# Effective Anomaly based Intrusion Detection using Rough Set Theory and Support Vector Machine

Shailendra Kumar Shrivastava
Samarat Ashok Technological Institute, Vidisha, (M. P.), India

Preeti Jain
Samarat Ashok Technological Institute, Vidisha, (M. P.), India

## ABSTRACT

Intrusion detection system is used to discover illegitimate and unnecessary behavior at accessing or manipulating computer systems. Subsequently, these behaviors are checked as an attack or normal behavior. Intrusion detection systems aim to identify attacks with a high detection rate and a low false positive. Most of the earlier IDs make use of all the features in the packet to analyze and look for well-known intrusive models. Some of these features are unrelated and superfluous. The disadvantage of these methods is degrading the performance of IDs. The proposed Rough Set Support Vector Machine (RSSVM) approach is extensively decreases the computer resources like memory and CPU utilization which are required to identify an attack. The approach uses rough set to find out feature reducts sets. Then reduct sets are sent to SVM to train and test data. The results showed that the proposed approach gives better and robust representation of data.

## KEYWORDS

Intrusion Detection, Rough Set Theory, Support Vector Machine, Feature Selection.

## 1. INTRODUCTION

The idea of intrusion detection came out in 1984 by Fred Cohen [1], and intrusion detection model was projected by Denning in 1986[2]. This is particularly with the raise of attacks on computers and on networks in recent years improved and automated surveillance has become a necessary addition to computer security. Intrusion detection is the process of monitoring the events occurring in a computer system and analyzing them for signs of intrusions. Intrusions are termed as attempts to compromise the confidentiality, integrity or availability of a computer or network or to bypass its security mechanisms [3]. They are caused by invaders accessing a system from the Internet, by authorized users of the systems who attempt to gain extra advantages for which they are not legitimated and by authorized users who misuse the privileges given to them.

Intrusion Detection is categorized into two approaches: Misuse Based Intrusion Detection and Anomaly Based Intrusion Detection. Misuse detection stores the signatures of known attacks in the dataset and compares new instances with the stored signatures to discover attacks, while Anomaly Detection studies the normal behavior of the monitored system and then looks out for any difference in it for signs of intrusions. It is comprehensible that IDS based on misuse detection cannot discover new attacks and we have to include manually any new attack signature in the list of known patterns. Anomaly based IDs are able to discover any new attacks as any attack is guessed to be dissimilar from normal activity. However anomaly based IDS sometimes sets false alarms because it cannot discriminate properly between deviations due to authentic user’s activity and that of an intruder [4].

In this paper, we design a RSSVM (Rough Set Support Vector Machine) algorithm for intrusion detection based on feature selection and classification. Among from all the 41 features, only six key features are used for classification. There are five main classes of attacks i.e. Normal, DoS, Probe, R2L and U2R in the dataset.

In this paper, the Rough Set Theory and the Support Vector Machine is used as a tool to enhance the accuracy of the present intrusion detection algorithms. The rest of paper is organized as follows. Introduction of Intrusion Detection System is briefly described in section two, followed by types of networking attacks. In section three, a brief introduction to Rough Set Theory. Introduction of Support Vector Machine is given Section four. Section five explains the proposed algorithm. In section six experimental setup and results are presented. Finally in section seven concludes the remark in this line of work.

## 2. INTRUSION DETECTION SYSTEM

Intrusion is an active sequence of related events that deliberately try to cause harm. This refers to both successful and unsuccessful attempts [5]. Intrusion detection is the act of detecting such action. Intrusion Detection System (IDS) is a software tool used to detect unauthorized access to a computer system or network [6]. The important advantages of network security are control and visibility. Control is achieved by firewalls and access control lists in routers, among other things. Control should be an instantiation of the operational policy, but because of individual fault or other reasons, indistinctness exist in policy. Visibility allows one to acknowledge the capability when and where those indistinctness exist and provides the intelligence to modify control systems aptly [5].

