NOVEL APPROACH OF INTRUSION CLASSIFICATION BY HYBRID GENETIC AND PARTICLE SWARM OPTIMIZATION

Er. Komal Saklani[1], Er. Poonam Chaudhary[2]
[1, 2] *Computer Science and Engineering, SIRDA Group of Institutions*

Abstract- Network security is becoming an essential need of modern society to protect the confidential information flowing over the networks. Detection of Intrusion over the network is one the most extremely important task to prevent their unlawful use by the attackers [1]. Efficient intrusion detection is needed as a defense of the network system to detect the attacks over the network. A feature selection and classification based Intrusion Detection model is presented, by implementing feature selection, the dimensions of NSL-KDD data set is reduced then by applying machine learning approach, we are able to build Intrusion detection model to find attacks on system and improve the intrusion detection using the captured data. With the increasing number of new unseen attacks the purpose of this model is to develop a system for intrusion detection, and the model will be capable of detecting new and previously unseen attacks using the basic signatures and the features of known attacks.

Keywords- Intrusion Detection System,

I. INTRODUCTION

In the present scenario the use of internet is growing at a large pace with is highly developed and emerging forms of ever growing network and its connectivity but the use of internet poses a great threat to cyber security. In order to maintain the high level of security there is an important need to overcome the cyber threats posing problems to various organizations, companies, and the firms. One of the major challenges among the cyber-security is to maintain the integrity of the intrusion detection system (IDS) thereby protecting it from major forms of attacks and to conquer the various form of risks of the intruded system [1] [3]. The main function of the IDS is to identify a more precise form of intrusion. The illegal hackers of the security have found a large number of ways to break the security of the system whether it is a cloud network or the wireless-based network. Many researches have been performed by the technologists to curb the security threats from distinct forms of intrusions done to the cloud computing systems and the wireless system. So, the main objective of IDS is to protect the information whether it is governmental, public or private entity [10]. The use of IDS is mainly required in detecting the false and the poor detection rates. Whenever an attack is observed by the system or a harmful activity is done to the system, it automatically generates an alarm resulting in a false-positive alarm [3] [5]. The research mainly focusses upon the enhanced capabilities of the intrusion detecting system and thereby reduces the occurrence of the false type alarms.

1.1 Overview

The main requirement of the IDS is not only to encounter the intruders in the data path but also to supervise the intruders of the data. The most important security aspects of an intrusion detection system consist of maintaining the following conditions.

(a) Confidentiality: Only an authorized user can detect the system.

(b) Availability: Here, the computer technology provides various forms of resources and the access to the legal users of the system without disturbing the working operation of the system.

(c) Integrity: The information must be protected from any kind of malicious act.

The following figure.1 explains the general structure of an intrusion dectection system.

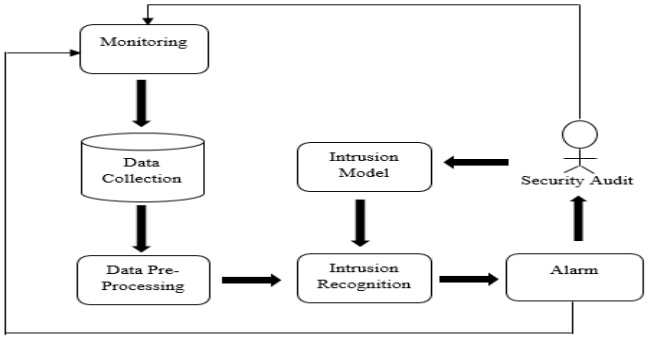


Fig.1.1 Basic structure of IDS

In order to increase the performance of the IDS, a method known as soft computation is done.The term “soft computing” refers to the process of different methods to get the best possible finite results. The eminent technology of Artificial Intelligence and the machine learning processes has resulted in accuracy anfd thereby providing the best suitable results as per the requirement. It has shown a great success in the IDS mechanism

[4]. The are various distinct forms of soft computing methods used in IDS detection such as Support Vector Machine [SVMs], Artificial Neural Network [ANNs], Genetic Algorithms [GA], Bayesian Networks, and Fuzzy Logic. In case of human eyes the reseachers use the AI techniques to identify the intrusions that is the main reson why the researchers use the data mining processes and the artifial intelligent techniques to explore the feasible intrusions.

1.2 IDS: Architecture

The architecture of IDS comprises of its unique core element i.e. sensor popularly known as the analyzing engine to pin-point the intrusions occurring in the system. The sensor consists of a mechanism that helps in detecting the intrusions. In the following figure.3 the sensor gets the data (raw) from the given sources as shown which consists of the audit trails, knowledge-based data and, syslog. The 'syslog' includes the authority to the particular system or the system file configuration [1] [2].

