International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Anomaly Intrusion Detection System in Real Time Environment using Ensemble Learning Model

Sharmila. K. Wagh, Anuradha S. Varal

Abstract*: *System security is of essential part now days for huge organizations. The Intrusion Detection System (IDS) are getting to be irreplaceable for successful assurance against intrusions that are continually changing in size and intricacy. With information honesty, privacy and accessibility, they must be solid, simple to oversee and with low upkeep cost. Different adjustments are being connected to IDS consistently to recognize new intrusions and handle them. This paper proposes model based on combination of ensemble classification for network traffic anomaly detection. Intrusion detection system is try to perform in real time, but they cannot improved due to the network connections. This research paper is trying to implement intrusion detection system (IDS) using ensemble method for misuse as well anomaly detection for HIDS and NIDS based also. This system used various individual classification methods and its ensemble model on KDD99 and NSL-KDD data set to check the performance of model. It also check the performance on creating real time network traffic using own attack creator and send this to the remote machine which has our proposed IDS system. This system used training rule set as a background knowledge which are generated by genetic algorithm. Ensemble approach contains three algorithms as Naive Bayes, Artificial Neural Network and J48. Ensemble classifiers apply on network packets mapping with GA rule set and generate the result. Finally our proposed model produces highest detection rate and lower false negative ratio compare to others. Also find the accuracy of each attack types.

Keywords* : *Anomaly Detection, Get-Distance, Intrusion Detection, Mutation, Network Security, Network traffic anomaly, Similarities

I. INTRODUCTION

Computer networks security is important domain of research for years. For protecting important information or data the network security technology has become very useful. Any fruitful endeavor or unsuccessful endeavor to trade off the honesty, privacy, and accessibility of any data asset or the data itself is viewed as a security assault or an interruption. Every day industries has to deal with the variety of attacks. Avoiding this problem with the help of Intrusion Detection System (IDS). The wide use of computer networks and the increase in web based business has made security of the host and network an important issue as these are vulnerable to attacks. These attacks can be passive that just reads confidential data or it can be active attack that also modifies or fabricates the data [10]. Since it is not possible to avoid these vulnerabilities and design a completely secure system. Intrusion detection has become a major challenge. The key objective of Intrusion detection system is to recognize the attack and in some matter examine it. Several techniques and methods have been developed. But with the progression of new attacks more robust systems need to be designed.

Basically Intrusion Detection System (IDS) ordered into two distinctive arranged Host Base Intrusion Detection System (HIDS) and Network base Intrusion Detection System (NIDS). Today's system security foundation promisingly relies on Network intrusion detection Framework (NIDS). NIDS gives security from known interruption assaults. It is unrealistic to stop interruption assaults, so associations should be prepared to handle them. ID is a cautious component whose main role is to keep work continuing considering every conceivable assault on a framework. Interruption recognition is a procedure used to distinguish suspicious movement both at system and host level. Two principle ID methods accessible are abnormality identification and abuse location. The oddity identification model depicts the typical conduct of a client to recognize this current client's irregular or unaccustomed activity [12]. Identification is the procedure of observing the activities happening in a network framework or organizes and breaking down them for indications of likely occurrence, which are infringement or looming dangers of infringement of network security arrangements, adequate use strategies, or normal security hones. Fundamentally when an interloper endeavor to break into a data framework or perform an activity not authoritatively permitted, we imply to this activity as an interruption. Interruption system may incorporate abusing programming bugs and plan mis-configurations, secret word incensed, sniffing unsecured exchange, or misusing the outline defect of express conventions. An Interruption Location Framework [14] is a plan for distinguishing interruptions and reporting them definitely to the best possible power.

Manuscript published on November 30, 2019.
* Correspondence Author
Dr. Sharmila. K. Wagh*, Department of Computer Engineering, Modern Education Society's College of engineering, Pune, Maharashtra ,India. Email:skwagh@mescoepune.org
Anuradha S. Varal, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra ,India. Email:Anuradhavaral4@gmail.com

II. LITERATURE SURVEY

In Kagan Tumer a, Adrian K. Agogino[1], proposed Ensemble Clustering with Voting Active Clusters (VACs). This paper consist various clustering models into one cluster that is ensemble clustering. In ensemble clustering does not require all collected data in one central location. The contribution of this proposed model is providing an adaptive voting method to maximize the quality measure by clustering update their votes.

This method achieved better performance than traditional cluster ensemble method. But this model works only on noise-free condition.

