ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
Volume 2, Issue 8, August 2016

An efficient Intrusion Detection System based on Random Particle Swarm Optimization (RPSO)

Shyam Cholera[1], Prof. Kailash Patidar[2] , Mr. Jitendra Rai[3]
PG Student, Dept. of CSE, School of Engineering, SSSUTMS, Sehore, M.P., India[1]
Professor & Head, Dept. of CSE/IT, School of Engineering, SSSUTMS, Sehore, M.P., India[2]
Asst. Prof. , Dept. of CSE, School of Engineering, SSSUTMs, Sehore , M.P., India[3]

ABSTRACT- Intrusion detection is a challenging area of research. As now there are several research work are already done and the result improvement is in progress. In this paper a hybrid combination of association rule mining and random particle swarm optimization (RPSO) has been applied. This approach is applied on NSL-KDD dataset. A limit set is provided by our framework which will be adapted as per the user choice to select the set of data for use. Our approach successfully differentiates the normal and attack node. Then we have applied a recheck frame for the normal node for finding the suspicious node. Then by the help of association rule associated values are passed for the next procedure. Then we apply RPSO to check the boundary value for the possible type of intrusion detection. If it is passed the velocity value then it will be listed in the attack type. Finally based on the attack category of Denial of Service (DoS), User to Root (U2R), Remote to User (R2L) and Probing (Probe) attacks are classified. The results of our method shows improvement in detection different type of attack in comparison to the previous method.

KEYWORDS- Association rule mining, RPSO, DoS, U2R, R2L, Probe

I. INTRODUCTION

Detection is an essential concern in interruption location. Intrusion Detection System (IDS)[1] by using KDD data set[2] had been demonstrated for keeping up information uprightness. It is a mix of pitfalls considering programming and equipment gear [3]. The assaults are for the most part isolated into four separate parts. 1) Where the aggressors relate to be occupied the hub which is asked for by a few clients by fake vicinity is called Denial of Service assault. 2) When the assailants pick up the root access of a client account then it is called User to Root assault. 3) When the aggressor increase illicit nearby get to then it is called Remote to User assault. 4) When any aggressors control the data then it is called examining. Anyhow just intrusion counteractive action is insufficient. As frameworks get to be more complex, there are constantly exploitable shortcomings in the frameworks because of outline and programming slips, or different entrance methods. Thusly Intrusion discovery is needed as another measure to secure our PC frameworks [4]. Information mining procedures have been effectively connected in numerous fields like Network Management, Education, Science, Business, Manufacturing, Process control, and Fraud Detection. Information Mining for IDS is the procedure which can be utilized mostly to recognize obscure assaults and to raise alerts when security infringement is distinguished [5].

The primary inspiration driving utilizing interruption location as a part of information digging [5][6][7][8][9][10][11] is for better relationship with the related assaults and ordinary information. Better order can likewise be performed when we separate it by utilizing characterization and affiliation rules Preliminary

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
Volume 2, Issue 8, August 2016

arrangement could be possible by backing and certainty values. The approaches with genetic algorithm, support vector machine etc. along with different data mining techniques are also applied. The related approaches are [12][13][14][15].

The remaining of this paper is organized as follows. In Section 2 we discuss about PSO. In section 3 we discuss about the implementation. In section 4 we discuss about the results. The conclusions are given in Section 5.

II. PSO [16][17]

PSO gained from the situation and utilized it to take care of the improvement issues. In PSO, each one single arrangement is a "fledgling" in the hunt space. We call it "particle". All of particles have wellness values which are assessed by the wellness capacity to be upgraded, and have speeds which administer the flying of the particles. The particles fly through the issue space by emulating the current ideal particles. PSO is introduced with a gathering of arbitrary particles (arrangements) and afterward hunt down optima by redesigning eras. In every cycle, every particle is overhauled by taking after two "best" values. The first is the best arrangement (wellness) it has accomplished as such. (The wellness quality is likewise put away.) This worth is called pbest. An alternate "best" esteem that is followed by the particle swarm analyzer is the best esteem, acquired so far by any particle in the populace. This best esteem is a worldwide best and called gbest. At the point when a particle partakes of the populace as its topological neighbors, the best esteem is a neighborhood best and is called lbest.

