An Efficient Approach for Intrusion Detection in Reduced Features of Network Data Using Particle Swarm Optimization

Sadiya Anjum
M.Tech Scholar, Department of CSE
PIT, Bhopal INDIA
E-mail- sadiyaanjum09@gmail.com

Prof. Surendra Vishwakarma
Department of CSE
PIT, Bhopal India
E-mail- s.vish83@gmail.com

ABSTRACT

The process of clustering technique plays an important role in intrusion detection system. The processes of clustering technique grouped the network traffic data on the basis of similarity and validate the traffic data. The process of clustering suffered from the problem of large number of iteration and loss of data. Now a day's various authors used various optimization technique for the controlling the number of iteration and selection of seed. In consequence of this used particle of swarm optimization technique. The particle of swarm optimization technique reduces the loss of data and increase the validation of cluster content. In this paper tried to propose a very simple and fast clustering method for intrusion detection. A hybrid scheme based on coupling two different algorithms one is particle of swarm optimization and other is k-means algorithm. The main originality of proposed approach relies on associating two techniques: extracting more information bits via specific linguistic techniques, space reduction mechanisms, and moreover arcing cluster to aggregate the best clustering result. For the validation and performance evaluation of proposed algorithm used MATLAB software and KDDCUP99 dataset 10%. This dataset contains approx 5 lacks number of instance.

Index Terms:- IDS, PSO, KDD, NIDS.

INTRODUCTION

The performance of intrusion detection system depends on classification of unknown types of attacks. The detection of unknown types of attack is very difficult due to large number of attribute and huge amount of network data. For the improvement of unknown attack feature reduction is important area of research. The reduction process reduces the large number of attribute and improved the detection of intrusion detection system. In the process of feature reduction various algorithm are used such algorithm are principle of component analysis and neural network. The reduction process used PCA method this method is static reduction technique, reduces only fixed number of attribute. The fixed number of feature reduction process not justify the value of feature it directly reduces the feature. On the consideration of computational time feature reduction is also an important aspects, the reduces feature increase the processing of detection ratio. Many methods have been proposed in the last decades on the designs of IDSs based on feature reduction technique. With the tremendous growth of network-based services and sensitive information on networks, network security is becoming more and more importance than ever before. Intrusion detection techniques are the last line of defenses against computer attacks behind secure network architecture design, firewalls, and personal screening. Despite the plethora of intrusion prevention techniques available, attacks against computer systems are still successful. Thus, intrusion detection systems (IDSs) play a vital role in network security. Symantec in a recent report uncovered that the number of fishing attacks targeted at stealing confidential information such as credit card numbers, passwords, and other financial information are on the rise, going from 9 million attacks in June2013 to over 33 millions in less than a year. One solution to this is the use of network intrusion detection systems (NIDS) that detect attacks by observing various network activities. It is therefore crucial that such systems are accurate in identifying attacks, quick to train and generate as few false positives as possible. Internet has rapidly become one of the main communication methods in our society. Various types of internet application and usage are available more and more. Increasing usages of network applications also increase security risks to internet users, to prevent unwanted or dangerous threats.

An Intrusion Detection System (IDS) inspects the activities in a system for suspicious behavior or patterns that may indicate system attack or misuse. There are two main categories of intrusion detection techniques; Anomaly detection and Misuse detection. The former analyses the information gathered and compares it to a defined baseline of what is seen as "normal" service behavior, so it has the ability to learn how to detect network attacks that are currently unknown. Misuse Detection is based on signatures for known attacks, so it is only as good as the database of attack signatures that it uses for comparison. Misuse detection has low false positive rate, but cannot detect novel attacks. However, anomaly detection can detect unknown attacks, but has high false positive rate.