According to Giorgio Giacinto et al[24] in the year 2003, a model recognition approach to set of connections intrusion detection based on the combination of manifold classifiers. Five choice blend techniques were assessed by tests and their introductions were thought about. The possibilities of classifier combination for the change of skilled invasion discovery frameworks were surveyed and contended. The announcement results clarified that the MCS approach offers an enhanced exchange off among speculation capacities and false alert age than that offered by an individual classifier instructed on the general list of capabilities.

According to Suseela T et al[25] have presented a multilevel hierarchical Kohonen Net (K-Map) for an intrusion detection system in 2005. Each level of the progressive guide was displayed as a clear victor take-all K-Map. The computational adequacy is the real favorable position of this staggered progressive K-Map. Measurable abnormality recognition strategies, for example, closest neighbor approach, K-implies bunching or probabilistic investigation that utilized separation calculation in the component space to recognize the exceptions. Anyway this staggered progressive K-Map's approach does not draw in costly point-to-point calculation in gathering the information into groups. The decreased system measure is an extra preferred standpoint. The classification capability of the K-Map on picked measurements of information was utilized to set for identifying inconsistencies. Sub-sets chosen heedlessly that encase the two assaults and ordinary records from the KDD Cup 1999 benchmark information were utilized to control the various leveled net.

Li Zheng Tao Li, Chris Ding[2], presents Hierarchical Ensemble Clustering model which works on both partitional clustering and hierarchical clustering. For hierarchical clustering they observed the importance of ultra-metric distance. They also used ultra-metric distance in proposed method from the aggregated distance matrices. Finally, they generated hierarchical clustered structure with separation of the cluster. This model ensemble both framework as partitional clustering and hierarchical clustering.

Sandro Vega-Pons, Jos Ruiez-Shulcloper[3], proposed A Survey of Clustering Ensemble Algorithms which provides alternative ensemble method when occurring cluster analysis problem. In that generating a number of clusters from the same dataset and finally ensemble all together. The main aim of this model is to improve the quality of individual clusters. This paper presents an overview of clustering ensemble methods that can be very important for the clustering practitioners for the purpose of community.

Wei Li [4] describe, Using Genetic algorithm for network intrusion detection. This genetic algorithm works on both temporal and spatial information of network connections during the encoding of the problem; therefore, it should be more helpful for identification of network anomalous behaviors. They also working on TCP/IP layer for detection. This genetic algorithm rule base system which including crossover, mutation, fitness and selection process and finally generate the rules for test data. But this system is rules dependent. If the behavior of the packets flowing in the network is new, then the system cannot take any decision. So they purely work in the basis of the initial rules provided. It cannot create its own rule depending on the current situation and also it requires manual energy to monitor the inflowing packets and analyze their behavior.

According to M. Bahrololum et al[27] in the year 2009 M. Bahrololum published a paper to plan the system using a hybrid of misuse and irregularity detection for training of normal and attack packets respectively. The utilized technique for assault preparing was the blend of unsupervised and managed Neural Network (NN) for Intrusion Detection System. Assaults was arranged into littler classes mulling over their comparable highlights by the unsupervised NN in light of Self Organizing Map (SOM), and taken after by unsupervised NN in light of Back engendering was used for gathering. Known bundles were perceived quick by abuse approach and obscure assaults will have the capacity to spot by this strategy.

In Kaur, P.[5], present Adaptive Intrusion Detection Based on K-SVMeans Algorithm, which overcome the problem of one individual algorithms for improving the efficiency. To improve the efficiency of clustering algorithm it is a challenging work without overcoming the generalization performance of Support Vector Machine (SVM). Have gone through this difficulties , so this model developed new hybrid method based on combination of SVM and k-means clustering. SVM is used to build classifiers which can help users to take business decision very well. Response time of support vector machine is concern in real-time task network traffic analysis. The KSVM algorithm combines the k-means clustering technique with SVM.

P. Jongsuebsuk and N. Wattanapongsakorn [6] describe, Network intrusion detection with fuzzy genetic algorithm for unknown attacks. Fuzzy rule is a machine learning algorithm that can classify network attack data, while a genetic algorithm is an optimization algorithm that can help finding appropriate fuzzy rule and give the best optimal solution. This system experimental results show that fuzzy GA can efficiently detect online network dataset within 2-3 seconds, Where 2 seconds belong to the preprocessing time and less than a second for the detection time, while it takes only a fraction of a second to detect attacks in the KDD99 dataset. In given approach having two experimental results first fuzzy algorithm classify the attack on online dataset and KDD dataset with high accuracy and false alarm rate, while second experiments illustrated detection rate of each attack. Demerits of this system that there is no optimal solution for threshold so sometime it will effect on actual detection rate and also they have not classified overall detection rate for all attacks.