In the wake of discovering the two best values, the particle upgrades its speed and positions with taking after comparison (an) and (b).

v[] = v[] + c1 * rand() * (pbest[] - present[]) + c2 * rand() * (gbest[] - present[]) (a)

present[] = present[] + v[] (b)

v[] is the particle velocity, present[] is the current particle (solution). pbest[] and gbest[] are defined as stated before. rand () is a random number between (0,1). c1, c2 are learning factors. usually c1 = c2 = 2.

The pseudo code of the procedure is as follows

```
For each particle
Initialize particle
END
Do
   For each particle
      Calculate fitness value
      If the fitness value is better than the best fitness value (pBest) in history
         set current value as the new pBest
   End
 Choose the particle with the best fitness value of all the particles as the gBest
   For each particle
     Calculate particle velocity according equation (a)
      Update particle position according equation (b)
   End
```

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
Volume 2, Issue 8, August 2016

While maximum iterations or minimum error criteria is not attained Particles' velocities on each dimension are clamped to a maximum velocity Vmax. If the sum of accelerations would cause the velocity on that dimension to exceed Vmax, which is a parameter specified by the user. Then the velocity on that dimension is limited to Vmax.

III. IMPLEMENTATION

The Association for Computing Machinery (ACM) has a particular vested party on Knowledge Discovery and Data mining (KDD) [20] for the information mining understudies and analysts. They gave set KDD Cup99 information sets for intrusion discovery.

In our methodology which is likewise better clarified by the flowchart as demonstrated in figure 1. We are first considering the NSL-KDD Dataset having 1025973 records with 41 highlights. Among the 41 highlights, 1-9 are utilized to speak to the essential highlights of a bundle, 10-22 utilize the substance emphasizes, 23-31 are utilized for movement highlights with two seconds of time window and 32-41 for host based highlights (Wenke Lee et al 1999). They are essentially assembled into three classes: essential highlights of individual association, substance offers inside an association, and movement highlights which are processed utilizing a two seconds time window. Additionally, the KDD Cup99 information involves ordinary and 22 separate sorts of assaults (Chi-Ho Tsang et al 2007). The highlights are named as Field1, Field2… .Field 41 for the helpful representation which will be advantageous for utilizing as a part of our proposed strategy as demonstrated in table 1. The field 4 has vital ramifications for deciding the sifting. It has 13 separate associations as indicated in table 2.

This approach is divided into five different parts as shown below.

1) Preprocessing

It is used to select random limit set from 1025973 records. This is then used for final detection of DoS, U2R, R2L, Probe along with the normal features.

2) Normal data Separation

Then normal data separation will take place on the selected database as selected from the preprocessing. It will be processed based on the fourth field and it is terminated based on the normal features and then the remaining filter node is processed. We first consider Normal foundation and end as a typical condition information and different as the assault information [18]. At that point we again channel the assault information taking into account the getting association as the ordinary and set up the introductory assault information.

3) Random Particle Swarm Optimization (RPSO)

Then we apply random particle swarm optimization for the better classification. The algorithm is shown below:

Input:

- PS(ps1,ps2….psn)
- OS(Os1,Os2….Osn)

Output:

- ET1…….ETn

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
Volume 2, Issue 8, August 2016

Ps→ Particles

OS→Optimal Set

ET→Efficient Trails

V→ Velocity

RV→Random Velocity

RV_p → Previous Random Velocity

Step 1: KDD

Step 2: Initialize particle

Step 3: Random Velocity Calculation

for i=0 ;i<=5;i++

RV_i=Math.random();

Step 4: Distribute PS for the below Iteration

do

$E_V=(PS_1*RV_1 + PS_2* RV_2 + PS_3 * RV_3 +\ldots + PS_n * RV_n)/n$

If ($V_{t1} > V_{tn-1}$)