There are basically three main types of IDS namely Network based Intrusion Detection System (NIDS), Host based Intrusion Detection System (HIDS), and a hybrid of the two. NIDS analyzes individual packets flowing through a network. The malicious packets that may be overlooked by a firewall simplistic filtering rules can be detected in the NIDS. On the other hand in HIDS, the IDS examine the activity of each individual computer or host. HIDS operates on information collected from within an individual computer system such as operating system. A hybrid IDS combines a HIDS, which monitors events occurring in the host system, and a NIDS, which monitors network traffic.

Specifically, the four broad classes of attack type defined in IDS [7] as: DoS, Probe, R2L and U2R.

Denial-of-Service (DoS): These are attacks designed to make some service accessible through the network unavailable to legitimate users.

Probe: A Probe is reconnaissance attack designed to uncover information about the network, which can be exploited by another attack.

Remote-to-Local (R2L): This is where an attacker with no privileges to access a private network attempts to gain access to that network from outside, e.g. over the internet.

User-to-Root (U2R): The attacker has a legitimate user account on the target network. However, the attack is designed to escalate his privileges so that one can perform unauthorized actions on the network.

Table 1 has the attack types and their respective classes.

**Table 1: Attack types and their respective classes**

| # | Attack Class | Attack Type |
|---|---|---|
| 1 | Normal | Normal |
| 2 | DoS | apache2, back, land, mailbomb, neptune , pod, processtable, smurf, teardrop, udpstrom |
| 3 | Probe | ipsweep, mscan, nmap, portsweep, saint, satan |
| 4 | R2L | ftp_write, guess_passwd, imap, multihop, named, phf, sendmail, spy, snmpgetattack, snmpguess , warezclient, warezmaster, worm, xlock, xsnoop |
| 5 | U2R | buffer_overflow, httptunnel, loadmodule, perl, ps, rootkit, sqlattack, xtern |

## 3. ROUGH SET THEORY

In research environments, output data are frequently indistinguishable, imperfect, and incompatible. Opportunely, the theory of Rough Sets has especially intended to handle these types of situations. In Rough Sets every object of significance is related with a piece of information representing relative relationship. This information is used to derive data classification and is the key issue of any reasoning, learning, and decision making [8].

Rough Set [9] is one of data-mining technique which decreases the features from large numbers of data. Knowledge, obtained from human or machine practice, is represented as a set of patterns describing features of two types, condition and decision [9]. Rough Set theory deals with unpredictability, vagueness and incompleteness by striking an upper and a lower approximation to set membership. It has effectively used as a selection method to determine data dependencies and find out all probable feature reduct subsets and eliminate redundant information. Therefore, a reduct is a nominal subset of features with the same capability of objects classification as the entire set of features [8, 10]. The definitions given below shows the reduct derivation for rough set theory.

### Definition 1:

Knowledge is acted for by means of a table called an Information System given by $S =<U,A,V,f>$; where $U = \{x1, x2, \ldots,xn\}$ is a finite set of objects of the universe ($n$ is the number of objects); $A$ is a non empty finite set of features, $A=\{a1, a2, \ldots, am\}$; $V= U_{a \in A} V_a$ and $V_a$ is a domain of feature $a$; $f:U\times A\rightarrow A$ is a total function such that $f(x, a) \in V_a$ for each $a \in A, x \in U$. If the features in $A$ can be separated into condition set $C$ and decision feature set $D$; i.e. $A=CUD$ and $C\cap D=\Phi$. The information system $A$ is known as decision system or decision table.

### Definition 2:

Every $B\subseteq A$ yields an equality relation up to indiscernibility, $IND_A (B) \subseteq (U\times U)$, given by: $IND_A(B) = \{(x,x') : \ a\in B \ a(x) = (x')\}$ a reduct of $A$ is the least $B\subseteq A$ that is equivalent to $A$ up to indiscernibility. i.e., $IND_A (B) = IND_A (A)$.