According to Iftikhar Ahmad, Azween et al[28] in the year 2011, proposed a paper to surmount introduction issues an improved obstruction recognition system by methods for delicate processing strategies. The KDD-container dataset was connected that was a benchmark for evaluating the security distinguishing proof instruments. To change the key in models into an element space the Principal Component Analysis (PCA) was connected. Choosing of an appropriate amount of chief segments was a critical issue.

As an option of utilizing traditional strategy, Genetic Algorithm (GA) was connected in the ideal decision of foremost segments as needs be. The Support Vector Machine (SVM) was utilized for classification reason. What's more, a corresponding report was set up with displayed approaches. In this manner, the procedure exhibited ideal obstruction discovery system was capable to limit measure of highlights and augment the distinguishing proof rates.

Basant Subba, Santosh Biswas, Sushanta Karmakar[9], proposed A Neural Network Based System for Intrusion Detection and Attack Classification. This paper used a simple Artificial Neural Network (ANN) based IDS model. Using different optimization techniques presents intrusion detection system with the help of feed forward and the back propagation algorithms. This method used to reduce the overall computational overhead and also maintain performance level. The proposed model results shows that high accuracy and higher detection rate on NSL-KDD dataset. This model is very useful for analyzing the detection of intrusions in real time environment.

In Wagh SK, Kolhe SR.[7], present Effective semi-supervised approach towards intrusion detection system using machine learning techniques. This system proposed new semi-supervised method using machine learning approach, which effectively increased detection rate and by default accuracy also improved. Semi-supervised learning method contains both unlabeled data and labeled data. This paper present a new method called as novel self-learning construction. In that input as a train data given to the supervised classifier with unlabeled dataset for testing purpose. Entropy is calculated for correct predicated label. Threshold calculation is done by adding confident data into the training data. For data selection number of statistical methods used. Alert mechanism is improved with this proposed model.

According to Adel Nadjaran Toosi and Mohsen Kahani [26] in 2007, an evolutionary soft computing approach for intrusion has been commenced and was successfully clarified its utility on the preparation and testing subset of KDD Cup 99 dataset. For obstruction recognition, the ANFIS arrange was used as a neuro-fluffy classifier. Without the assistance of human specialists ANFIS was skilled of creating fluffy tenets. Besides, to discover the quantity of guidelines and enrollment capacities with their underlying areas for improved characterization subtractive bunching had been utilized. To make the framework all the more overarching for assault identification utilizing the fluffy derivation approach, a fluffy basic leadership motor was developed. At last, they arranged a framework to use hereditary calculations to enhance the fluffy basic leadership motor.

According to Naila Belhadj Aissa, Mohamed Guerroumi[8], present A Genetic Clustering Technique for Anomaly-Based Intrusion Detection Systems, which focus on identifying the anomaly by using genetic clustering algorithm. The proposed model is Genetic Clustering for Anomaly-based Detection (GC-AD). GC-AD utilizes a divergence measure to frame k cluster and find centroid of the k groups by applying on hereditary procedure. It worked on KDD99 dataset for gaining the exactness of frameworks system. The output is calculated with kmeans grouping. The main goal for CG-AD (Clustering Genetic for Anomaly-based Detection) is to get an ideal homogenous apportioning of typical and oddity cases.

According to K. M. Faraoun and A. Boukelif [29] in the year 2004, utilizing the K-mean bunching calculation a procedure have developed by to enhance the learning limits and diminishing the calculation quality of a focused learning multi-layered neural system. Through a back engendering learning implies the suggested display utilized multi-layered system auxiliary plan. To diminish the measure of cases to be offered to the neural system, the K- mean calculation was at first used to the preparation dataset via naturally picking a most positive arrangement of tests. The obtained results demonstrated that the proposed strategy executes extraordinarily as far as both exactness and calculation time when related to the KDD99 dataset coordinate to an ordinary learning pattern that used the full dataset.

