, we work on training of the classifier by these reduced feature subsets and test that classifier for anomaly and normal activity detection. Our next aim is to reduce the false positive of anomaly based intrusion detection.

REFERNCES

[1] Chen L., Shi L., Jiang Q. and Wang S., "Supervised Feature Selection for Dos Detection Problems Using a New Clustering Criterion," *Journal of Computational Information Systems*, 1983-1992

[2] Li J., Zhang G. and Gu G., "The Research and Implementation of Intelligent Intrusion Detection System Based on Artificial Neural Network," *IEEE Proceedings of the 3rd.International Conference on Machine Learning and Cybernetics*, pp. 3178-3182, 2004

[3] Sung A. and Mukkamala S., "Identifying Important Features for Intrusion Detection Using Support Vector Machines and Neural Networks," In *Proceedings of the 2003 Symposium on Applications and the Internet*, pp. 209-216, 2003

[4] Zhang C., Jiang J. and Kamel M., "Intrusion Detection using Hierarchical Neural Networks," *Pattern Recognition Letters,* Vol. 26, pp. 779-791, 2005

[5] Xu X. and Wang X., "An Adaptive Network Intrusion Detection Method Based on PCA and Support Vector Machines," In *proceedings of First International Conference on Advanced Data Mining and Applications ADMA*, Wuhan, China, Vol. 3584, pp. 696-703, 2005

[6] Gao H., Yang H. and Wang X., "Lecture Notes in Computer Science, Topic: Kernel PCA Based Network Intrusion Feature Extraction and Detection Using SVM,"Springer- Verlag, Berlin Heidelberg, New York, Vol. 3611, pp. 89-94, 2005

[7] Jensen R. and Shen Q., "Rough set based feature selection: A review," *Rough Computing: Theories, Technologies and Applications*, 2007

[8] Pawlak Z., "Rough sets," *International Journal of Computer and Information Science*, Vol. 11, No. 5, pp. 341-356, 1982

[9] Pavel J., "Using Rough Sets in Data Mining," *In Proceedings of the 12th Conference and Competition student EEICT*, Vol. 4, pp. 475-480, 2004

[10] Jensen R. and Shen Q., "Fuzzy-Rough Attribute Reduction with Application to Web Categorization,"*Fuzzy Sets and Systems*, Vol. 141, No. 3, pp. 469-485, 2004

[11] Quinlan J., "Programs for Machine Learning: The Morgan Kaufmann Series in Machine Learning,," *Morgan Kaufmann Publishers*, San Mateo, CA, 1993

[12] Chouchoulas A. and Shen Q., " Rough set-aided keyword reduction for text Categorisation," *Applied Artificial Intelligence*, Vol. 15, No. 9, pp. 843–873, 2001

[13] Świniarski and Roman W., "Rough sets methods in feature reduction and classification," *International Journal of Applied Mathematics and Computer Science*, Vol. 11, No. 3, pp. 565-582, 2001

[14] KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[15] Stolfo S., Fan W., Lee W., Prodromidis A. and Chan P., "Cost based modeling for fraud and intrusion detection: Results from the JAM project," *Proceedings, DARPA Information Survivability Conference and Exposition, DISCEX '00*, Vol. 2, pp. 130-144,2000

[16] Lippmann R., Fried D., Graf I., Haines J., Kendall K., McClung D., Weber D., Webster S., Wyschogrod D., Cunningham R. and Zissman M., "Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation," In *Proceeding ofDARPA Information Survivability Conference and Exposition, DISCEX '00*, Vol. 2, pp. 12-26, 2000

[17] Mukkamala S., Janoski G. and Sung A., "Intrusion detection using neural networks and support vector machines,"*In Proceedings of the IEEE,International Joint Conference onNeural Networks, IJCNN'02*, Vol. 2, pp. 1702-1707, 2002

[18] Shaik A., Nageswara and Chandulal, "Intrusion Detection System Methodologies Based on Data Analysis," *International Journal of Computer Applications*, Vol. 5, No.2,pp.0975–8887,2010