Selective Iteration based Particle Swarm Optimization (SIPSO) for Intrusion Detection System

Sana Warsi
M.Tech Research Scholar
Computer Science
SIST, Bhopal

Yogesh Rai
Assistant Professor
Computer Science
SIST, Bhopal

Santosh Kushwaha
HOD and Assistant Professor
Computer Science
SIST, Bhopal

ABSTRACT

In the current age Intrusion detection is an interest in and challenging area. As there are now a few exploration works are as of now done and the outcome change is in advancement. In this paper a hybrid approach has been proposed which is based on association rule mining and Selective Iteration based Particle Swarm Optimization (SIPSO). The NSL-KDD dataset is used. First normal and attack nodes are separated. Then normal node is checked for suspicious behavior. Then association rule mining is applied to form the associated for the next preprocessing. Then we apply SIPSO to check the threshold value obtained for the different intrusion types. If it is passed the threshold velocity assigned, then it will be categorized as the specific attack. We have considered a Denial of Service (DoS), User to Root (U2R), Remote to User (R2L) and Probing (Probe) attacks in this research work. The results show the improvement in detection as compared to the previous method.

Keywords

Association rule mining, SIPSO, DoS, U2R, R2L, Probe

1. INTRODUCTION

The Association for Computing Machinery (ACM) hosts a specific vested get-together on Knowledge Discovery and Data mining (KDD) [1] for the data mining understudies and investigators. They gave set KDD Cup99 data sets for interruption disclosure [2]. This gathering is utilized for interruption discovery and a few analysts had considered this as the benchmark data set for result correlation.

As of late, various specialists are focusing to use data burrowing thoughts for Intrusion Detection [3]. This is a methodology to think the undeniable information and learning.

Interruption disclosure is the procedure of malicious ambush in the structure and framework when we are instantly correspondence or isolating data in the steady environment [4][5]. Since its development, the intrusion area has been one of the key parts in fulfilling information security. It goes about as the second-line boundary which supplements the passage controls. Right when the controls failed, the intrusion distinguishing proof systems should have the ability to remember it consistent and alert the security officers to take incite and suitable exercises [5][6].

Interference acknowledgment structure oversees administering the scenes happening in PC system or framework circumstances and taking a gander at them for signs of possible events, which are certain threats to PC security, or standard security sharpens Intrusion recognizable proof structures (IDS) have ascended to recognize exercises which risk the uprightness, protection or openness of are sourced as a push to give a response for existing security issues [7].

So in the above course we contemplate a couple of points of view in the ensuing fragments. We in like manner discuss data mining and progression techniques, in light of the fact that it can be used as a piece of forming the structure which conveys better recognizable proof system.

As we are analyzing this study toward a prevalent framework with the blend of data mining and streamlining. These systems are useful and has been used as a piece of assorted approaches like [8][9][10][11][12][13]. So the usage of these counts can enhance an impact.

2. RELATED WORK

In 2012, LI Yin–huan [14] focuses on an improved FP-Growth computation. According to inventor Preprocessing of data mining can extend capability on looking the typical prefix of center and decrease the time capriciousness of building FP-tree. In perspective of the improved FP Growth count and other data mining frameworks, an interference area model is finished by inventors. Their exploratory results are fruitful and possible.

In 2012, P. Prasenna et al. [15] recommended that in standard framework security just relies on upon numerical estimations and low counter measures to taken to deflect intrusion recognizable proof system, but the greater part of this procedures to the extent speculatively tried to execute. Makers recommend that rather than creating generous number of standards the headway change systems like Genetic Network Programming (GNP) can be used .The GNP is in perspective of composed chart. They focus on the security issues related to send a data mining-based IDS in a ceaseless circumstance. They entirety up the issue of GNP with alliance rule mining and propose a cushy weighted connection rule mining with GNP framework suitable for both steady and discrete qualities.

In 2011, LI Han [16] focuses on interference revelation in light of collection examination. The fact of the matter is to improve the acknowledgment rate and reduction the false alert rate. A balanced component K-suggests computation called MDKM to distinguish irregularity activities is proposed and relating reenactment examinations are presented. Firstly, the MDKM computation channels the tumult and isolated spotlights on the data set. Likewise by finding out the detachments between all sample data centers, they procure the high-thickness parameters and gathering part parameters, using component iterative technique we get the k grouping concentrate accurately, then a peculiarity revelation model is shown. They used KDD CUP 1999 data set to test the execution of the model. Their results show the structure has a higher acknowledgment rate and a lower false alert rate, it achieves confident point.