III. SYSTEM OVERVIEW

Existing system has three phases. The first phase is choosing a proper dataset and applied on preprocessing phase for minimizing or elimination of the noise forced on the data. Second is building the hybrid model which consisting number of classifiers which produced accuracy of each individual classifier. Finally in Analysis phase, results is generated by comparing accuracy of each individual model and select the best model as ensemble. Advantage of existing system that it shows higher detection rate and it improves accuracy rate. It also shows that majority of attacks are done using the TCP protocols but this system works only on host based. Problem definition of the propose system is to develop the Intrusion Detection System for distributed architecture for the detection and correctly classification of the incoming network packet attacks using ensemble method.

The proposed system worked with ensemble model. When two or more models are ensemble together to form a new model called as ensemble model. This ensemble model combines the output of several classifiers and produced a single composite classification. Our proposed model consisting number of classifiers. System first collects data from different online as well as offline sources. Once data has collected by system it will apply some data mining strategies with different classification approaches.

Objectives of Proposed System:

1) To study existing Network Intrusion Detection Systems (NIDSs) and types of NIDSs.
2) Execute the same system on HIDS as well NIDS base environment.
3) To propose a new integrated approach for network anomaly detection using ensemble approach, compare the result with individual algorithmic results.

4) To compare the experimental results of existing methodology with proposed system for network anomaly detection.

IV. SYSTEM ARCHITECTURE

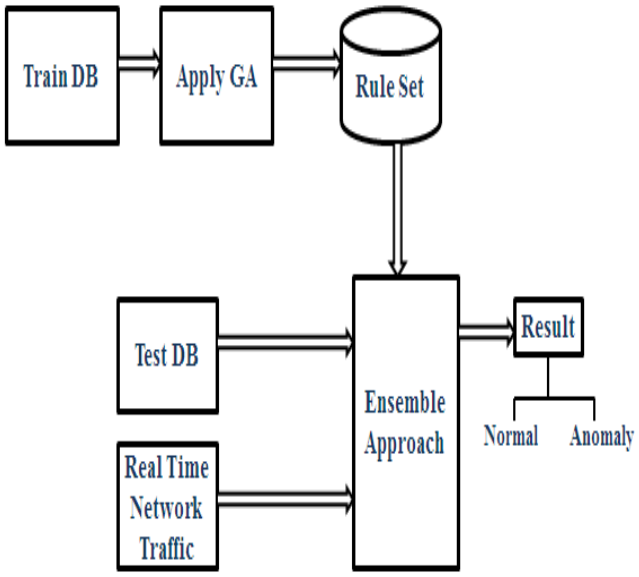


Fig.1 Proposed System

Dataset Details

The KDD cup 99 dataset contains several statistical analyses which affects to detect the accuracy of many IDS model. The KDD dataset has training and testing dataset. The total number of training dataset is approximately 4, 900, 000. It contains 41 features with labeled as normal or specific attack type. 300, 000 instances contains in testing dataset with twenty four training attack types and extra fourteen attack types in the test set only. NSL-KDD data set is a modified version of its precursor. It consist, important records of the KDD data set. There are 4 attack classes of anomaly which are further classified as DoS, Probe, R2L and U2R. In that there are subtypes of each anomaly attack types.

System Modules:

1) *Data Preprocessing:*

Data preprocessing done by Weka tool. This is offline method. Data preprocessing includes following three main task:

- Converting non-numerical features of NSL-KDD dataset into numerical values.
- At the end transferring attack types into numerical values.
- Finally preparing proper dataset.

2) Rule Creation:

Genetic algorithm is used for creating a rule set as a normal pool and intrusion pool. Apply the genetic algorithm on training dataset and generate the proper rule set. Our system works on main six attribute. This rule set is used as a background knowledge when ensemble model is running in system.

3) Ensemble Model:

Building the hybrid model consisting number of classifiers. Our system uses Naive Bayes, Artificial Neural Network and J48 algorithm for ensemble model. Input to the ensemble model is real time input packets and current rule set. Generate the each algorithm values on the basis of rule set. Check the final values and detects the master attack.

4) Result Generation:

Finally results are generated whether incoming packet is normal or anomaly. If it is anomaly then it also finds subclasses of that anomaly.

V. SYSTEM ANALYSIS

A. For Standalone System:

Input: Training and Testing Dataset
Output: Attack Detection

Step 1: Load Training dataset and Testing dataset.
Step 2: Apply Genetic Algorithm on training dataset for rule creation purpose.

- Select initial population size and iteration size.
- Proposed system works on main six attributes of packets.

G= {duration, protocol, service, flag, src_byte, dst_byte}

-Find the fitness function of each chromosomes and select best fitness value which gives the best quality rule set.

