ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 8, August 2016

Intrusion Detection Using Efficient Swarm Intelligence

[1] A Shriya [2] B Harshitha [3] K Archana [4] B.Sujatha
[1][2][3] Student [4] Assistant professor
[1][2][3][4]From Matrusri Engineering College

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 dissertation a hybrid approach has been proposed which is based on association rule mining and Intrusion Detection Using Swarm Intelligence Based on Iterative Selection. 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 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, DoS, U2R, R2L, Probe**

I. 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.

II. LITERATURE SURVEY

In 2012, LI Yin–huan [14] concentrates on an enhanced FP-Growth calculation. As per creator Preprocessing of information mining can expand proficiency on looking the normal prefix of hub and diminish the time unpredictability of building FP-tree. In view of the enhanced FP Growth calculation and other information mining systems, an interruption location model is completed by creators. Their exploratory results are successful and doable.

In 2012, P. Prasenna et al. [15] proposed that in ordinary system security just depends on numerical calculations and low counter measures to taken to avert

ISSN (Online) 2394-6849
International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 8, August 2016

interruption identification framework, albeit the majority of this methodologies as far as hypothetically tested to execute. Creators propose that as opposed to producing substantial number of principles the advancement improvement procedures like Genetic Network Programming (GNP) can be utilized .The GNP is in view of coordinated diagram. They concentrate on the security issues identified with send an information mining-based IDS in a continuous situation. They sum up the issue of GNP with affiliation principle mining and propose a fluffy weighted affiliation guideline mining with GNP system suitable for both constant and discrete qualities.

In 2011, LI Han [16] concentrates on interruption discovery in light of grouping examination. The point is to enhance the recognition rate and abatement the false caution rate. An adjusted element K-implies calculation called MDKM to identify inconsistency exercises is proposed and relating reenactment analyses are introduced. Firstly, the MDKM calculation channels the commotion and segregated focuses on the information set. Also by ascertaining the separations between all example information focuses, they acquire the high-thickness parameters and group part parameters, utilizing An easy way to comply with the conference paper formatting requirements is to use this document as a template andsimply type your text into it.

Element iterative methodology we get the k bunching focus precisely, then an oddity discovery model is displayed. They utilized KDD CUP 1999 information set to test the execution of the model. Their outcomes demonstrate the framework has a higher recognition rate and a lower false caution rate, it attains to hopeful point.
In 2011, Z. Muda et al. [17] talk about the issue of current irregularity identification that it not able to distinguish a wide range of assaults effectively. To beat this issue, they propose a half breed learning approach through blend of K-Means bunching and Naïve Bayes characterization. The proposed methodology will be grouping all information into the comparing gathering before applying a classifier for order reason. An examination is done to assess the execution of the proposed methodology utilizing KDD Cup '99 dataset. Results demonstrate that the proposed methodology performed better in term of exactness, location rate with sensible false caution rate. In 2014, Deshmukh et al. [18] presents a Data Mining system in which different preprocessing techniques will be included such as Normalization, Discretization and Feature choice. With the help of these techniques the information will be preprocessed and obliged highlights are chosen. They utilized NaIve Bayes system in directed learning strategy which groups different system occasions for the KDD cup'99 Dataset.

In 2014, Benaicha et al. [19] present a Genetic Algorithm (GA) approach with an enhanced starting populace and choice administrator, to proficiently identify different sorts of system interruptions. They utilized GA to enhance the look of assault situations in review documents, thanks to its great offset investigation / misuse; as per the creators it gives the subset of potential assaults which are display in the review document in a sensible preparing time. The testing period of the Network Security Laboratory Knowledge Discovery and Data Mining (NSL-KDD99) benchmark dataset has been utilized to identify the abuse exercises. Their methodology of IDS with Genetic calculation increments the execution of the identification rate of the Network Intrusion Detection Model and decreases the false positive rate.

In 2014 Kiss et al. [20] recommend that Modern Networked Critical Infrastructures (NCI), including digital and physical frameworks, are presented to keen digital assaults focusing on the steady operation of these frameworks. To guarantee abnormality mindfulness, their watched information can be utilized as a part of agreement with information mining procedures to create Intrusion Detection Systems (IDS) or Anomaly Detection Systems (ADS). They proposed a grouping based methodology for identifying digital assaults that cause peculiarities in NCI. Different bunching strategies are investigated to pick the most suitable for grouping the time-arrangement information highlights, consequently characterizing the states and potential digital assaults to the physical framework. The Hadoop execution of MapReduce standard is utilized to give a suitable preparing environment to extensive datasets.

In 2014, Thaseen et al. [21] proposed a novel technique for coordinating vital segment investigation (PCA) and bolster vector machine (SVM) by upgrading the piece parameters utilizing programmed parameter determination system. Their methodology lessens the preparation and testing time to distinguish interruptions consequently enhancing the exactness. Their proposed

ISSN (Online) 2394-6849
International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 8, August 2016

strategy was tried on KDD information set. The datasets were painstakingly partitioned into preparing and testing considering the minority assaults, for example, U2R and R2L to be exhibit in the testing set to distinguish the event of obscure assault. Their outcomes show that the proposed strategy is effective in recognizing interruptions. Their exploratory results demonstrate that the order exactness of the proposed system outflanks other arrangement strategies utilizing SVM as the classifier and other dimensionality decrease or highlight choice systems.

In 2014, Wagh et al. [22] proposed Network security is an essential part of web empowered frameworks in the present world situation. As per the creators because of perplexing chain of PCs the open doors for interruptions and assaults have expanded. In this way it is need of great importance to locate the most ideal routes conceivable to secure our frameworks. So the creators propose interruption identification framework is assuming basic part for PC security. The best strategy used to tackle issue of IDS is machine learning. Thy watched that the rising field of semi regulated learning offers a guaranteed route for corresponding exploration. So they proposed a semi-managed system to diminish false alert rate and to enhance discovery rate for IDS.