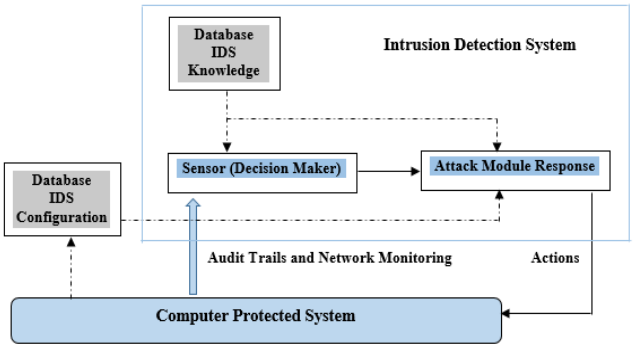


Fig1.2 Sample IDS (arrow width ∞ information between system components)

The sensor consists of a component known as event generator which performs the data collection shown in figure.3. It detects the way of collecting the data. The event generator consists of network, operating system and the network applications where it generates a set of events including audit (log) of the system or the packets of the network. This form of set events also involves the policy of information collection i.e. in or out of the system. Sometimes it is not necessary to store the data as it reaches simply to the analyser. So, basically the key role of the sensor is to extract or filter the data and remove the unwanted form of the data that is achieved from the event data set system [6] [13]. Additionally, the database holds the configurational parameters of IDS that includes its mode of communication methods based on the response module. The sensor itself contains its own data observing all the historical multiplex forms of intrusions. Practically, the IDS may follow a structure based on an 'agent' principle where small modules (autonomous) are designed on 'per-host' basis approach. The agent mainly monitors and filters the activities scheduled within the area i.e. fully protected and further starting its initial analysis by undertaking a response action [10]. The Interfacing of the IDS results in linking or providing the interactions between its components. These can be saved for a long period of time but the monitoring process requires synchronization of these components.

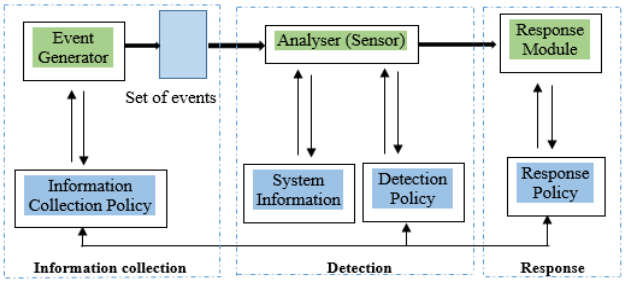


Fig1.3 IDS components

1.3 Data Mining Techniques

The processing of data from the different sources results in gigantic data-sets that cannot be analysed properly [9]. So, by analysing the sources of data-set, the data-mining techniques plays a significant role in revealing the hidden data source and the normal or abnormal forms of patterns. This particular section states the different forms of data-mining techniques in order to detect the various forms of attack observed in the network [11].

1.3.1 Association rules: This is method which identifies the connection or association between the variables in large amount of data-sets, association among the data attributes and helps in determining the system values. As the nature of this rule is based on pattern discovery so, we cannot rectify the problems related to classification and prediction. In association rule mining process two of the threshold values are considered. One is the maximum support and the other is the minimum confidence

1.3.2 Classification: When each sample of data set is assigned to a unique form of class then it is termed as the process of classification. Generally, it is used for signature-based technique but it can also be used for anomaly-based detection technique. In this type of technique, firstly, the datasets which are available are predefined. There are various types of classification techniques as explained below:

(a) Decision Tree: It is well known recursive method forming a structure like a tree. Here, the divide and conquer methods are adapted for segregating the attribute value. The process of classification starts from the root-node towards the path of the leaf node. The root-node denotes the values of the attribute whereas the leaf node denotes the class-label. A large set of data tree gives the excellent performance rate.

(b) *ID3 Algorithm:* It is an algorithm based on attributes creating a decision-tree on the basis of trained data-sets. It is used in natural as well as the machine leaning methodologies. The mechanism of ID3 helps in constructing the information and the entropy gains to design a decision tree.

(c) *J48 Algorithm:* This is a form of C4.5 algorithm which constructs a decision tree based on the information gain of an attribute denoting the high level gain. But the disadvantage of using this algorithm is that it require more time for central processing unit to run and needs a huge space for memory [7]. In, J48-algorithm, set of rules are produced by analyzing decision- based tree

(d) *NB Algorithm:* It uses both the classifier methods i.e. the Naïve Bayes and the decision tree methods. Naïve Bayes is used in leaf nodes and the root-node uses a classifier based on decision tree.

(e) *Random Forest:* This technique is based on random analysis where each tree is designed by distinct data-sets on random based selection. A high-quality dimensional data can be handled easily in this form of method [8, 12].

(f) *K-Nearest Neighbor:* This represents a simple form of classification technique where it describes the distance among different data points and locates the data points that are not labelled. Ro its nearest neighboring class. It is based on the some of the important conditions i.e. if 'm' denotes the value equal to one, then object gets simply assigned to its neighboring value. But if the value of 'm' is large then its prediction is very difficult in such case.