Fitness= F(x)/sum(F(x))

F(x)= TP+FN

- Finally extract the rule set.

Step 3: Apply ensemble model on testing dataset (input packet) for detection attack

- Find the similarity between input packets & current rules and calculate the weight for each classifiers.
- Calculate similarity from following formula:

Weight = getsimilarity(S, T);

Where, S= Input Packet

T= Rule set

- For ensemble model three classifiers used
- EM= {J48, NB, NN}
- W1 for NB, W2 for NN, W3 for J48 classifier.

- Calculate W1 value using NaiveBayes:

W1= getsimilarity(S, T);

Where, S= Input Packet

T= Rule set

- Calculate W2 value using Neural Network:

W2= CalculateWeight (HiddenLayer[i], InputNeurons);

Where,

HiddenLayer= Rule set.

InputNeurons= Input Packets

Weight= ((srccount*100)/length);

srccount= getsimilarity(InputNeurons, HiddenLayer);

- Calculate W3 value using J48:

W3= GetDistance (InputPacket, Rule set);

Weight= min+ Math.random()*(max-min);
Where, max= 0.35, min= 0.01

- et the threshold value as Th= 0.35.
- Threshold value is the limit on a scale. If the threshold is reached, our detection phase generated on following condition:
  If (W1>=Th || W2>=Th || W3>=Th)
  Then check master attacks i.e, DoS, Probe, U2R, R2L.
  Else
  Normal connection;
- Then Find Attacks which check the sub-types of masters attack and increased the particular master attack types.

Step 4: Shows the final result and generate analysis graph of detection ratio.

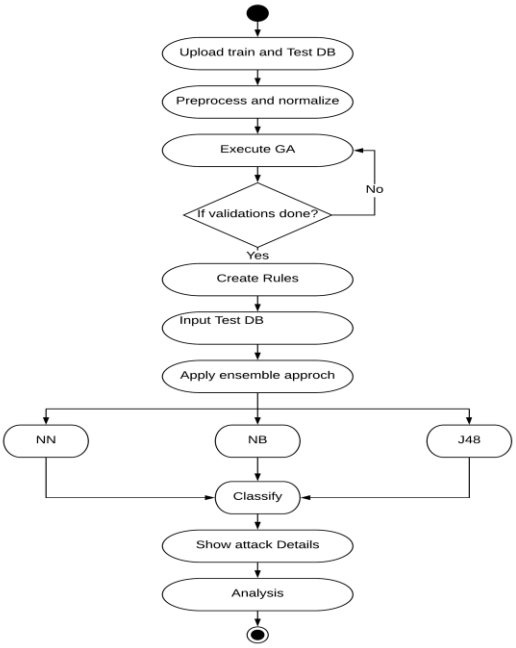
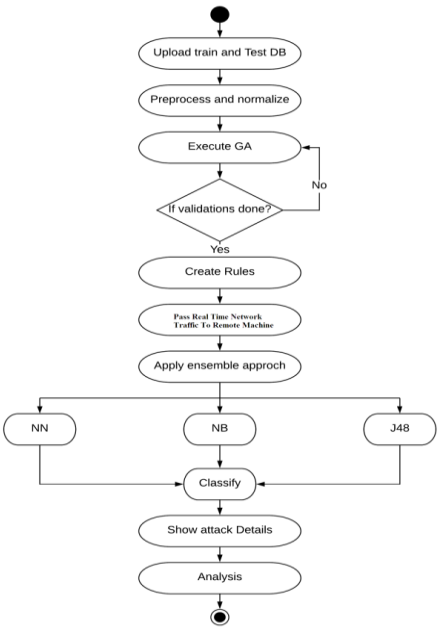


Fig.2 Activity Diagram of Standalone System



B. For Real-Time System:

Input: Training Dataset, Network Traffic Packets
Output: Attack Detection
Step 1: Load Training dataset on remote machine.
Step 2: Apply Genetic Algorithm on training dataset for rule creation purpose.
- Select initial population size and iteration size.
- Proposed system works on main six attributes of packets.
  G= {duration, protocol, service, flag, src_byte, dst_byte}
  - Find the fitness function of each packets and select best fitness value which gives the rule set.
    Fitness= F(x)/sum(F(x))
    Where, F(x)= TP+FN
  - Finally extract the rule set.

Step 3: On other machine called as remote attacker machine, generate own attack creator machine using following main seven attributes:
RA = {Source IP, Source Port, MAC Address, Destination Port Number, Source_Byte, Time-Stamp}

- Remote Attacker Machine connect to the remote machine using socket programming within same network.
- Whenever server started on remote machine, remote attacker generate random packets on above attributes and send to the remote machine.