$V_{t1} = V_{tn-1}$

$RV_p = RV_i$

while;

For 2 to 5

$T_V=(PS_1*RV_1 + PS_2* RV_2 + PS_3 * RV_3 +\ldots + PS_n * RV_n)/n$ - value(RV_p)

$V_{t1} = V_{tn-1}$

If ($V_{t1} > V_{tn-1}$)

$V_{t1} = V_{tn-1}$

Step 5: Overall Accuracy

$O_{AC}=\sum PS_i / n$

Step 6: Finish

The above algorithm clearly shows the working phenomena based on support and RPSO.

4) Attack Classification

This classification is based on the table 4 details. We have considered four different types of attack. These attacks are DoS: back, land, neptune, smurf, teardrop, pod. Then in U2R the attacks are loadmodule,buffer_overflow and rootkit. Then in R2L the attacks are phf, guess_passwd, warezmaster, imap, multihop, ftp_write",warezclient. Then in Probe the attacks are "satan","nmap","portsweep","ipsweep". The result comparisons are considering perl and spy in both the databases because it is not defined specifically in R2L and U2R separately.

5) Final Analysis

Last investigation is done on the premise of contrasting the last assault database and the aggregate database. It will be better clarified in our outcome investigation. The outcome demonstrates the better characterization as far as DoS and test.

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
IJARSMT
Volume 2, Issue 8, August 2016

Table 1: NSL-KDD Dataset [20]

ID	Field1	Field2	Field3	Field4	Field5	Field6	Field7	Field10	Field8	……	Field 41
1	0	tcp	ftp_data	SF	491	0	0	0	0		20
2	0	udp	other	SF	146	0	0	0	0		15
3	0	tcp	private	S0	0	0	0	0	0		19
4	0	tcp	http	SF	232	8153	0	0	0		21
5	0	tcp	http	SF	199	420	0	0	0		21
6	0	tcp	private	REJ	0	0	0	0	0		21
7	0	tcp	private	S0	0	0	0	0	0		21
8	0	tcp	private	S0	0	0	0	0	0		21
9	0	tcp	remote_job	S0	0	0	0	0	0		21
10	0	tcp	private	S0	0	0	0	0	0		21
…	….	…	..	…..	..	..	.	..	..	.	.

Table 2: Connection State Summary [21]

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).

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
Volume 2, Issue 8, August 2016

Table 3: Attack Detection

Node	T1	T2	T3	T4	T5	T6
66622	1	1	0.2222	0.6667	0.3	0.6
66663	1	1	0.3333	0.5556	0.6	0.5
66684	1	1	0.3333	0.6667	0.4	0.7
66697	1	1	0.2222	0.6667	0.3	0.6
66706	1	1	0.3333	0.5556	0.6	0.5
66723	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
66773	1	1	0.3333	0.6667	0.6	0.5
66811	1	1	0.4444	0.5556	0.6	0.5
66814	0.8462	0.9231	0.3333	0.6667	0.4	0.6
66830	1	1	0.2222	0.6667	0.3	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
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
66910	1	1	0.2222	0.6667	0.3	0.6
66934	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
66980	1	1	0.2222	0.6667	0.3	0.6
66995	0.9231	1	0.3333	0.6667	0.5	0.7
67013	1	1	0.4444	0.5556	0.6	0.5
67042	0.9231	1	0.4444	0.6667	0.5	0.5
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
67085	1	1	0.4444	0.5556	0.7	0.5
67107	1	1	0.2222	0.6667	0.3	0.6

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
Volume 2, Issue 8, August 2016