(g) *Naive Bayes classifier:* This a probability-based classifier method with the assumption based on the membership probability. It works typically on the relation among variables i.e. dependent and independent variables that derives the probable conditions: -

$$P(H/X) = \{(P(X/H).\ P(X))\}/\ P(H) \quad \ldots\ldots (i)$$

where,

X = recorded data

H = hypothesis

P (H) = prior probability

P (H/X) and P(X/H) = posterior probability

The Naive-Bayes classifier can be easily designed without the use of iterative complex parameters.

(h) Support Vector Machine: It is generally used for the process of classification and prediction. It represents the two main classes of data-points using the method of hyperplane which denotes the +1(normal-data) and -1(suspicious-data) values [12, 14]. The hyperplane condition is stated as below:

$$(W^* X) + b = 0 \quad \ldots\ldots(ii)$$

Where,

W (weight vector) = w1, w2………………………… wn

X (attribute values) = x1, x2………………………… xn

b = a scalar

Here, the main objective of SVM is to use some part of data to train the system and to identify linear-optimal hyper-plane in order to maximize the gap between the margins of separation [14].

1.3.3 Genetic Algorithm: Genetic algorithm represents a best technology for data-mining technology that selects can hold the information from a vast collection of data or a data-box, further finding the different operating modes to gather the accurate results. This is based on the theory of natural evolution. The fitness function evaluates the quality of each and every rule [8].

1.3.4 Neural Network: The term neural network represents a paradigm for the process of information system i.e. based on working of the biological nervous systems. It represents a set of elements that are processed highly consisting of linked or interconnected nodes which produces an alteration to the input-nodes creating the desired form of output, where every node is connected such that it forms an adequate connection in its neighboring-layers.

1.3.5 Markov Model: This method is based on the approaches of learning techniques. Here, the states that are definite in nature in HMM i.e. Hidden Markov Model are controlled by the transition-probability sets. After, the probability-distribution mechanism, and output gets generated and this process repeated again and again till the desired results are not achieved. The HMM uses it calling methodology to detect the intrusions of the system. Hidden Markov Model (HMM) is also used to detect intrusions using the system calls.

1.3.6 Hierarchical Clustering: The main advantage of using such kind of clustering is that it is very helpful in dealing with noise based applications. It possesses efficient memory and provides a high standard quality of clustering at a minimal cost.

1.*3.7 K-Mean Clustering:* This is most extensive form of clustering algorithm and depicts an easy and simple way to deal with different processes. The first step is the identification of number of clusters 'k' that are stated to distribute the samples or instances into a number of clusters that are pre-defined. The first method is to select the 'k' samples denoted as clustering center. Secondly, each and every instance gets assigned to its nearest cluster.

II. RELATED WORK

Dias GV et.al [1] conducted a study indicated an intrusion detection system based on SVM methodology that combines an algorithm (hierarchical clustering), feature selection method and the technique of SVM. The algorithm i.e. used helps in providing the support vector machine with maintaining an abstracted form of high level of trained examples obtained from the trained set-up of KDD Cup 1999. The study indicates high level performance of SVM based technology which further resulted in a reduced form of training-time. The method of feature-based selection was adopted to remove the un-necessary features of the training set in order to maintain the levels of accuracy. So, the methodology based on this dataset showed better analysis in detection of probe and DoS based attacks, maintaining accuracy globally. Cannady et.al [2] proposed a study on the process of misuse detection which is defined as a process to recognize the instances of different types of attacks by measuring the unexpected activity and the activity that is going currently. Mostly, the present processes based on misuse detection uses a technology of rule-based systems with