Step 4: On remote machine, apply ensemble method on real time input packets which comes from remote attacker for detection particular attack types.

- Find the similarity between input packets & current rules and calculate the weight for each classifiers.
- Calculate similarity from following formula:
  Weight = getsimilarity(S, T);
  Where, S= Input Packet
  T= Rule set
- For ensemble model three classifiers used.
- EM= {J48, NB, NN}
- W1 for NB, W2 for NN, W3 for J48 classifier.
- Calculate W1 value using NaiveBayes:
  W1= getsimilarity(S, T);
  Where, S= Input Packet
  T= Rule set
- Calculate W2 value using Neural Network:
  W2=CalculateWeight(HiddenLayer[i], InputNeurons);
  Where,
  HiddenLayer= Rule set.
  InputNeurons= Input Packets

Weight= ((srccount*100)/length);
srccount= getsimilarity(InputNeurons, HiddenLayer);

- Calculate W3 value using J48:
  W3= GetDistance (InputPacket, Rule set);
  Weight= min+ Math.random()*(max-min);
  Where, max= 0.35, min= 0.01
- Set the threshold value as Th= 0.45.
- Threshold value is the limit on a scale.

If the threshold is reached, our detection phase generated on following condition:

If (W1>=Th || W2>=Th || W3>=Th)

Then check master attacks i.e, DoS, Probe, U2R, R2L.

Else

Normal connection;

- Then Find Attacks which check the sub-types of master attack and increased the particular master attack types.

Step 5: Shows the final results of detection and generate Result.txt file which saves the attack detection output.

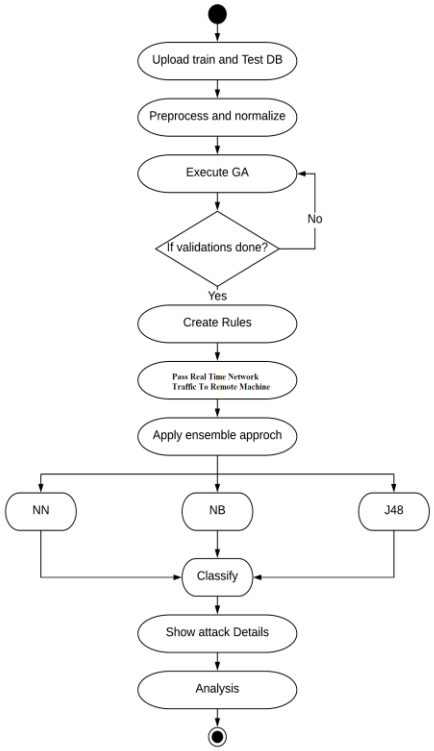


Fig.3 Activity Diagram of Real-Time System

C. **Analysis of Ensemble Model on Real-Time:**

Table 1: Ensemble model values of three classifiers

Packet No.	Threshold Value	W1 Value	W2 Value	W3 Value	Class Type
1	0.35	0.31	0.1	0.17	Normal
2	0.35	0.13	0.13	0.3	Normal
3	0.35	0.13	0.19	0.34	Normal
4	0.35	0.13	0.1	0.19	Normal
5	0.35	0.12	0.22	0.32	Normal
6	0.35	0.13	0.23	0.13	Normal
7	0.35	0.12	0.22	0.33	Normal
8	0.35	0.12	0.19	0.29	Normal
9	0.35	0.11	0.15	0.26	Normal
10	0.35	0.12	0.16	0.02	Normal

Note:

Total number of packets= 10

Threshold value= 0.35

W1 = NaiveBayes classifier

W2 = Neural Network classifier

W3 = J48 classifier

Calculate W1 value using NaiveBayes:

W1= getsimilarity(S, T);

Where, S= Input Packet

T= Rule set

Similarity=longerLength-editDistance(longer, shorter)/longerLength;

Calculate W2 value using Neural Network:

W2= CalculateWeight (HiddenLayer[i], InputNeurons);

Where,

HiddenLayer= Rule set

InputNeurons= Input Packets

Weight= ((srccount*100)/length);

srccount= getsimilarity(InputNeurons, HiddenLayer);

Calculate W3 value using J48:

W3= GetDistance (InputPacket, Rule set);

Weight= min+ Math.random()*(max-min);