In 2014, Masarat et al. [23] presented a novel multistep structure taking into account machine learning procedures to make a proficient classifier. In first step, the highlight choice technique will execute taking into account pick up proportion of highlights by the creators. Their technique can enhance the execution of classifiers which are made taking into account these highlights. In classifiers mix step, we will exhibit a novel fluffy gathering technique. In this way, classifiers with more execution and lower expense have more impact to make the last classifier.

III. METHOD

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 values. 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) Swarm Intelligence (SI)

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

ISSN (Online) 2394-6849
International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 8, August 2016

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
EV=(ID1*PRV1 + ID2* PRV2 + ID3 * PRV3 +…. + IDn * PRVn)/n
If (Vt1 > Vtn-1)
Vt1 = Vtn-1
PRV = PRV
while;
For 2 to 4
TV= Ev + (ID1*PRV1 + ID2* PRV2 + ID3 * PRV3 +…. + IDn * PRVn)/n + PRV
If (Vt1 > Vtn-1)
Vt1 = Vtn-1
PRV = PRV
while;
Step 5: Overall Accuracy
OAC=ΣIDi / n
Step 6: Finish
The above algorithm shows the working phenomena based on association rule mining and 3) Swarm Intelligence
Based on
Iterative Selection

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

ID	Field1	Field2	Field3	Field4	..	..	Field39	Field40	Field41	Field42
1	0	tcp	ftp_data	SF			0	0.05	0	normal
2	0	udp	other	SF			0	0	0	normal
3	0	tcp	private	S0			1	0	0	neptune
4	0	tcp	http	SF			0.01	0	0.01	normal
5	0	tcp	http	SF			0	0	0	normal
6	0	tcp	private	REJ			0	1	1	neptune
7	0	tcp	private	S0			1	0	0	neptune
8	0	tcp	private	S0			1	0	0	neptune
9	0	tcp	remote_job	S0			1	0	0	neptune
10	..	....	..	..	..	..	.	..	.	....
11	..	..	.	..	.	..	..	.....	.....	....

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 try by originator seen (but no reply from responder)
6	S3	Connection established and close attempt by answered seen (but no reply from originator).
7	RSTO	Connection established, originator aborted (sent a RST).
8	RSTR	Established, answered aborted.
9	RSTOS0	Originator sent a SYN ensue by a RST, we nevermore saw a SYN ACK from the answered.
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 answer (hence the connection was "half" open).
12	SHR	Responder sent a SYN ACK ensue by a FIN, we never saw a SYN from the originator.
13	OTH	No SYN seen, just mid chance 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

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

ISSN (Online) 2394-6849
International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 8, August 2016

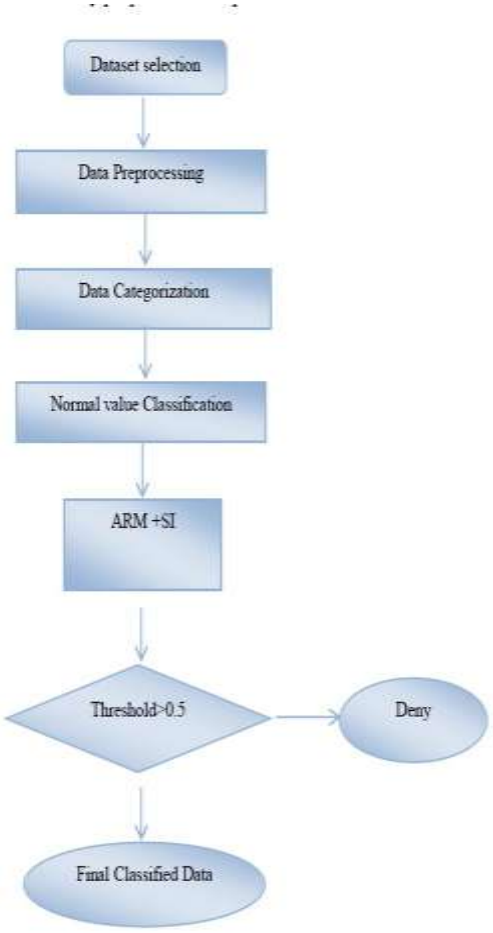

Figure 1: Process Flowchart

IV. 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

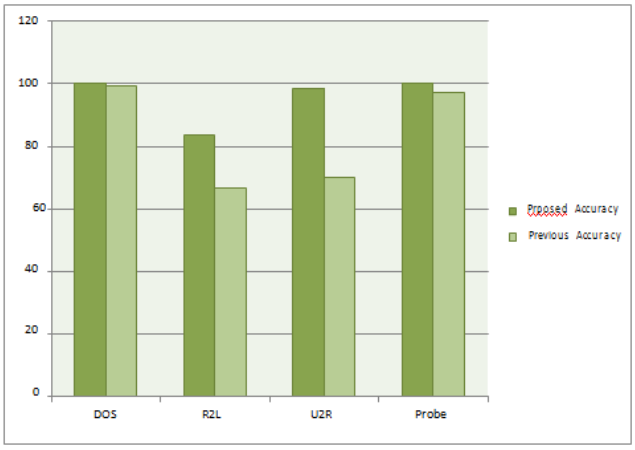

Figure 2: Classification accuracy

V. CONCLUSION

In this paper we have applied Swarm Intelligence Based on Iterative Selection 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.

REFERENCES

1) Alexander O. Tarakanov, Sergei V. Kvachev, Alexander V. Sukhorukov ," A Formal Immune Network and ItsnImplementation 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, " 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
    a. 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 e Conference 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