## 4. SUPPORT VECTOR MACHINE

The Support Vector Machine (SVM) was primary proposed by Vapnik [11] and has since attracted a high level of interest in the machine learning research area. Several recent studies have reported that the SVM generally are capable of delivering higher performance in terms of classification accuracy than the other data classification algorithms. Occasionally, we like to classify data into two sets. There are numbers of techniques for classification for instance the Artificial Neural Network and Fuzzy Logics. When applied correctly, these techniques provide satisfactory results. Most significant benefit of SVM is simple to use and high accuracy rate and its error rate is minimum.

Basic input data design and output data areas are given as follows:

$$( x_i , y_i ),\ldots, (x_n ,y_n ), x\in R^m ,y\in\{+1,-1\}$$

where $( x_i , y_i ),\ldots, (x_n ,y_n )$ are a train data, n is the numbers of samples, m is the inputs vector, and y fits in to category of +1 or -1 respectively. On the problem of linear, a hyper plan can divided into the two categories. The hyper plan formula is:

$$(w \, . \, x) + b =0$$

The category formulae are:

$$(w.x) + b \geq \text{if } y_i = +1$$

$$(w.x) + b \leq \text{if } y_i = -1$$

SVM is a statistical learning theory based on machine learning methods. SVM is extensively used in the field of bioinformatics, data mining, image recognition, text categorization, hand-written digit recognition. The SVM was designed to solve binary classification problems [13]. In this section, we create an SVM model for classification. Though an intrusion behavior occurs, SVM will discover the intrusion. There are two main causes for choosing the SVMs method for intrusion detection. The first is speed because real time performance is one of key importance to intrusion detection systems, and any classifier that can potentially exceed ANN is worth considering. Another cause is scalability: SVMs are comparatively insensitive to the number of data points and the classification intricacy does not depend on the dimensionality of the feature space [13].

## 5. PROPOSED APPROACH

Proposed Intrusion detection technique is represented in figure 1, which contains following steps:

First, Data preprocessing is done to convert the non-numeric value to numeric value, and then feature selection method is used to uncover valuable features using Johnson’s and genetic algorithm of RST. Lastly, the reduct sets are sent to Support Vector Machine for classification of data and to find accuracy of system.

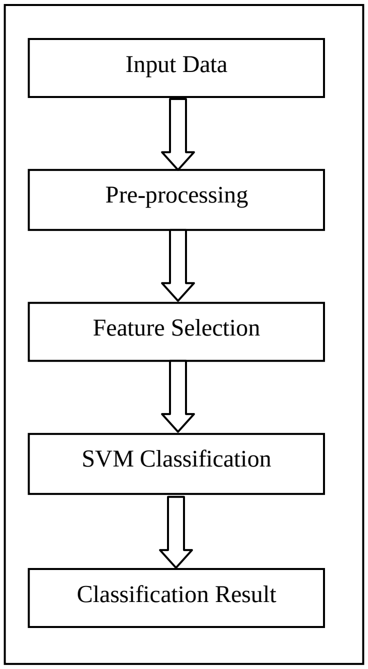

**Figure 1: Flowchart of Proposed RSSVM Approach for Intrusion Detection.**

**Table 2: KDD Cup’99 Data Set 41 Features**

| # | Feature Name | # | Feature Name | # | Feature Name |
|---|---|---|---|---|---|
| 1 | duration | 15 | su_attempted | 29 | same_srv_rate |
| 2 | protocol_type | 16 | num_root | 30 | diff_srv_rate |
| 3 | service | 17 | num_file_creations | 31 | srv_diff_host_rate |
| 4 | flag | 18 | num_shells | 32 | dst_host_count |
| 5 | src_bytes | 19 | num_access_files | 33 | dst_host_srv_count |
| 6 | dst_bytes | 20 | num_outbound_cmds | 34 | dst_host_same_srv_rate |
| 7 | land | 21 | is_hot_login | 35 | dst_host_diff_srv_rate |
| 8 | wrong_fragment | 22 | is_guest_login | 36 | dst_host_same_src_port_rate |
| 9 | urgent | 23 | Count | 37 | dst_host_srv_diff_host_rate |
| 10 | hot | 24 | srv_count | 38 | dst_host_serror_rate |
| 11 | num_failed_logins | 25 | serror_rate | 39 | dst_host_srv_serror_rate |
| 12 | logged_in | 26 | srv_serror_rate | 40 | dst_host_rerror_rate |
| 13 | num_compromised | 27 | rerror_rate | 41 | dst_host_srv_rerror_rate |
| 14 | root_shell | 28 | srv_rerror_rate | | |