the aim to identify the provoked nature of the attacks known to us. Kemmerer et.al [3] presented a study by framing a simple question of why there is a need of intrusion detection system. Suppose, the owner of a house is out of town and he has locked his home with all the windows and doors closed. But, there is someone outside his home who wants to enter. The reason to install these detective systems is that the intrusions still exist because sometimes the people may forget to lock their doors or windows, the same case occurs with the computer based networks which do not provide us 100% security of the system to work accurately. So, based on this study the researchers has tried to explain the techniques based on IDS to deal with these kind of intrusions present in the network. Steven T et.al [4] proposed a study on an application of STATL that represents a descriptive language based on a transition-based attacking system that is constructed to support the IDS. This form of descriptive language describes a process of penetration done to the computer network implemented by a hacker. These type of penetrations includes attacking activities performed by the hacker. There is a deep study of syntax based on the STATL language. Common real examples of both the network and the host are also described in the paper. Pi-Cheng et.al [5] conducted a research based on two of its issues related to the IDS designs. The two issues include the selection based on optimization of rule-based selection and the discovery in case of attack. This type of approach provides a connection between the junked packets. An algorithm is implemented for the attack identification and the rule based selection. Cavusoglu et.al [6] conducted a research on security systems of IT. The information technology firms rely on various forms of technologies such as IDS and the firewalls to manage the risks of the organizations. There exists some most interesting facts related to security alerts in IT industries. This paper presented a study to demonstrate the values of IDS adopted in an IT company. The configuration of IT was represented by the true-positive and the false-positive rates which further consists of determining the negative or the positive rates of an organization. Kim, Jungwon, et.al [7] conducted a research on the use of artificial immune systems in IDS which is an interesting concept that relied on two main reasons. Here, the researchers have used various distinct algorithms for the development of the systems and the best possible outcomes. The analysis has been done based on the important developments within this area of research, in addition to forming suggestions for future research options. Zhang, J., et.al [8] proposed new frameworks that involved the use of a data mining algorithms such as the hybrid-network-based IDSs, random-forests in misuse, and an anomaly based detection. The hybrid mechanism has improved the performance of detection with the combination of misuse advantages. The results demonstrate that the use of misuse detection approach was much better than the best KDD'99 data-set approach that provided low false rate, high amount of detection rate that resulted in an overall increased performance of the IDS system. Muhammad Hilmi Kamarudin, et.al [9] proposed their study on technology of network security that has become a supreme method for the protection of information or the data. With the excessive growth of internet technology, various forms of attack cases are observed in a day to day life. So, to tackle such kind of attacks, a methodology of Intrusion Detection System (IDS) is adopted and the process of Machine Learning is the most used technology in the IDS. The study based on recent years has shown that the Machine Learning Intrusion Detection system provides a good detection rate and a high accuracy. Muamer N., et.al [10] conducted a study on using smart and intelligent form of data-mining approaches to observe the intrusion occurring in the local-networks. This paper suggested an improved strategy for Intrusion Detection System (IDS) that combines the expert systems, the processes of data mining as implemented in WEKA. The classification generally consists of the detection principle as well as some of the aspects of WEKA such as open-source data-mining processes. The combining methodology gives better performance of IDS based systems, and helps to maintain the detection more effectively. The result was based on evaluating a new design produced a better form of detection based on efficiency. So, the study presented a good approach to analyse the experiments on behalf of intrusion detection. Nadiammai, et.al [11] focused upon the security issue of the networks and various developments in applications running on distinct platforms capturing an attention towards security of the network. This type of paradigm exploited the vulnerabilities of security that are technically difficult and expensive to solve. In this work, data mining concept is integrated with an IDS to identify the relevant, hidden data of interest for the user effectively and with less execution time. Four issues such as Classification of Data, High Level of Human Interaction, Lack of Labeled Data, and Effectiveness of Distributed Denial of Service Attack are being solved using the proposed algorithms like EDADT algorithm, Hybrid IDS model, Semi-Supervised Approach and Varying HOPERAA Algorithm respectively. All the proposed algorithm shows better accuracy and reduced false alarm rate when compared with existing algorithms. M. A. Jabbar, et.al [12] proposed the research based on the intrusion detection system to notify and identify the type of activities or normal users or the hackers performing malicious operations. The IDS represents complicated and a linear problem dealing with traffic-data of the network. Many forms of IDS classes have been developed and proposed which further produced distinct levels of accuracy with the aim to maintain a robust and effective Intrusion detection system that is a necessary requirement. In this paper, a model has been designed for intrusion detection system (IDS) using a classifier based on random forest where, the Random Forest (RF) denoted an ensemble classifier and that performed very well as compared to the other classifiers that worked traditionally for an effective

classification of different forms of attacks. Aafreen K. et.al [13] proposed a work using the IDS tool for anomaly detection that provides network security to the system. The IDS represents a method to detect the processes of cyber-attacks and this process of detection is based on the amount of distinct forms of intrusive activities occurring in the operation of the system as the detection of an intrusion denotes a very complicated process. Some of the attacks are known while some of them are not known. The detection process of a known attack is not a difficult task as it can use a rule based or signature-based method but to pin-point an unknown attack is very challenging process.

III. THE PROPOSED METHOD

3.1 Proposed Methodology

The author has proposed a hybrid model which consists of SVM i.e. Support Vector Machine combined different classification-algorithm to mitigate the rates of the false-positive alarms. To obtain the pre-thesis objective a methodology has been proposed which is further divided into three types of phases.

Phase 1: Collection and preprocessing

- Data-set collection
- Extraction of features through a data i.e."tcpdump"
- Converting the obtained features into binary representation
- Preparation of the input for its classification

Phase 2: Classification

- To find the best classifier from the available classifier.
- To test and train the tool of classification by the dataset-partitioning process.

Phase 3: Result analysis

- To compare the obtained results with their existing work.

The proposed working methodology is designed as below in figure.3.1

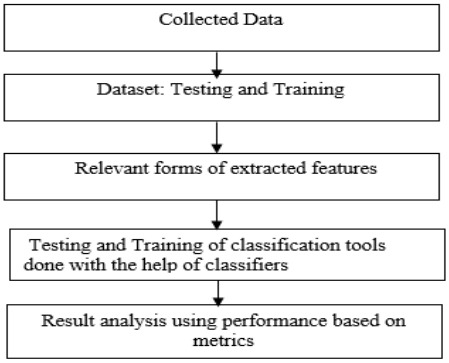


Fig 3.1 Proposed Methodology

In order to start proposed work. The first step is to study all the data-set obtained from the different sources, to eagerly check the data format of the data and further to analyze that which form of mining technique should be applied over the data. When the set of data gets collected then the process of feature extraction is carried out. Further, the process is will be classified into two classes. The first one is testing and the other one is training of classification tools i.e. done with the help of classifiers. Then the results is further analyzed based on various forms of performance metrics.