In 2011, Z. Muda et al. [17] discuss the issue of current abnormality recognizable proof that it not ready to recognize an extensive variety of strikes viably. To beat this issue, they propose a crossbreed learning approach through mix of K-Means grouping and Naïve Bayes portrayal. The proposed system will be gathering all data into the looking at get-together before applying a classifier for request reason. An examination is done to evaluate the execution of the proposed technique using KDD Cup '99 dataset. Result exhibit that the proposed system performed better in term of precision, area rate with sensible false alert rate.

In 2014, Deshmukh et al. [18] presents a Data Mining framework in which distinctive preprocessing methods will be incorporated, for example, Normalization, Discretization and Feature decision. With the assistance of these strategies the data will be preprocessed and obliged highlights are picked. They used NaIve Bayes framework in coordinated learning procedure which bunches diverse framework events for the KDD cup'99 Dataset.

In 2014, Benaicha et al. [19] present a Genetic Algorithm (GA) approach with an upgraded beginning masses and decision director, to capably distinguish diverse sorts of framework interferences. They used GA to upgrade the look of attack circumstances in survey archives, on account of its awesome counterbalance examination/abuse; according to the inventors it gives the subset of potential strikes which are show in the audit report in a sensible planning time. The testing time of the Network Security Laboratory Knowledge Discovery and Data Mining (NSL-KDD99) benchmark dataset has been used to recognize the misuse works out. Their technique of IDS with Genetic estimation augments the execution of the recognizable proof rate of the Network Intrusion Detection Model and reductions the false positive rate.

In 2014 Kiss et al. [20] prescribe that Modern Networked Critical Infrastructures (NCI), including computerized and physical structures, are exhibited to sharp advanced strikes concentrating on the relentless operation of these systems. To ensure variation from the norm care, their watched data can be used as a piece of concurrence with data mining methodology to make Intrusion Detection Systems (IDS) or Anomaly Detection Systems (ADS). They proposed a gathering based approach for recognizing advanced strikes that cause idiosyncrasies in NCI. Distinctive clustering methods are examined to pick the most suitable for gathering the time-course of action data highlights, thusly portraying the states and potential advanced ambushes to the physical structure. The Hadoop execution of MapReduce standard is used to give a suitable get ready environment to broad datasets.

In 2014, Thaseen et al. [21] proposed a novel method for arranging crucial fragment examination (PCA) and support vector machine (SVM) by updating the piece parameters using customized parameter determination framework. Their approach reduces the planning and testing time to recognize interferences thus improving the precision. Their proposed method was attempted on

KDD data set. The datasets were carefully parceled into get ready and testing considering the minority strikes, for instance, U2R and R2L to be show in the testing set to recognize the occasion of dark ambush. Their results demonstrate that the proposed system is powerful in perceiving interferences. Their exploratory results exhibit that the request precision of the proposed framework defeats other course of action methods using SVM as the classifier and other dimensionality diminishing or highlight decision frameworks.

In 2014, Wagh et al. [22] proposed Network security is a fundamental piece of web enabled systems in the present world circumstance. According to the makers due to bewildering chain of PCs the open entryways for interferences and attacks have extended. Along these lines it is need of incredible significance to find the most perfect courses possible to secure our structures. So the inventors propose intrusion distinguishing proof system is expecting fundamental part for PC security. The best method used to handle issue of IDS is machine learning. Thy watched that the rising field of semi controlled learning offers an ensured course for relating investigation. So they proposed a semi-oversaw framework to lessen false ready rate and to upgrade disclosure rate for IDS.

In 2014, Masarat et al. [23] exhibited a novel multistep structure considering machine learning systems to make a capable classifier. In first step, the highlight decision procedure will execute considering get extent of highlights by the makers. Their method can upgrade the execution of classifiers which are made considering these highlights. In classifiers blend step, we will show a novel soft assembling procedure. Along these lines, classifiers with more execution and lower cost have more effect to make the last classification.