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

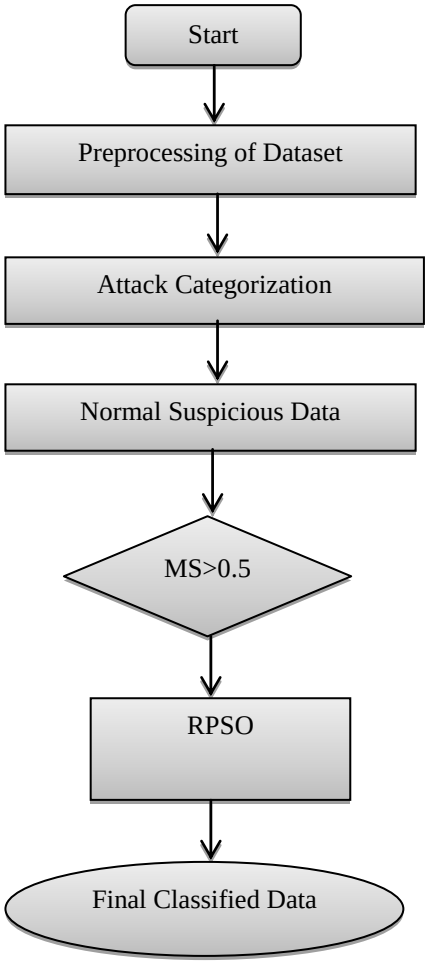

Fig. 1 Working Flowchart

1. Result

The final attack data is scanned from the remaining normal node find. As those data are not received normal but we cannot say confirm as it is attacked. The comparison is based on table 5, Table 6 and table 7. Then the support value is divided in six different parts. It is T1, T2… T6. Then RPSO is applied on them. We put 0.5 as the support value. If the node crosses or equivalent of the global optimum value then we will pass it into the attack database.

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
Volume 2, Issue 8, August 2016

In this manner we will create our final database.

Then we check the classifications based on the four attacks. We have considered the starting set from 66622 to 76312. The result is shown in figure 2. The results shown by our approach depicts better accuracy in terms of DoS and Probe accuracy. Table 8 shows the overall comparison from the previous results.