## 5.1 Data Pre-Processing

KDD CUP’99 data set is used as a database to test the system performance, which is the data set used for the third International Knowledge Discovery and Data mining tools competition, which was held in conjunction with KDD-99, the fifth international conference on Knowledge Discovery and Data Mining [7]. The information obtained by KDD Cup’99 can be a combination of many system calls. A system call is a text base record. Every text record in the database has 41 features as listed in table 2. Since SVM classification uses only numerical data for testing and training, so text features are needed to be converted into numerical values. Therefore, we have assumed some numerical values for different text features, like ‘protocol_type’ feature ‘tcp’ as 3, ‘udp’ as 7, and ‘icmp’ as 9 etc. as shown in table 3.

**Table 3: Transformation Table for translating the Text data to numeric data in KDD cup'99 Data Set.**

| Type | Class | No. | Type | Class | No. |
|---|---|---|---|---|---|
| Attack/ Normal | Attack | 1 | Service | imap4 | 23 |
| | Normal | 0 | | iso_tsap | 24 |
| Protocol Type | TCP | 3 | | Klogin | 25 |
| | UDP | 7 | | Kshell | 26 |
| | ICMP | 9 | | Ldap | 27 |
| Flag | OTH | 1 | | Link | 28 |
| | REJ | 2 | | Login | 29 |
| | RSTO | 3 | | Mtp | 30 |
| | RSTOS0 | 4 | | Name | 31 |
| | RSTR | 5 | | netbios_dgm | 32 |
| | S0 | 6 | | netbios_ns | 33 |
| | S1 | 7 | | netbios_ssn | 34 |
| | S2 | 8 | | Netstat | 35 |
| | S3 | 9 | | Nnsp | 36 |
| | SF | 10 | | nntp | 37 |
| | SH | 11 | | telnet | 38 |
| Service | Auth | 1 | | Time | 39 |
| | Bgp | 2 | | Uucp | 40 |
| | Courier | 3 | | uucp_path | 41 |
| | csnet_ns | 4 | | Vmnet | 42 |
| | Ctf | 5 | | Whois | 43 |
| | Daytime | 6 | | Z39_50 | 44 |
| | Discard | 7 | | ntp_u | 45 |
| | Domain | 8 | | Other | 46 |
| | domain_u | 9 | | pop_2 | 47 |
| | Echo | 10 | | pop_3 | 48 |
| | eco_i | 11 | | Printer | 49 |
| | ecr_i | 12 | | Private | 50 |
| | Efs | 13 | | remote_job | 51 |
| | Exec | 14 | | Rje | 52 |
| | Finger | 15 | | Shell | 53 |
| | ftp | 16 | | Smtp | 54 |
| | ftp_data | 17 | | sql_net | 55 |
| | Gopher | 18 | | Ssh | 56 |
| | Hostnames | 19 | | Sunrpc | 57 |
| | http | 20 | | Supdup | 58 |
| | http_443 | 21 | | Systat | 59 |
| | IRC | 22 | | X11 | 60 |

## 5.2 Feature Selection using Rough set Theory

For feature selection from KDD Cup'99 dataset, ROSETTA a data mining tool, invented by Ohrn[10], is used. The algorithm used by ROSETTA library supports two categories of discrenibility:

1). Full: In this category of discrenibility, reducts are selected relative to the system as a whole.

2). Objects: In this category of discrenibility, reducts are selected relative to a single object.