An intrusion detection system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches. In other words, intrusion detection is the act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a system/network. Traditionally, intrusion detection systems have been classified as a signature detection system, an anomaly detection system or a hybrid/compound detection system. A signature detection system identifies patterns of traffic or application data presumed to be malicious while anomaly detection systems compare activities against a "normal" baseline. On the other hand, a hybrid intrusion detection system combines the techniques of the two approaches. Both signature detection and anomaly detection systems have their share of advantages and drawbacks. The primary advantage of signature detection is that known attacks can be detected fairly reliably with a low false positive rate. The major drawback of the signature detection approach is that such systems typically require a signature to be defined for all of the possible attacks that an attacker may launch against a network. Anomaly detection systems have two major advantages over signature based intrusion detection systems. The first advantage that differentiates anomaly detection systems from signature detection systems is their ability to detect unknown attacks as well as "zero days" attacks. This advantage is because of the ability of anomaly detection systems to model the normal operation of a system/network and detect deviations from them. A second advantage of anomaly detection systems is that the aforementioned profiles of normal activity are customized for every system, application and/or network, and therefore making it very difficult for an attacker to know with certainty what activities it can carry out without getting detected. However, the anomaly detection approach has its share of drawbacks as well.

II TYPES OF IDS TECHNIQUES:

There are two primary approaches to analyze events to detect attacks, namely misuse detection and anomaly detection. Misuse detection is based on the extensive knowledge of known attacks and system vulnerabilities provided by a human expert, looking for hackers who attempt to perform these attacks and/or to exploit known vulnerabilities. Although misuse detection can be very accurate in detecting known attacks, it cannot detect unknown and emerging cyber threats this shortcoming makes them vulnerable to the reactivity of attackers. In other words, when attackers change their behavior in response to detection techniques, these techniques become useless and need major redesign. One solution for this problem would be to use adaptive approaches which are inherently designed to be resilient to small changes in the environment and adapt easily. On the other hand, anomaly detection is based on the analysis of profiles that represent normal behavior of users, hosts, or network connections. Anomaly detectors characterize normal "legitimate" computer activity using different techniques and then use a variety of measures to detect deviations from defined normal behavior. The major benefit of anomaly detection algorithms is their potential to recognize unforeseen attacks. However, the major limitation is the possibly high false alarm rate. Note that deviations detected by anomaly detection algorithms may not necessarily represent actual attacks as they may simply be new or unusual but still legitimate network behavior. Anomaly detection techniques fall into the following five groups: statistical methods, rule-based methods, distance-based methods, profiling methods, and model-based approaches. It should be mentioned that many IDSs, such as snort, use both misuse detection and anomaly detection to benefit from their respective advantages. There are two general categories of intrusion detection systems (IDSs): misuse detection and anomaly based. Misuse detection systems are most widely used and they detect intruders with known patterns. The signatures and patterns used to identify attacks consist of various fields of a network packet, like source address, destination address, source and destination ports or even some key words of the payload of a packet. These systems exhibit a drawback in the sense that only the attacks that already exist in the attack database can be detected, so this model needs continuous updating, but they have a virtue of having very low false positive rate. Anomaly detection systems identify deviations from normal behavior and alert to potential unknown or novel attacks without having any prior knowledge of them. They exhibit higher rate of false alarms, but they have the ability of detecting unknown attacks and perform their task of looking for deviations much faster. Application and development of specialized machine learning techniques is gaining increasing attention in the intrusion detection community. Soft computing is a collection of methodologies, which aim to exploit tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness and low solution, cost. As soft-computing techniques can also be used for machine learning, different soft-computing techniques have been used for intrusion detection (Fuzzy Logic, Artificial Neural Networks, Genetic Algorithms), but their possibilities are still under-utilized. In this work we have realized a misuse detection system that is based on Genetic Algorithm (GA). We have exploited both possibilities, either to classify network traffic as normal or abnormal, or to further classify the attacks by their type. Many features of GA make it very suitable for intrusion detection. Like robustness to noise, self learning capabilities and the fact that initial rules can be built randomly so there is no need of knowing the exact way of attack machinery at the beginning. Further classification of the attacks is not very important for intrusion detection, but it is important for network forensics because knowing the exact type of a threat and the way it performs its attack, the recovery after an attack would be more successful.