3. METHODS

The Association for Computing Machinery (ACM) has devised a Knowledge Discovery and Data mining (KDD) database[1] for the intrusion detection analysis and detection. They gave set KDD Cup99 data sets for interruption disclosure.

The flowchart in figure1 represents the methodology properly. The dataset considerd is NSL-KDD having 1025973 records with 41 attributes vlues. Among the 41 highlights, 1-9 are used to address the crucial highlights of a package, 10-22 use the substance accentuates, 23-31 are used for development highlights with two seconds of time window and 32-41 for host based highlights (Wenke Lee et al 1999). They are basically gathered into three classes: vital highlights of individual affiliation, substance offers inside an affiliation, and development highlights which are handled using a two seconds time window. Moreover, the KDD Cup99 data includes common and 22 different sorts of ambushes (Chi-Ho Tsang et al 2007). The attributes are Field1, Field2… .Field 41 for the supportive representation which will be profitable for using as a piece of our proposed methodology as exhibited in table 1. The field 4 has fundamental implications for choosing the filtering. It has 13 different relationship as demonstrated in table2.

The whole procedure is divided into following procedures.

1) Preprocessing

The data is preprocessed randomly and selected from 1025973 records. The detection are based on 4 different types of attacks name DoS, U2R, R2L, Probe.

2) Normal data Separation

At that point typical information division will occur on the selected record from the database as chose from the preprocessing. It will be handled in view of the fourth field and it is ended in light of the typical elements and afterward the remaining channel hub is prepared. We first consider Normal establishment and end as a run of the termination condition data and distinctive as the attack data [18]. By then we again channel the attack data considering the getting relationship as the conventional and set up the starting strike data.

3) Selective Iteration based Particle Swarm Optimization (SIPSO)

Then we apply Selective Iteration based Particle Swarm Optimization for the better classification. The algorithm is shown below:

```
Input:
    • ID(id1,id2….idn)
    • IDOS(idos1,idos2….idosn)
Output:
    • DN1…….DNn
ID→identification node
IDOS→Intrusion detection outputs
DN→ Deetection node
V→ Velocity
PRV→ Particle Random Velocity
PPRV → Previous Particle Random Velocity

Step 1: KDD dataset selection
Step 2: Initialize vlocity
Step 3: Particle Random Velocity
PRV= geerated vlue.
for i=1 ;i<4;i++
Step 4: Distribute ID for the below Iteration
        do
        $E_V=(ID_1*PRV_1 + ID_2* PRV_2 + ID_3 * PRV_3 +…. + ID_n * PRV_n)/n$
        If $(V_{t1} > V_{tn-1})$
        $V_{t1} = V_{tn-1}$
        PRV = PRV
        while;
For 2 to 4
        $T_V= Ev + (ID_1*PRV_1 + ID_2* PRV_2 + ID_3 * PRV_3 +…. + ID_n * PRV_n)/n + PRV$
        If $(V_{t1} > V_{tn-1})$
        $V_{t1} = V_{tn-1}$
        PRV = PRV
        while;
Step 5: Overall Accuracy
        $O_{AC}=\sum ID_i / n$
Step 6: Finish
```

The above algorithm shows the working phenomena based on association rule mining and 3) Selective iteration based Particle Swarm Optimization.

4) Attack Classification

This arrangement is taking into account the table 4 subtle elements. We have considered four unique sorts of assault. These assaults are DoS: back, area, neptune, smurf, teardrop, case. At that point in U2R the assaults are loadmodule,buffer_overflow and rootkit. At that point in R2L the assaults are phf, guess_passwd, warezmaster, imap, multihop, ftp_write",warezclient. At that point in Probe the assaults are "satan","nmap","portsweep","ipsweep". The outcome correlations are considering perl and spy in both the databases in light of the fact that it is not characterized particularly in R2L and U2R independently.

5) Final Analysis

The checking is done on the reason of differentiating the last strike database and the total database. It will be better cleared up in our result examination. The result exhibits the better portrayal to the extent DoS and test.