3.2 Proposed methodology: Flowchart

The proposed steps of flow chart are given as below:

1. KDD-99 Data Set: This is a type of data-set used for the (Third International Knowledge Discovery and Data Mining Tools) Competition, held in conjunction with KDD-99 (The Fifth International Conference on Knowledge Discovery and Data Mining). The main task was to build a network based on intrusion detection, and to predict a model i.e. capable of discriminating a good or a bad form of data-set. This form of data-set maintains a standard including a wide variety of network based intrusions.
2. Label Features: A label helps in providing a complete information regarding the set of data.
3. Input in PSO: Each of the particle has its velocity and position to search for better solution. So, the velocity and position are the inputs used in PSO.
4. Initialize particles: The PSO-based technique is initialized with a population of random solution.
5. Update fitness function: It helps in judging the individual solutions based on how well they can handle the problem.
6. Optimize Objective Form: Here, the objective is optimized.
7. Initialize chromosomes: The process is initialized by building a population of chromosomes which is a set of possible solutions to the optimization problem.

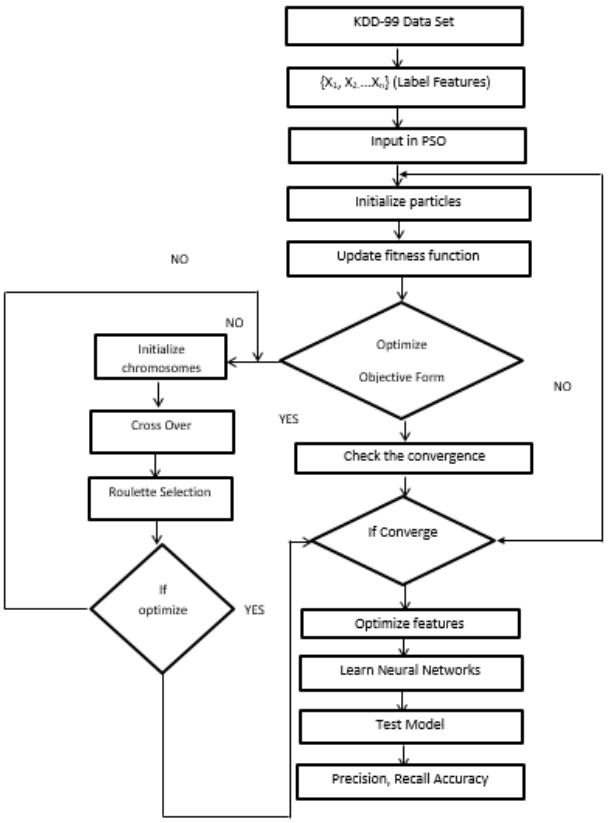


Fig.3.2 Proposed Flowchart

8. Check the convergence: These type of methods helps in testing the conditional-convergence, absolute-convergence, interval of convergence or divergence of an infinite series.
9. Cross Over: A point or place of crossing from one side to the other.
10. Roulette Selection: It is a method used in genetic-algorithms for selection of potentially useful solutions for the purpose of recombination.
11. Optimize features: This type of method achieves the best designing technique.
12. Neural Networks: It represents biologically inspired information processing system.
13. Test Model: It performs a system or software system.
14. Precision, Recall Accuracy: The precision is a good measure that determines the costs of False Positive is high.

3.3 Proposed Algorithm

Optimization is done by two algorithm PSO and learning approach. Initially feature is applied on PSO for optimization to obtain fitness value. If any feature is not optimized by PSO then these un-optimized feature is allow on learning algorithm for further optimization process. If all the features are optimized by these two algorithms then it further moves to check convergence. If convergence is not accepted then these feature again used for optimization by PSO and Learning algorithm. If convergence check is ok then we move to next phase which is learning.