III PROBLEM STATEMENT

The environment in which the feature extraction is done is a mobile operator's network with real people using it. This means that the network traffic contains user confidential information. For example in Finland user network traffic is protected by the data protection law. Because of this, only a limited analysis for the network traffic can be done, meaning that a deep packet analysis cannot be done. In general, only the header fields of the packets can be checked but not the user data in the payload. Scalability is an issue with IDS. Because of the huge amount of data flowing through the mobile operator's network, it is not an easy task to find out the right information needed for IDS. The problem is to find an answer to the question: "What features need to be taken into account when calculating or analyzing whether the activity is malicious or not?" Based on prior research on IDS it is clear that either one of the techniques alone cannot detect everything but the combination of the both is the most promising approach. For example misuse detection can be used to filter known threats from the traffic to make it easier for the anomaly detection system to focus on the unknown. Even though IDS have been researched over 20 years, we still do not have an answer to the question of what features should be monitored. So far different kinds of methods and algorithms have been developed for anomaly detection but the focus has been on making them more efficient. Almost all of them are lacking the same information; what features are important for IDS, especially in telecommunications networks? For some reason information on the used features is not easily found from IDS research publications. No matter what the reason is the result is the same; every researcher has to figure out by themselves which features should be used for the monitoring.

1. The pre-processing of KDDCUP99 takes more time.
2. The rate of false alarm generation is high.
3. Some clustering technique is used such as k-means and genetic algorithm
4. Entropy based intrusion detection system suffered by high false rate
5. The detection of dynamic feature evaluation as normal data.

IV PROPOSED WORK

In this section discuss the proposed algorithm for intrusion detection. The proposed algorithm is combination of k-means algorithm and particle of swarm optimization algorithm. The process of seed selection is done by particle of swarm optimization.

The proposed algorithm of intrusion detection describe as

Step 1 initialized the rand function as artificial particle and the range of particle is range of data. The state of particle and velocity of particle select as random fashion. Select random particle as cluster center. Here describe s

$$X_i^{(0)} = \left(p_{i1}^{(0)}, p_{i2}^{(0)}, \dots\dots\dots\dots p_{ik}^{(0)}\right) \dots\dots\dots\dots\dots\dots\dots\dots..(4.5.1)$$

Where $p_{i1}^{(0)}$ refers to the jth cluster centroid in solution suggested by the ith particle. Now intelligence of swarm suggests the value of center point.

Step 2 estimate the fitness constraints of every particle on given clustering condition the fitness constraints define as

$$F(i) = \frac{\sum_{x=1}^{k}\sum_{cij}^{p}(yp-pij)^2}{T_p} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(4.5.2)$$

Where T_p is total number of data point proceeding for the clustering?

Step 3: the total number of iteration of clustering technique is maximized go to Step 7, if it is minimum go to next step

Step 4: The value of Pbest and Gbest stored in swarm search space and otherwise estimate with equation 4.5.1 and equation 4.5.2

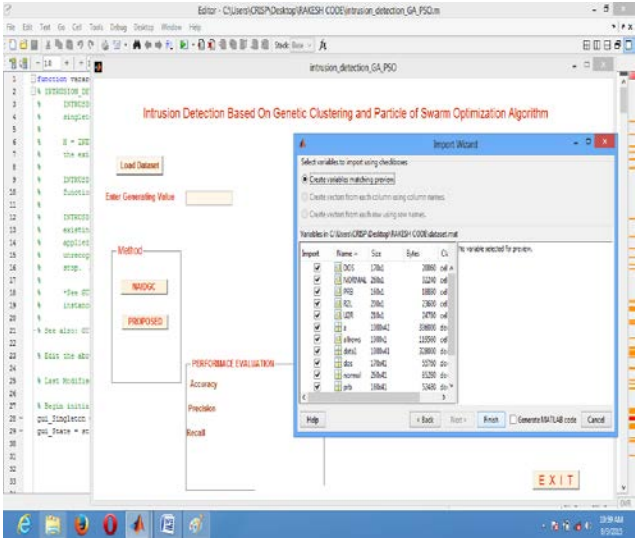
Step 5 minimized the eight value of particle W.