There are four algorithms, namely Johnson's, Genetic, Holte's and Manual reducer. Johnson's algorithm [12] uses a simple greedy algorithm to compute single reduct only; on the other hand genetic algorithm is used to select the minimum hitting sets [11, 14]. Holte's algorithm gives singleton reduct and Manual reducer algorithm reduct are depends on the features selected by human being.

In this paper, to find the reduct sets, we have used Johnson's and Genetic Algorithm. Former has given the single reduct of six features out of 41 features, however the Genetic Algorithm gives the 39 reduct set, out of which we have used only 4 reduct set of six features. All feature subsets are listed in table 4 which are used to classify the data.

**Table 4: Different features sets used in the Classification of Attack.**

| # | Method Name | Features used | Features Set |
|---|---|---|---|
| 1 | Proposed Johnson | 6 | 5, 6, 23, 24, 33, 36 |
| 2 | Proposed Genetic A | 6 | 5, 6, 24, 29, 33, 36 |
| 3 | Proposed Genetic B | 6 | 5, 6, 23, 24, 33, 36 |
| 4 | Proposed Genetic C | 6 | 5, 6, 23, 29, 33, 36 |
| 5 | Proposed Genetic D | 6 | 5, 6, 23, 30, 33, 36 |

## 5.3 Intrusion Evaluation by Support Vector Machine

We have divided the behavior of user into two classes namely attack and normal, where the behavior of user is the collection of different attacks belonging to the five classes as explained in table 1. The aim of our SVM experiment is to differentiate between normal and attack behavior of user. In our experiments normal data are classified as -1 and all attacks are classified as +1.We have used the LIBSVM 3.0 tool [15] for classification. There are four kernel function namely, linear function, polynomial function, radial basis function, and sigmoid function. It is used RBF kernel as a default function.

## 6. EXPERIMENTS AND DISCUSSIONS

We run our experiments on a system with 2.00 GHz, Core ™ 2 Duo processor and 1.99 GB RAM running Windows XP. All the processing is done using MATLAB® 2008b. The database is gathered from The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99, The Fifth International Conference on Knowledge Discovery and Data Mining. The database has seven week of training data and two weeks of testing data. We have selected some data for training and testing randomly. We have performed the experiment on 10K, 20K, 30K, 40K processes respectively.

The objective of this research work is to enhance the accuracy of the system. We have used the Johnson's and Genetic Algorithm of RST to find the reduct set. SVM is used for classification by using the suitable kernel of SVM classification. System performance is evaluated using the test data.

In SVM classification, first we have done the pre-processing step by converting the text values of 41 features into numeric values using the transformation table 3. By applying the feature selection algorithm Johnson and Genetic algorithm, we get the reduct set. The resulting selected features are only 6 (Table 4), it reduces the data by 83%. Then data is forwarded to LIBSVM tool [15]. The output of SVM is 1or -1. If the output is 1, it indicates the intrusion behavior. If the output is -1, then it indicates normal behavior.

Following fundamental formulas are used to estimate the performance of the system: accuracy rate (AR) and false positive rate (FPR).

$$\textbf{Accuracy Rate} = \frac{\text{Total number of correct classified process}}{\text{Total number of processes}} \text{X } 100\ \%$$

$$\textbf{False Positive Rate} = \frac{\text{Total number of misclassified process}}{\text{Total number of normal processes}} \text{X } 100\ \%$$

Our experiments have tested the accuracy rate, false positive rate and attack detection rate by using 41 features to SVM, 29 features to SVM, five proposed set of 6 Features to SVM (one using Johnson and four using genetic algorithm). The accuracy of proposed system is 95.98 % and false positive rate is 7.52%.

Table 5 shows the frequency of attacks accordingly their respective class occurred in the data set. Table 6 and Table 7 show the accuracy rate and false positive rate respectively. Figure 2, Figure 3 shows the comparative graphical analysis of accuracy rate using Johnson and genetic algorithm with 41 features and 29 features. Figure 4, Figure 5 shows the comparative graphical analysis of false positive rate using Johnson and genetic algorithm with 41 features and 29 features.