Normal	Dos	R2L	U2R	Probe
Normal	Smurf Process table Pod Land	PHF Xlock Send-mail Guess_password	Root-kit Eject Perl Buffer overflow	Portsweep Satan Saint M-scan

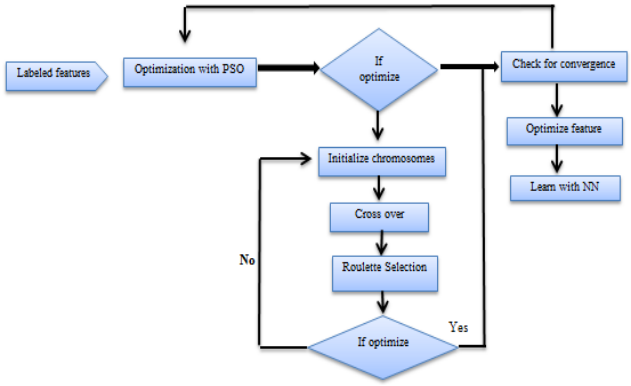


Fig.3.3 Optimization process

After labeling the dataset feature optimization is done by PSO stands for particle swarm optimization where swarm denotes to collection of particles. In the process of PSO Particles are float through the hyper-dimensional search space. PSO is a population based search algorithm which is based on simulation on the social behavior of birds within a flock. Variation in the position of particle in a search space is depending upon the psychological tendency of each particle to imitate the development of other. After optimization with Genetic algorithm if, all the features are optimizes they proceed to further step otherwise the optimization with GA is repeated. This procedure is repeated unless the entire features are optimized. At last after optimization every feature convergence is tally if it is ok then our feature are allow for learning by neural networks [13][14].

IV. RESULT ANALYSIS

4.1 Description of dataset: As discussed above experiments is executed by using KDD-99 which having 41 feature sets. These features are used for optimization and then learning and now they are used to analyze in terms of attack. In this work we use to evaluate the accuracy rate in an intrusion detection system. In the analysis we take data on the basis of number of intrusions. Attacks are generally fall into four categories 1) Dos, 2) Probe, 3) R2L 4) U2R. In our analysis we uses three categories 1) other attack which consist of probe, R2L and U2R 2) DoS-attack 3) Normal attacks (non-attacks). In this work we

evaluate the accuracy, precision, recall and F-measure in various cases:

Case 1: Evaluation of accuracy, precision, f-measure and recall is given by ANN individually, ANN with GA, ANN with PSO and ANN with combined GA_PSO which is represented in table 1.4. In this case we evaluate the efficiency of IDS by applying ANN individually or with GA and PSO or by hybrid of both algorithms with ANN.

Table.1. Evaluation Process

	ANN	ANN with GA	ANN with PSO	ANN with GA_PSO
Accuracy	89	92.23	92.34	94.23
Precision	88	89.23	90.32	92.33
Recall	87	88.56	91.26	96.33
F-measure	85	88.25	86.23	91.13

Case 2: In this case we evaluate the efficiency for single ANN on three attack condition a) other attack consist of probe, R2L and U2R b) Dos attack c) Normal or non-attacks condition. Similarly we evaluate the efficiency for ANN with PSO, ANN with GA and at last ANN with both GA_PSO. Here we evaluation efficiency of algorithms in terms of accuracy, precision, recall and F-measure. If we spot some light on attacks we are considered. Dos attack which stands for denial of service attack in Dos attack hidden attacks is done by user which is shown in the system. This type of attack may be done by single intruder or a group of intruders. It makes the system unavailable to its real user. Probe attack is a kind of attack where intruder used to break the security by trial methods. R2L attack stands for remote to user attack. And at last U2R attacks it is the type of attack where intruder starts on the system as a normal user and spoil all the activities of the systems.

Table.2 shows the attack type from KDD CUP 99 dataset

Table.3 shows the static data to analyze the efficiency of the approaches for above discussed attack.

Algorithm type	Types of attack	Accuracy	Precision	Recall	F-measure
a) ANN	Other attack	86.23	87.23	86	83
	Dos Attack	87.23	88.33	88	84
	Normal Attack	91.23	90.13	87	86
b)	Other attack	91.13	87.23	86.75	87.23
	Dos Attack	90.23	90.13	84.23	86.23
ANN with GA	Normal Attack	94.11	89.99	94.23	89.13
c) ANN with PSO	Other attack	93.13	89.23	92.23	84.23
	Dos Attack	92.13	88.13	89.13	83.23
	Normal Attack	90.13	84.23	92.13	87.34
d) ANN with GA and PSO	Other attack	95.62	90.23	92.33	89.23
	Dos Attack	89.23	89.23	93.33	90.13
	Normal Attack	96.23	93.13	99.23	90.23

1. In this section we analyze the statistical data by simulation. Graphical result of table 1 and 2 is given by simulation process.