Table 1: NSL-KDD Dataset [1]

ID	Field1	Field2	Field3	Field4	Field5	Field6	Field7	Field10	Field8	……
1	0	tcp	ftp_data	SF	491	0	0	0	0	
2	0	udp	other	SF	146	0	0	0	0	
3	0	tcp	private	S0	0	0	0	0	0	
4	0	tcp	http	SF	232	8153	0	0	0	
5	0	tcp	http	SF	199	420	0	0	0	
6	0	tcp	private	REJ	0	0	0	0	0	
7	0	tcp	private	S0	0	0	0	0	0	
8	0	tcp	private	S0	0	0	0	0	0	
9	0	tcp	remote_job	S0	0	0	0	0	0	
10	0	tcp	private	S0	0	0	0	0	0	
…	….	…	..	…..	..	..	.	..	..	.

Table 2: Connection State Summary [24]

S.No	State	Description
1	S0	Connection attempt seen no reply.
2	S1	Connection established, not terminated.
3	SF	Normal establishment and termination.

4	REJ	Connection attempt rejected.
5	S2	Connection established and close attempt by originator seen (but no reply from responder).
6	S3	Connection established and close attempt by responder seen (but no reply from originator).
7	RSTO	Connection established, originator aborted (sent a RST).
8	RSTR	Established, responder aborted.
9	RSTOS0	Originator sent a SYN followed by a RST, we never saw a SYN ACK from the responder.
10	RSTRH	Responder sent a SYN ACK followed by a RST, we never saw a SYN from the (purported) originator.
11	SH	Originator sent a SYN followed by a FIN, we never saw a SYN ACK from the responder (hence the connection was “half” open).
12	SHR	Responder sent a SYN ACK followed by a FIN, we never saw a SYN from the originator.
13	OTH	No SYN seen, just midstream traffic (a “partial connection” that was not later closed).

Table 3: Associative Items

Node	T1	T2	T3	T4	T5	T6
66663	1	1	0.3333	0.5556	0.6	0.5
66723	1	1	0.3333	0.6667	0.6	0.5
66811	1	1	0.4444	0.5556	0.6	0.5
66830	1	1	0.2222	0.6667	0.3	0.6
66684	1	1	0.3333	0.6667	0.4	0.7
66706	1	1	0.3333	0.5556	0.6	0.5
66814	0.8462	0.9231	0.3333	0.6667	0.4	0.6
66857	1	1	0.2222	0.6667	0.3	0.6
66859	1	1	0.3333	0.6667	0.3	0.6
66863	1	1	0.2222	0.6667	0.3	0.6
66948	1	1	0.2222	0.6667	0.3	0.6
66951	1	1	0.4444	0.6667	0.5	0.5
66995	0.9231	1	0.3333	0.6667	0.5	0.7
67051	1	1	0.2222	0.6667	0.3	0.6
67053	1	1	0.3333	0.6667	0.6	0.5
67063	1	1	0.3333	0.6667	0.6	0.5
66697	1	1	0.2222	0.6667	0.3	0.6
66773	1	1	0.3333	0.6667	0.6	0.5
66729	1	1	0.3333	0.6667	0.3	0.6
66730	0.9231	1	0.4444	0.6667	0.3	0.6
66732	1	1	0.3333	0.6667	0.6	0.5
66733	1	1	0.4444	0.5556	0.7	0.5
66740	0.8462	0.9231	0.3333	0.6667	0.4	0.6
66758	1	1	0.3333	0.6667	0.6	0.5
66910	1	1	0.2222	0.6667	0.3	0.6
66875	1	1	0.2222	0.6667	0.3	0.6
66879	1	1	0.3333	0.6667	0.6	0.5
66897	1	1	0.2222	0.6667	0.3	0.6
66934	1	1	0.2222	0.6667	0.3	0.6
67013	1	1	0.4444	0.5556	0.6	0.5
67042	0.9231	1	0.4444	0.6667	0.5	0.5
67107	1	1	0.2222	0.6667	0.3	0.6
67140	0.9231	1	0.4444	0.6667	0.3	0.6
66663	1	1	0.3333	0.5556	0.6	0.5
66723	1	1	0.3333	0.6667	0.6	0.5
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…

Table 4: Types of Attack

TCP	back , buffer_overflow, ftp_write , guess_passwd, imap, ipsweep, land, loadmodule, multihop, neptune, nmap, normal, perl, phf, portsweep,rootkit, satan, spy, warezclient, warezmaster
UDP	Nmap, normal, rootkit, satan, teardrop
ICMP	Ipsweep, nmap, normal, pod, portsweep, satan, smurf