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

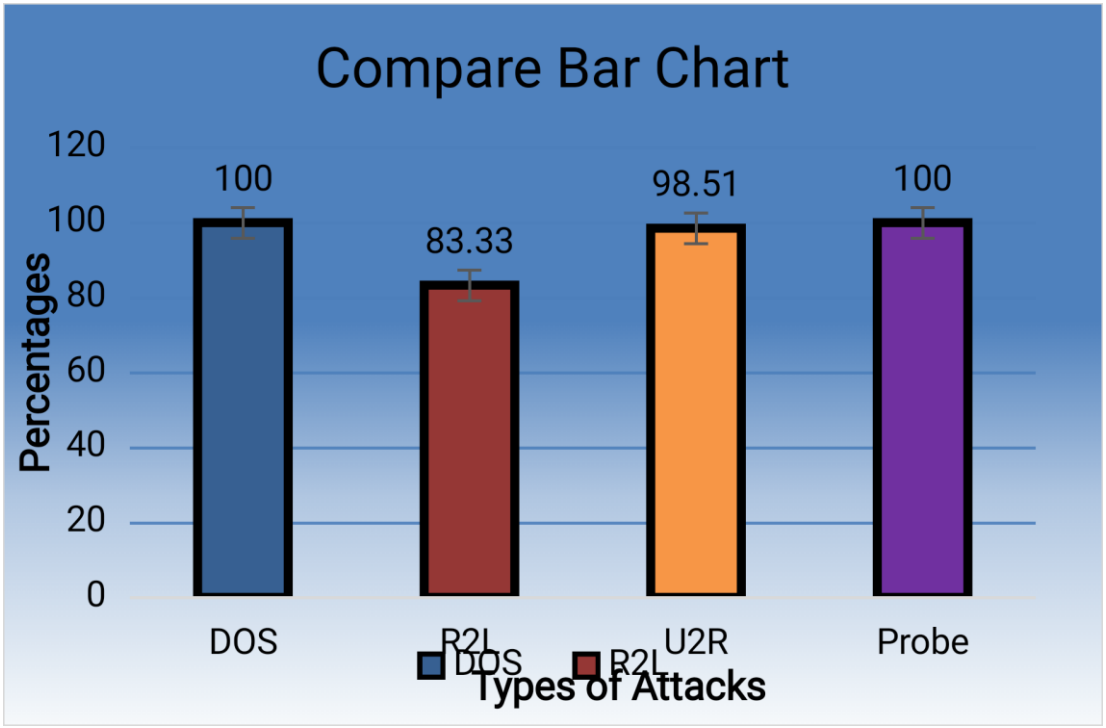

Fig. 2 Classification Accuracy

Table 8: Comparison

Model	Accuracy
Proposed Approach	**95.46 %**
Fuzzy Ensemble	**93 %**
Random Forest [22]	**92.93 %**
JRip [23]	**92.30 %**
SVM [24]	**92.18 %**

IV. CONCLUSION

In this paper we have applied RPSO along with the association rule mining approach. We have applied this approach on the classified normal data so that the suspicious normal node can be identified. We have identified four different types of attack name DoS, U2R, R2L and probe. Our approach produce better results in terms of DoS and probe and the overall accuracy is also better in comparison to the previous approach.

REFERENCES

[1] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

[2] R. Bane, N. Shivsharan, "Network intrusion detection system (NIDS)", pp. 1272-1277, 2008.

[3] Gudadhe, M.; Prasad, P.; Wankhade, K., "A new data mining based network Intrusion Detection model," Computer and Communication Technology (ICCCT), 2010 International Conference on , vol., no., pp.731,735, 17-19 Sept. 2010.

[4] Vitthal Manekar, Kalyani Waghmare," Intrusion Detection System using Support Vector Machine (SVM) and Particle Swarm Optimization (PSO) " , International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-16, September-2014 ,pp.808-812.

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

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

[8] 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.

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

[10] 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.

[11] W. Lee, S. J. Stolfo, K. W. Mok, "A data mining framework for building intrusion detection models", Proc. of the 1999 IEEE Symp.on Security and Privacy, pp. 120--132. Oakland, CA, 1999.

[12] 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.

[13] Manish Somani, Roshni Dubey, " Design of Intrusion Detection Model Based on FP-Growth and Dynamic Rule Generation with Clustering " , International Journal of Advanced Computer Research (IJACR), Volume-3, Issue-10, June-2013 ,pp.146-150.

[14] M. Panda, M. Patra, "Ensemble rule based classifiers for detecting network intrusions", pp 19-22, 2009.

[15] Z. Yu, J. Chen, T. Q. Zhu, "A novel adaptive intrusion detection system based on data mining", pp.2390-2395, 2005.

[16] Keon-Myung Lee,Mining Generalized Fuzzy Quantitative Associaton Rules wih Fuzzy Generalization Hierarchies, IEEE 2011.

[17] Kennedy, James. "Particle swarm optimization." In Encyclopedia of Machine Learning, pp. 760-766. Springer US, 2010.

[18] Ruchita Gupta, C.S.Satsangi, "An Efficient Range Partitioning Method for Finding Frequent Patterns from Huge Database", International Journal of Advanced Computer Research (IJACR), Volume-2, Issue-4, June-2012, pp.62-69.

[19] Shushma Lata, "An Iterative PSO for Web worth Optimization through random velocity", International Journal of Advanced Technology and Engineering Exploration (IJATEE), Volume-2, Issue-3, February-2015, pp.31-36.

[20] 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.

[21] http://www.takakura.com/Kyoto_data/BenchmarkData-Description-v3.pdf

[22] J. Zhang, M. Zulkernine, and A. Haque, "Random-forestsbased network intrusion detection systems," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, vol. 38, pp. 649-659, 2008.

[23] H. A. Nguyen and D. Choi, "Application of data mining to network intrusion detection: classifier selection model," in Challenges for Next Generation Network Operations and Service Management, ed: Springer, 2008, pp. 399-408.

[24] T. Ambwani, "Multi class support vector machine implementation to intrusion detection", Proceedings of the International Joint Conference on Neural Networks, 2003, pp. 2300-2305.