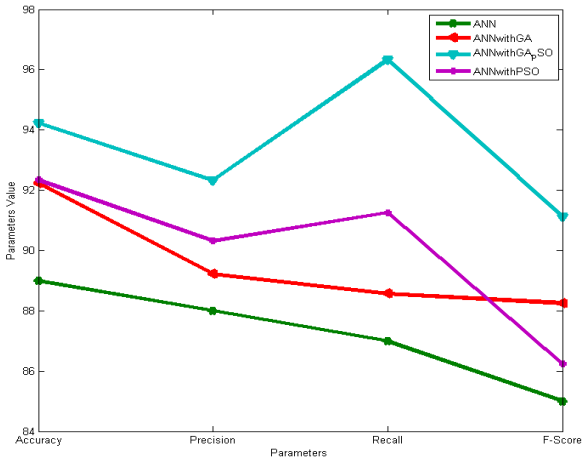
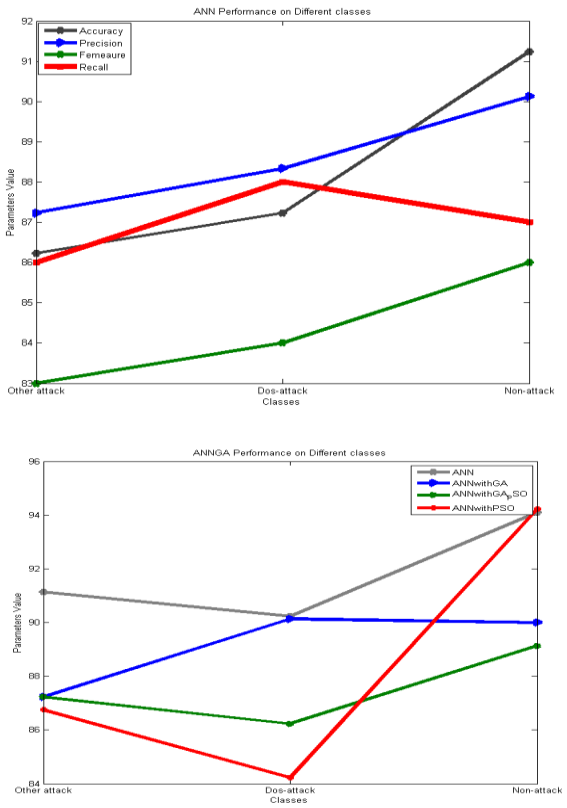


Fig.4.1 Simulated graph of table 1

Fig. 4.1 shows the simulated analysis of table 1 in terms of accuracy, precision, recall and f-measure. In this figure analysis on efficiency is demonstrated from all the four algorithm that are ANN represented by green line, ANN with PSO represented by purple line, ANN with GA represented by red line and ANN with both PSO and GA represented by blue line. Analysis demonstrates that ANN with both SPO and GA giver better result in terms of all the four parameters (accuracy, precision, recall, f-measure).



\

Fig. 4.2: Analysis with ANN and ANN _GA

1. Observation1: In fig. 4.1 parameters analysis of different classifier and proposed approach. In analysis parameters like precision, recall, accuracy and f measure vary according to classifier but one analysis very clear about proposed approach (PSO with GA in neural network) show significant improve all parameters.

Fig. 4.1 and 4.2 gives the analytic result in terms of accuracy (represented by black line) , recall (represented by red line), precision (represented by blue line) and F-measure (represented by green line) of two approaches that are ANN and ANN with GA for the parameters (Dos attack, other attack and Non or normal attack). The entire four graphs demonstrate the better efficiency of the algorithm ANN with PSO_GA.

2. Observation2: In fig. 4.2 depth analysis of all three classes in ANN and ANN_GA. In this analysis we try to show what the significance of our approach is. This discussion we continue in observation (3) also. So first point which analysis by normal class n which not any attack working and in both cases ANN and ANN with GA perform well compare to other parameter like precision, recall and f-measure but ANN_GA still better accuracy than ANN so feature weighted by optimization somehow perform because of reducing overlapping information learning. If analysis through DOS attack it also show higher accuracy in ANN with GA.so we can conclude Feature optimize weight is better approach so how can improve optimization these observation discuss in next part.

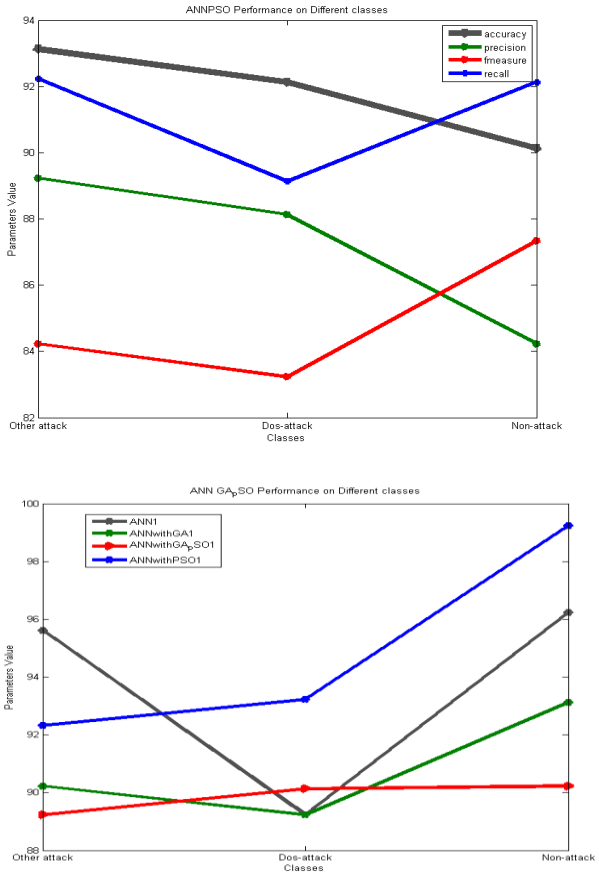


Fig.4.3 Analysis with ANN _PSO and ANN with PSO_GA

At last from the whole analysis it can be concluded that algorithm ANN with PSO_GA gives better result for all the attacks we examined in our work.