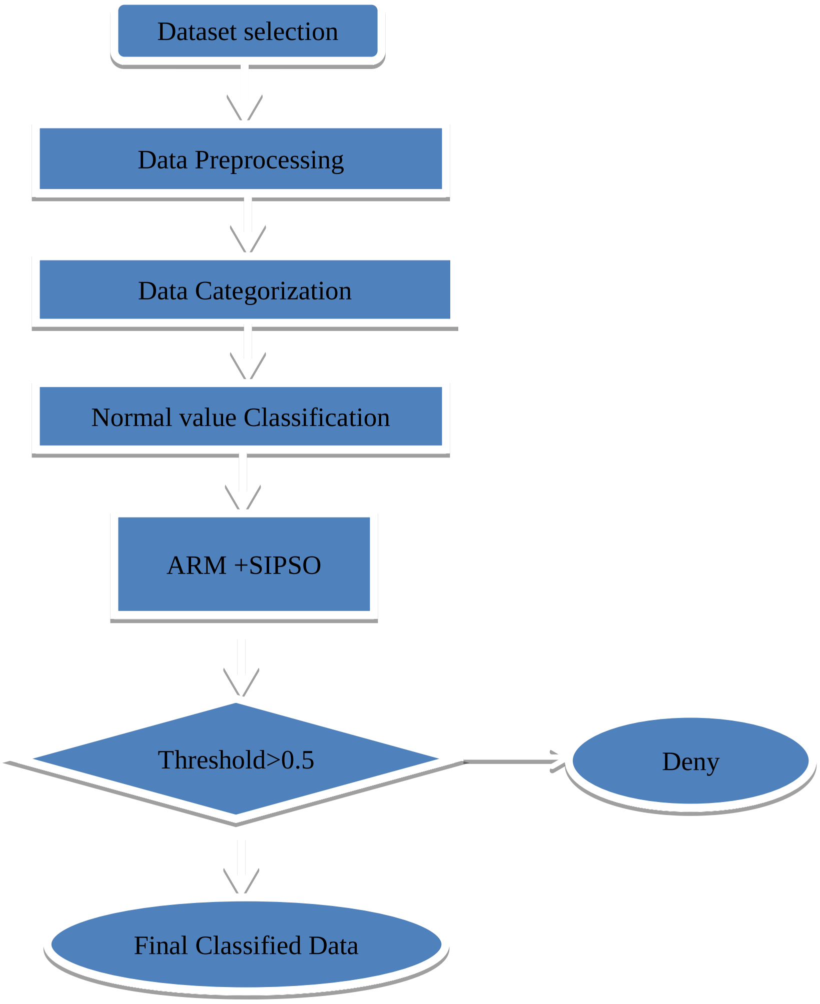

Figure 1: Process

4. RESULT

The last steps of the information is examined from the staying ordinary hub find. As those information are not got ordinary, but rather we can't say affirm as it is assaulted. The correlation is taking into account table 5, Table 6 and table 7. At that point the bolster quality is partitioned in six distinct parts. It is T1, T2… T6. At that point RPSO is connected on them. We put 0.5 as the bolster esteem. In the event that the hub crosses or likeness the worldwide ideal esteem then we will pass it into the assault database. In this way we will make our last database.

The final classifications taken for the result comparison is based on the four different attacks. The records are considered from 66630 to 763127. The result is shown in figure 2. DoS and Probe accuracy achived by our result is better.

Table 5: Content Features1 (10-22)

0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0

Table 6: Traffic Features1 (23-31)

1	1	0.00	0.00	1.00	1.000	0.01	0.06	0.00

1	1	0.00	0.00	0.00	0.000	1.00	0.00	0.4

Table 7: Host –Based Features1 (32-41)