**Table 5: Frequency of Attacks in the data set accordingly to their respective classes.**

| Attack Class | Data Set | | | |
|---|---|---|---|---|
| | 10 K | 20 K | 30K | 40K |
| Normal | 7829 | 2055 | 16549 | 12768 |
| DoS | 541 | 17520 | 7168 | 21622 |
| Probe | 618 | 45 | 3114 | 2509 |
| R2L | 1004 | 332 | 3135 | 3068 |
| U2R | 8 | 48 | 34 | 33 |

**Table 6: Comparison of Accuracy Rate**

| No. of Features | Data Set | | | |
|---|---|---|---|---|
| | 10K | 20K | 30K | 40K |
| 41 Feature | 85.23% | 85.32% | 86.48% | 86.79% |
| 29 Features | 89.07% | 89.10% | 89.36% | 89.13% |
| **Proposed** | | | | |
| 6 Features | 94.11% | 95.25% | 95.59% | 95.98% |
| 6 Features | 93.51% | 93.24% | 93.73% | 94.67% |
| 6 Features | 94.11% | 95.25% | 95.59% | 95.98% |
| 6 Features | 93.23% | 93.58% | 93.75% | 95.36% |
| 6 Features | 93.59% | 93.56% | 93.78% | 95.04% |

**Table 7: Comparison of False Positive Rate**

| No. of Features | Data Set | | | |
|---|---|---|---|---|
| | 10K | 20K | 30K | 40K |
| 41 Feature | 30.01% | 28.75% | 24.51% | 29.97% |
| 29 Features | 12.46% | 16.87% | 19.29% | 13.27% |
| **Proposed** | | | | |
| 6 Features | 7.70% | 7.99% | 8.76% | 7.52% |
| 6 Features | 9.79% | 12.56% | 11.38% | 8.29% |
| 6 Features | 7.70% | 7.99% | 8.76% | 7.52% |
| 6 Features | 8.64% | 11.32% | 13.45% | 8.68% |
| 6 Features | 8.19% | 14.70% | 11.28% | 10.87% |

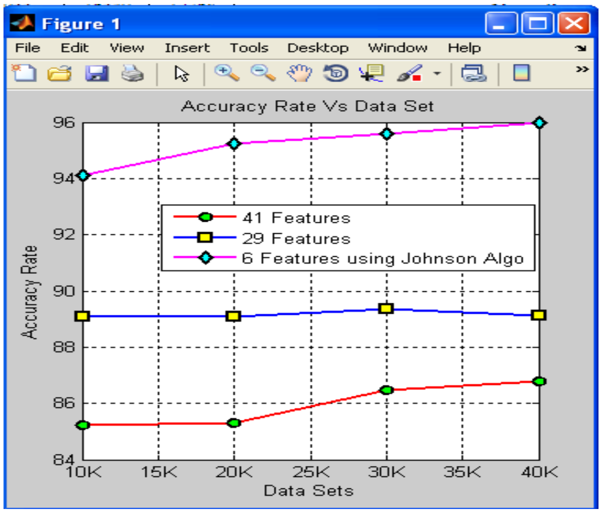

**Figure 2: Analysis of classification Accuracy Rate of ID using three feature sets that are 41 features, 29 features RST [16] and 6 features Johnson's Algorithm**