Where, max= 0.35, min= 0.01

For example,

For 1st packet,

W1= getsimilarity(Input Packets, Rules);

Similarity= (22- editDistance(31, 16))/22;

= (22-15)/22

W1 =0.31

W2= CalculateWeight(Rules, Input Packets);

Srccount= getsimilarity(Rules, Input Packets);

Srccount= 16

Weight= ((16*100)/15947)

W2= 0.10

W3= GetDistance (InputPacket, Rule set);

Weight= 0.01+0.49*(0.35-0.01);

Weight= 0.17

W3=0.17

D. Threshold Value:

Proposed system test on 3 to 4 threshold values in simulation and pick the one that gives the optimum results, such as good detection accuracy.

- Th= 0.15 value gives bad result, it shows all normal traffic as attack.
- Th=0.30 value gives better result than 0.15, it shows some packets are normal and some anomaly, but actual all packets are normal.
- Th= 0.35 value gives best result than other, it shows all normal traffic in network correct as a normal.

That's why our proposed system used 0.35 threshold value for real time traffic detection.

This depends on the approach you use for detecting attacks. A general problem with a threshold value, is that it may be valid only for a limited period of time if attack pattern changes.

E. Experimental Results

In proposed model, firstly apply the training dataset. Training dataset is accepted as shown in Fig.4. Training dataset is used for extracting the features. The feature extraction done automatically, when rule creation process start with the help of genetic algorithm.

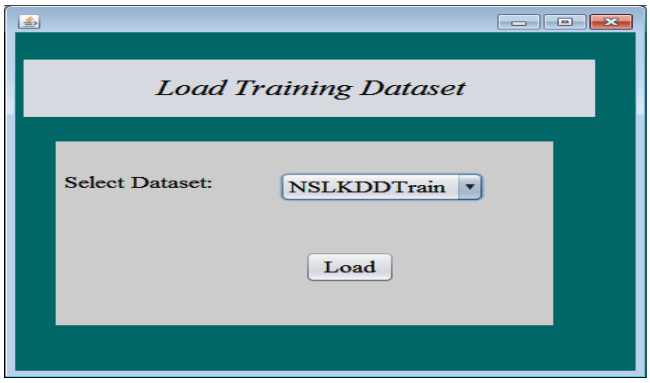


Fig. 4 Upload Training Dataset Form

After training dataset loaded, apply genetic algorithm for rule creation. Whenever applies genetic algorithm, it ask for population size and generation size. Select the random population size of chromosomes and set the variation size which shows that how many times our genetic algorithm run. Fig.5 shows that the population size and generation size form.

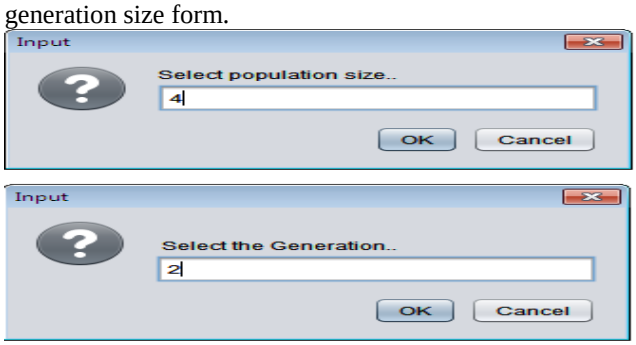


Fig. 5 Population and Generation Size Form

Genetic algorithm use biological concept like mutation, selection and crossovers. After selecting random population and generation size, it checks generation reached or not. If the generation is reached, then it generates the result as a best solution, otherwise apply above biological terms. Fig.6 shows output of the genetic algorithm. Output creates normal pool and intrusion pool.

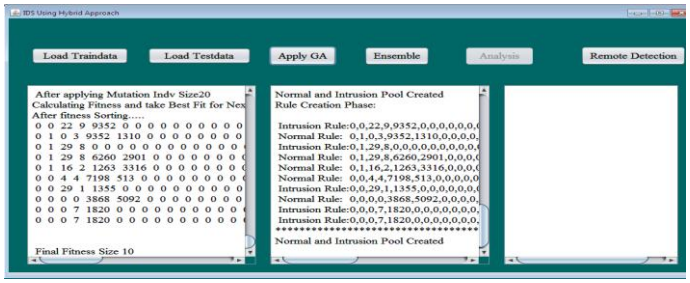


Fig. 6 Rule Creation Form

Finally, apply testing dataset as NSL-KDD dataset or real time network traffic on ensemble model. Ensemble model maps test dada with rule set and generate final result as packet is normal or anomaly shows in Fig.7.