1	1	0.00	0.06	0.00	0.00	0.00	0.00	1.00	1.00

1	1	1.00	0.00	0.01	0.03	0.00	0.00	0.00	0.00

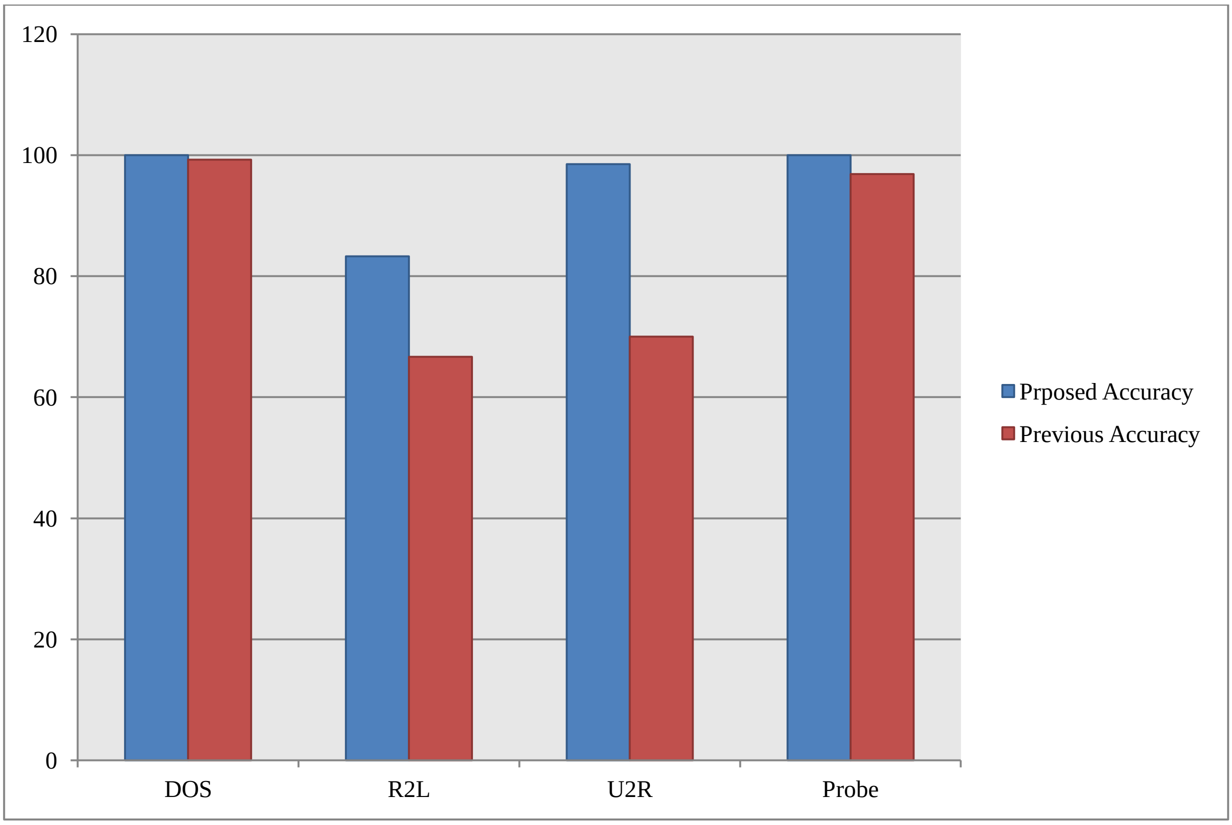

Figure 2: Classification accuracy

5. CONCLUSION

In this paper we have applied SIPSO which is based on association rule mining. This approach has been applied on normal data which is preprocessed and classified. It is so as to find the suspicious normal node to identified it correctly. The attacks identified are DoS, U2R, R2L and probe. DoS and Probe accuracy achived by our result is better. In future hybrid evolutionary algorithm can be applied to improve the detection.

6. REFERENCES

[1] Alexander O. Tarakanov, Sergei V. Kvachev, Alexander V. Sukhorukov ," A Formal Immune Network and Its Implementation for On-line Intrusion Detection", Lecture Notes in Computer Science Volume 3685, pp 394- 405, 2005.

[2] Ranjna Patel, Deepa Bakhshi and Tripti Arjariya,"Random Particle Swarm Optimization (RPSO) based Intrusion Detection System " , International Journal of Advanced Technology and Engineering Exploration (IJATEE), Volume-2, Issue-5, April-2015 ,pp.60-66.

[3] Meng Jianliang,Shang Haikun,Bian Ling," The Application on Intrusion Detection Based on K-means Cluster Algorithm", International Forum on Information Technology and Applications, 2009.