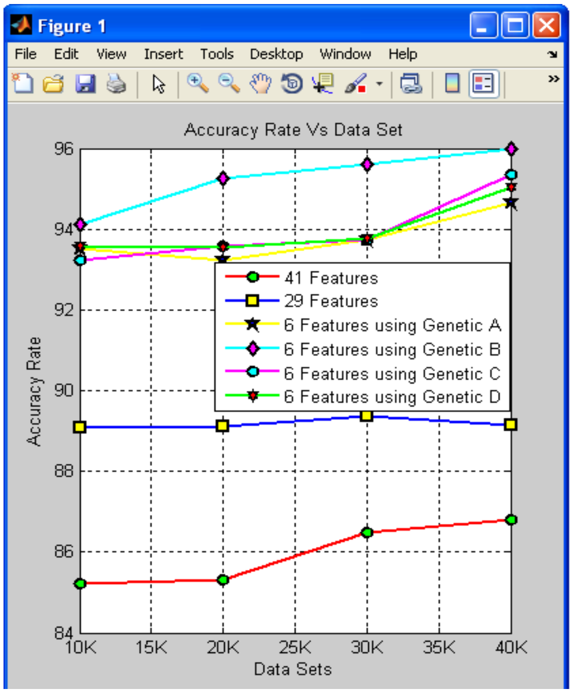

**Figure 3: Analysis of classification Accuracy of Full features, 29 RST [16], and Genetic Algorithm (A, B, C, D) set**

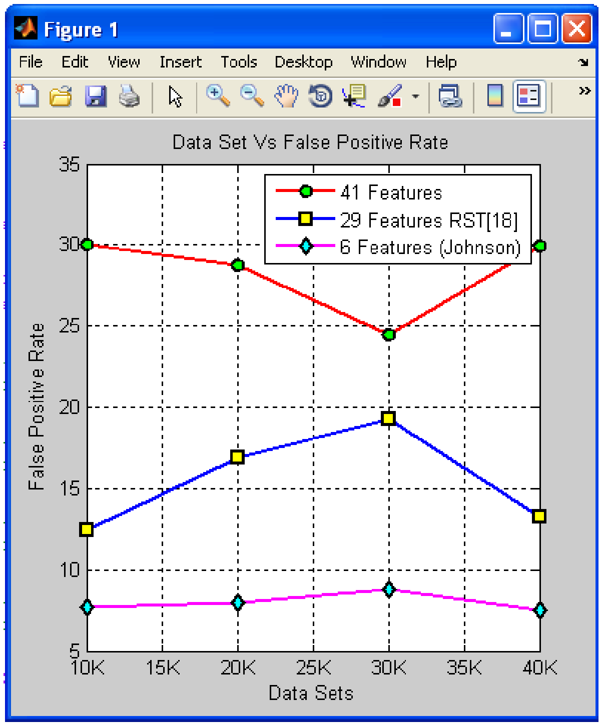

**Figure 4: Graphical Analysis of false positive rate using three feature sets that are 41 features, 29 features RST [16] and 6 features Johnson's Algorithm**

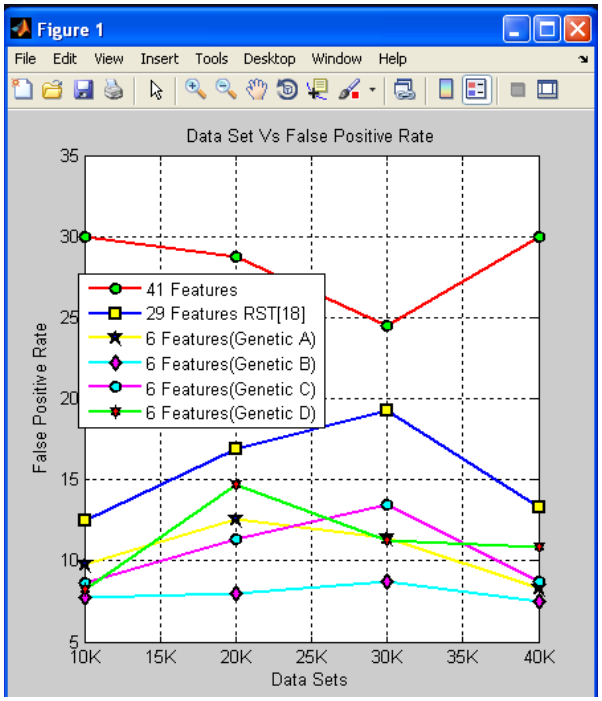

**Figure 5: Graphical Analysis of false positive rate using three feature sets that are 41 features, 29 features RST [16] and Genetic Algorithm (A, B, C, D) set**

## 7. CONCLUDING REMARKS

This paper focuses on the dimensionality reduction using feature selection. The RSSVM (Rough Set Support Vector Machine) approach deploy Johnson's and Genetic algorithm of rough set theory tool to find the reduct sets and sent to SVM to identify any type of new behavior either normal or attack. The accuracy of SVM is compared with full 41 features, 29 features, and 6 features (with five set of different attributes) respectively. Here, the proposed algorithm is used only 6 features out of 41 features; hence the CPU and memory utilization is decreased. Thus, proposed algorithm is very apt and reliable for intrusion detection.