Step 6: If the value of Gbest is constant, unchanged for a number of iterations go to Step 7 otherwise go to Step 3.

Step 7: Use the k-means algorithm to finish clustering task. The clustering terminates when one of conditions meet according to fitness function.

V RESULT ANALYSIS

THE PROPOSED INTRUSION DETECTION ALGORITHM IMPLEMENTED IN MATLAB 7.8.347. MATLAB IS A STRONG MATHEMATICAL TOOL WHICH PROVIDES HELP TO ENGINEERS TO SOLVE, MODEL, SIMULATE THE PROBLEMS AND FIND SOLUTIONS ASSUMING ENVIRONMENT IN TO MATHEMATICAL EQUATIONS.


Editor - C:\Users\CRISP\Desktop\RAKESH CODE\intrusion_detection_GA_PSO.m
intrusion_detection_GA_PSO
Intrusion Detection Based On Genetic Clustering and Particle of Swarm Optimization Algorithm
Load Dataset
Enter Generating Value
Method
NAIDGC
PROPOSED
PERFORMANCE EVALUATION
Accuracy
Precision
Recall
EXIT
Import Wizard
Help
< Back
Next >
Finish
Generate MATLAB code
Cancel

Select variables to import using checkboxes

● Create variables matching preview

○ Create vectors from each column using column names.

○ Create vectors from each row using row names.

Variables in C:\Users\CRISP\Desktop\RAKESH CODE\dataset.mat

No variable selected for preview.

Import	Name	Size	Bytes	Cl
☑	DOS	170x1	20060	cell
☑	NORMAL	290x1	33240	cell
☑	PRB	150x1	18880	cell
☑	R2L	250x1	23600	cell
☑	U2R	250x1	24790	cell
☑	a	1080x41	336800	do
☑	allrows	1080x1	110560	cell
☑	data1	1080x41	328800	do
☑	dos	170x41	55790	do
☑	normal	250x41	85290	do
☑	prb	150x41	53480	do

Figure 1: Shows that the main windows for implementation of intrusion detection system to load of data set.

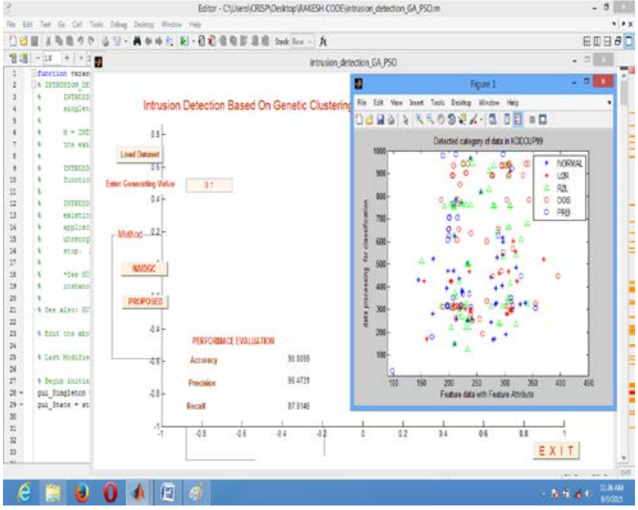

Editor - C:\Users\CRISP\Desktop\RAKESH CODE\intrusion_detection_GA_PSO.m
intrusion_detection_GA_PSO
Intrusion Detection Based On Genetic Clustering
Load Dataset
Enter Generating Value
0.1
Method
NAIDGC
PROPOSED
PERFORMANCE EVALUATION
Accuracy
90.8099
Precision
90.4731
Recall
87.9146
EXIT
Figure 1
Detected category of data in KDDCUP99
data processing for classification
Feature data with Feature Attribute
NORMAL
U2R
R2L
DOS
PRB

Figure 2: Shows that the intrusion detection system enter generating value of 0.1 image for the method NAIDGC and find the value of accuracy, precision and recall.