[4] Lundin, E. and Jonsson, E. "Survey of research in the intrusion detection area", Technical Report, Department of Computer Engineering, Chalmers University of Technology, Göteborg, Sweden. January 2002.

[5] R.Venkatesan, R. Ganesan, A. Arul Lawrence Selvakumar, " A Comprehensive Study in Data Mining Frameworks for Intrusion Detection " , International Journal of Advanced Computer Research (IJACR), Volume-2, Issue-7, December-2012 ,pp.29-34.

[6] S.Devaraju, S.Ramakrishnan:,"Analysis of Intrusion Detection System Using Various Neural Network classifiers, IEEE 2011.

[7] Moriteru Ishida, Hiroki Takakura and Yasuo Okabe," High-Performance Intrusion Detection Using OptiGrid Clustering and Grid-based Labelling", IEEE/IPSJ International Symposium on Applications and the Internet, 2011.

[8] S. T. Brugger, "Data mining methods for network intrusion detection",pp. 1-65, 2004.

[9] W. Lee, S. J. Stolfo, "Data Mining Approaches for Intrusion Detection",Proceedings of the 1998 USENIX Security Symposium, 1998.

[10] Kamini Nalavade, B.B. Meshram, "Mining Association Rules to Evade Network Intrusion in Network Audit Data " , International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-15, June-2014 ,pp.560-567.

[11] W. Lee, S. J. Stolfo, "Data mining approaches for intrusion detection" Proc. of the 7th USENIX Security Symp.. San Antonio, TX, 1998.

[12] Reyadh Naoum, Shatha Aziz, Firas Alabsi, "An Enhancement of the Replacement Steady State Genetic Algorithm for Intrusion Detection", International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-15, June-2014, pp.487-493.

[13] Aditya Shrivastava, Mukesh Baghel, Hitesh Gupta, " A Review of Intrusion Detection Technique by Soft Computing and Data Mining Approach " , International Journal of Advanced Computer Research (IJACR), Volume-3, Issue-12, September-2013 ,pp.224-228.

[14] LI Yin–huan , "Design of Intrusion Detection Model Based on Data Mining Technology", International Conference on Industrial Control and Electronics Engineering, 2012.

[15] P. Prasenna, R. Krishna Kumar, A.V.T Raghav Ramana and A. Devanbu "Network Programming And Mining Classifier For Intrusion Detection Using Probability Classification", Pattern Recognition, Informatics and Medical Engineering, March 21-23, 2012.

[16] LI Han, "Using a Dynamic K-means Algorithm to Detect Anomaly Activities", Seventh International Conference on Computational Intelligence and Security, 2011.

[17] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir," Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification", 7th International Conference on IT in Asia (CITA), 2011.

[18] Deshmukh, D.H.; Ghorpade, T.; Padiya, P., "Intrusion detection system by improved preprocessing methods and Naïve Bayes classifier using NSL-KDD 99 Dataset," Electronics and Communication Systems (ICECS), 2014 International Conference on , vol., no., pp.1,7, 13-14 Feb. 2014.

[19] Benaicha, S.E.; Saoudi, L.; Bouhouita Guermeche, S.E.; Lounis, O., "Intrusion detection system using genetic algorithm," Science and Information Conference (SAI), 2014 , vol., no., pp.564,568, 27-29 Aug. 2014.

[20] Kiss, I.; Genge, B.; Haller, P.; Sebestyen, G., "Data clustering-based anomaly detection in industrial control systems," Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on , vol., no., pp.275,281, 4-6 Sept. 2014.

[21] Thaseen, I.S.; Kumar, C.A., "Intrusion detection model using fusion of PCA and optimized SVM," Contemporary Computing and Informatics (IC3I), 2014 International Conference on , vol., no., pp.879,884, 27-29 Nov. 2014.

[22] Wagh, S.K.; Kolhe, S.R., "Effective intrusion detection system using semi-supervised learning," Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on , vol., no., pp.1,5, 5-6 Sept. 2014.

[23] Masarat, S.; Taheri, H.; Sharifian, S., "A novel framework, based on fuzzy ensemble of classifiers for intrusion detection systems," Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on , vol., no., pp.165,170, 29-30 Oct. 2014.

[24] Description of Kyoto University Benchmark Data http://www.takakura.com/Kyoto_data/BenchmarkData-Description-v3.pdf