Observation3: In fig 4.3 analysis continue from observation (2) and try to finding significance of optimization improvement effect on different classes' detection by classification. If analysis the both graph show the effective recall but for normal class so reduce the false positive rate this improvement happening with all classes like DOS attack and other attacks but the effective result show in other attack which increase significantly in proposed approach. So PSO optimization is good but PSO with GA more improve in other attack and normal class.

IV CONCLUSION

The present scenario experiences various forms of developments and a huge growth in advanced processing technologies consisting

of connectivity among different networks but methodology is vulnerable by the activities of the intruders or the attackers of the system. These specifically smart attackers interrupt the operation with new and fascinating methods of data-breaching among large networks. Though there are various forms of available intrusion of intrusion detection systems that can detect the intrusions occurring in the network i.e. based on the false positive detection rate and the alert rates but with the detection rate of intrusions, they also have a high false-positive rate resulting in an adequate system comprising of low accuracy level of the system and are generally more prone to different kinds of attack. This usually helps the intruder to enter into the system and perform a pre-planned attack. So, this pre-thesis will propose a hybrid approach to reduce the false positive alarms. The experimental analysis consists of a specified particular form of data-set and the process of feature-based selection will be done to improve the analysis. These features obtained will be used for the classification-tool training and testing the performance of the system. Finally, the result obtained will be compared with the results that already exist.

V. REFERENCES

[1] Snapp SR, Brentano J, Dias GV, Goan TL, Heberlein LT, Ho CL, Levitt KN, Mukherjee B, Smaha SE, Grance T, Teal DM. DIDS (distributed intrusion detection system)-motivation, architecture, and an early prototype. In Proceedings of the 14th national computer security conference 1991 Oct (Vol. 1, pp. 167-176).

[2] Cannady, James. "Artificial neural networks for misuse detection." In National information systems security conference, vol. 26. pp. 368–381, 1998.

[3] Kemmerer, Richard A., and Giovanni Vigna. "Intrusion detection: a brief history and overview." *Computer* 35, no. 4 (2002): supl27-supl30.

[4] Eckmann, Steven T., Giovanni Vigna, and Richard A. Kemmerer. "STATL: An attack language for state-based intrusion detection." *Journal of computer security* 10, no. 1-2 (2002): 71-103.

[5] Hsiu, Pi-Cheng, Chin-Fu Kuo, Tei-Wei Kuo, and Eric YT Juan. "Scenario based threat detection and attack analysis." In *Security Technology, 2005. CCST'05. 39th Annual 2005 International Carnahan Conference on*, pp. 279-282. IEEE, 2005.

[6] Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan. "The value of intrusion detection systems in information technology security architecture." *Information Systems Research* 16, no. 1 (2005): 28-46.

[7] Kim, Jungwon, Peter J. Bentley, Uwe Aickelin, Julie Greensmith, Gianni Tedesco, and Jamie Twycross. "Immune system approaches to intrusion detection–a review." *Natural computing* 6, no. 4 (2007): 413-466.

[8] Zhang, J., Zulkernine, M., & Haque, A. (2008). Random-forests-based network intrusion detection systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *38*(5), 649-659.

[9] Jalil, Kamarularifin Abd, Muhammad Hilmi Kamarudin, and Mohamad Noorman Masrek. "Comparison of machine learning algorithms performance in detecting network intrusion." In *Networking and Information Technology (ICNIT), 2010 International Conference on*, pp. 221-226. IEEE, 2010.

[10] Mohammad, Muamer N., Norrozila Sulaiman, and Osama Abdulkarim Muhsin. "A novel intrusion detection system by using intelligent data mining in weka environment." *Procedia Computer Science* 3 (2011): 1237-1242.

[11] Nadiammai, G. V., and M. Hemalatha. "Effective approach toward Intrusion Detection System using data mining techniques." *Egyptian Informatics Journal* 15, no. 1 (2014): 37-50.

[12] Farnaaz, Nabila, and M. A. Jabbar. "Random forest modeling for network intrusion detection system." *Procedia Computer Science* 89 (2016): 213-217.

[13] Siddiqui, Aafreen K., and Tanveer Farooqui. "Improved Ensemble Technique based on Support Vector Machine and Neural Network for Intrusion Detection System." *International Journal Online of Science* 3, no. 11 (2017).

[14] Wang, Huiwen, Jie Gu, and Shanshan Wang. "An effective intrusion detection framework based on SVM with feature augmentation." *Knowledge-Based Systems* 136 (2017): 130-139.