Input value	Method	Accuracy	Precision	Recall
0.1	NAIDGC	90.8099	90.4731	87.8146
	PROPOSED	97.3994	91.5004	90.3957

Table 1: Shows that the comparative results enter of the value 0.1 image for the NAIDGC and PROPOSED method and finds the value of accuracy, precision and recall

Input Value	Method	Accuracy	Precision	Recall
0.2	NAIDGC	92.4517	92.1149	89.4565
	PROPOSED	99.0412	93.1422	92.0375

Table 2: Shows that the comparative results enter of the value 0.2 image for the NAIDGC and PROPOSED method and finds the value of accuracy, precision and recall.

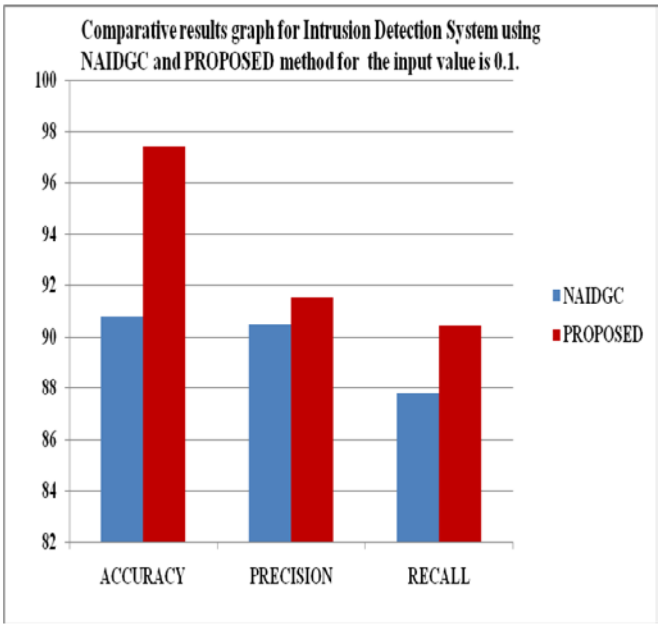


Figure 3: shows that comparative result analysis of intrusion detection system and NAIDGC and PROPOSED method.

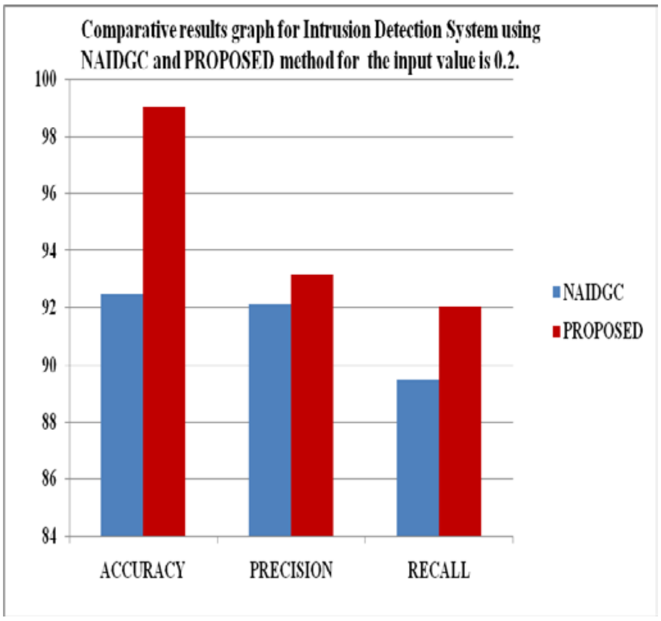


Figure 4: shows that comparative result analysis of intrusion detection system and NAIDGC and PROPOSED method .

VI CONCLUSION

In this paper proposed intrusion detection technique based on particle of swarm optimization along with k-means clustering technique. The modified clustering algorithm perform better clustering task instead of k-means-GA. Particle of swarm optimization control the iteration and also provide the facility of center point selection. The proposed clustering technique implemented in MATLAB 7.8.0 computational software and tested with valid intrusion dataset KDDCUP99. The KDDCUP99 dataset consist of five categories of data such as denial of service attack, prob attack, U2R attack, R2L attack and finally normal data connection for network traffic data.