## 8. REFERENCES

[1]. Cohen, Fred, "Computer Viruses: Theory and Experiments," 7th DOD/NBS Computer Security Conference, Gaithersburg, MD, September 24-26, 1984.

[2]. Denning, Dorothy E., "An Intrusion Detection Model," Proceedings of the Seventh IEEE Symposium on Security and Privacy, May 1986, pages 119–131.

[3]. R. Power, "CSI/FBI computer crime & security survey," Computer Security Journal, Vol.18, No.2, 2002, pp: 7-30.

[4]. Rawat, S., Gulati V., Pujari A., A Fast Host-based Intrusion Detection System Using Rough Set Theory, Transaction on Rough Sets IV, LNCS 3700, 2005.

[5]. Judy Weng, Glen Qin, "Network Intrusion Prevention Systems", JTB_Journal of Technology and Business. October 2007.

[6]. Helman, Paul, et al, Wynette, "Foundations of Intrusion Detection," The IEEE Computer Security Foundations Workshop V, 1992.

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

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

[9]. Wang Xuren, He Famei ," Improving Intrusion Detection Performance Using Rough Set Theory and Association Rule Mining", Hybrid Information Technology, 2006.

[10].Ohrn, A., Komorowski, J., A Rough Set Toolkit for Analysis of Data, In Proceedings of the third Joint conference on Information Sciences, Vol(3), USA, 1997, pp.403- 407, available on http://www.idi.ntnu.no/~aleks/rosetta.

[11].V. Vapnik, "The Nature of Statistical Learning Theory". NY: Springer-Verlag. 1995.

[12].Zia Akbar," Marketing data classification using Johnson's Algorithm", 2003

[13].Joachims T. "Estimating the Generalization Performance of a SVM Efficiently." Proceedings of the International Conference on Machine Learning, Morgan Kaufman, 2000.

[14].Godinez, F., Hutter, D., Monroy R., Attribute Reduction for Effective Intrusion Detection, AWIC 2004, LNAI 3034, 2004.

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

[16].Rung-Ching Chen, Kai-Fan Cheng and Chia-Fen Hsieh, "Using Rough Set And Support Vector Machine for Network Intrusion Detection", International Journal of Network Security & Its Applications (IJNSA),Vol 1 C. Chang and C. J. Lin, LIBSVM, No 1, April 2009.

## AUTHORS PROFILE

**Prof. Shailendra Kumar Shrivatava** has received his B.E. in Computer Science & Engineering from Barkatulla Univerisity, India in 1991 and M.E. in Computer Sceince & Engineering from University Of Rajiv Gandhi Proudyogiki Vishwavidhyalaya, India in 2001. He has 20 years experience in teaching field. He is currently working as an Associate Professor and HOD in Information Technology Department, Samrat Ashok Technological Institute (SATI), Vidisha, (M.P.), India. His research interest includes Artificial Intelligence and Machine Learning.

**Ms. Preeti Jain** has received her B.E. degree from Jiwaji University, India, in 2000. She has 3 Years 2 Months in Teaching & 2 Years 7 Months in Industry, (TOTAL: 5 Years 9 Months) Experience. Presently, she is a Research Scholar in Information Technology from SATI, Vidisha, (M.P.), and Rajiv Gandhi Proudyogiki Vishwavidhyalaya University, India. Her research interest includes Intrusion Detection System, Support Vector Machine and Rough Set Theory.