Compared the results of the proposed method with some of commonly used clustering technique, the standard k-means-GA methods used as baseline methods are the big-bang and Local cluster group. The proposed method of intrusion detection clustering improved the cluster validation of attack data during the process of clustering. A hybrid technique of intrusion detection clustering suffered a problem of low range data.

REFERENCES:-

[1] Huiling Guo, Weichen, Fang Zhang”Research of Intrusion Detection based on genetic clustering algorithm” IEEE, 2012, Pp 1204-1207.

[2] Feng Du ”An effective pattern matching algorithm for intrusion detection” ICCSE 2012, Pp 34-38.

[3] Prasanthi S, Sang-Hwa Chung and Won-Suk Kim ”An enhanced TCP scheme for distinguishing non-congestion losses from packet reordering over wireless mash networks” IEEE, 2011, Pp 440-447.

[4] Xie Yong He Fubao, Zhang Yilai" A descending suffix tree based pattern matching algorithm for intrusion detection" IEEE, 2012, Pp 21-28.

[5] Yang Li, Li Guo" An active learning based TCM-KNN algorithm for supervised network intrusion detection" IEEE, 2007, Pp 459-46.

[6] Yang Lia,d, Jun-Li Wangb, Zhi-Hong Tiand, Tian-Bo Luc, Chen Youngc "Building lightweight intrusion detection system using wapper based feature selection mechanisms" IEEE, 2009, Pp 466-475.

[7] Hu Han "performance improvement of TCP reno based on monitoring the wireless packet loss rate" IEEE, 2011, Pp 469-472.

[8] Shen li "An efficient architecture for network intrusion detection based on ensemble rough classifiers" ICCSE, 2013, Pp 1411-1415.

[9] Zhenwei Yu, Jeffrey J.P.Tsai "An automatically tuning intrusion detection system" IEEE,2007, Pp 373-384.

[10] Mohammad saniee Abadeh and Jafar Hakibi "Computer intrusion detection using an interactive fuzzy rule learning approach"IEEE,2007, Pp 2345-2351.

[11] Cichen Shingo mabu, Chuan Yue, Kaoru Shimeda and kotoro Hirasawa "Network Intrusion Detection using fuzzy class association rule mining based on genetic network programming" IEEE,2009, Pp 60-67.

[12] Ambareen siraj, Susanm Bridges, Rayford B.Vaughu"Fuzzy cognitive maps for decision supporting an intelligent intrusion detection system" IEEE, 2012.

[13] Mansour Sheikhan , Zahra Jadidi "Misuse detection using hybrid of association rule mining and connectionist modeling" world applied science journal,2009, Pp 31-38.

[14] R. Shanmugavadivu,Dr.N-Nagarajan "Network intrusion detection system using fuzzy logic" IEEE2011, Pp 101-111.

[15] Perminder kaur, Dhavlesh Rattan, Amit kumar bhardwaj" An analysis of mechanism for making IDS fault tolerant" International journal application, Vol-10, 2010, Pp 22-25.

[16] Muna Elsadig Mohmed, Brahim Belhaovari Samir, Azween Abdullah "Immune multi-agent system for network intrusion detection using non-linear classification algorithm" International journal computer application, Vol-12, 2010, Pp 7-12.

[17] Sunita Patel, Jyoti Sondhi, Anand Motvani, Anurag Shrivastava "Improved Intrusion Detection Technique based on Feature Reduction and Classification using Support Vector Machine and Particle of Swarm Optimization" International Journal of Computer Applications, Vol-100, 2014. Pp 34-37.

[18] Alberto Fernandez, Maria Jose del Jesus and Francisco Herrera "On the influence of an adaptive inference system in fuzzy rule based classification system for imbalanced data-sets" in Elsevier Ltd. All rights reserved 2009.

[19] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E.Vazquez "Anomaly-based network intrusion detection: Techniques, Systems and challenges" in Elsevier Ltd. All rights reserved 2008.

[20] Terrence P. Fries "A Fuzzy-Genetic Approach to Network Intrusion Detection" in GECCO 08, July12–16, 2008, Atlanta, Georgia, USA.