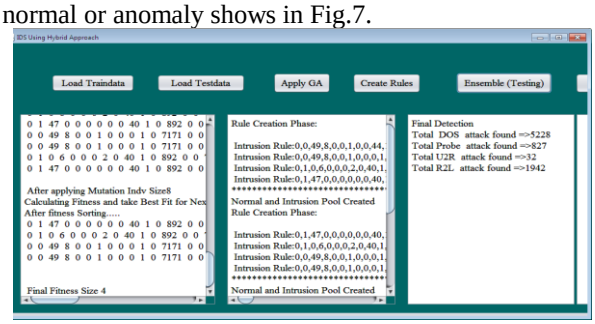


Fig.7. Attack Detection using Ensemble Model

On real time system, Firstly create the remote attacker shown in Fig.8. On this remote attacker capture the real time traffic with seven main attributes. E.g. Source IP, Protocol, Source Port, Mac address, Destination Port Number, Source Bytes, Time Stamp etc. Put the IP address of remote machine which receives the attack packets from remote attack creator machine. In remote attack creator frame, we put genera te attack button, stop generation and count of that attack packets.

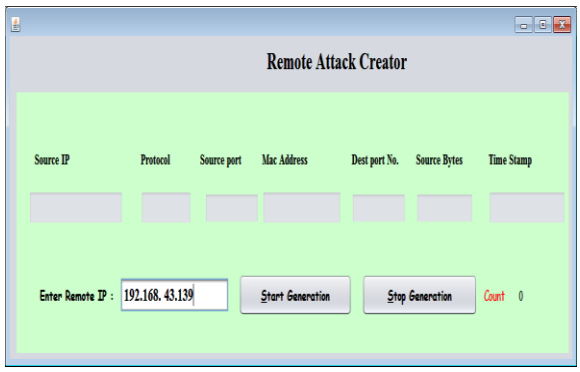


Fig.8. Remote Attack Creator

Fig.9. shows the start generation of real time attacks to the remote machine.

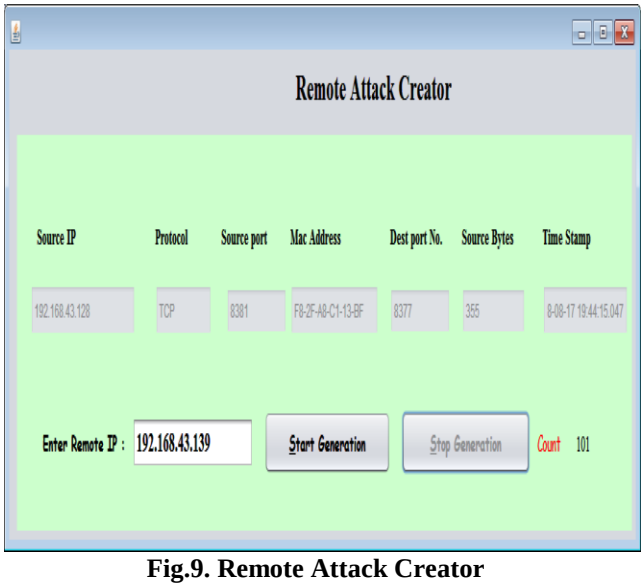


Fig.9. Remote Attack Creator

Remote machine receives the packets and apply the ensemble model with current rules and detects the receiving packets are normal or anomaly shown in Fig.10.

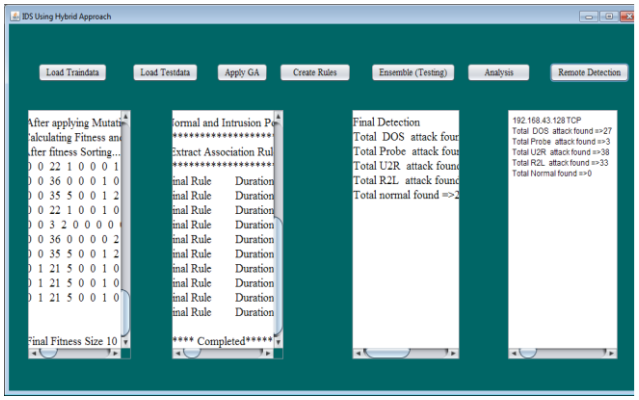


Fig.10. Remote Attack Detection

Finally generate the .txt file which save the final detection result shown in Fig. 11.

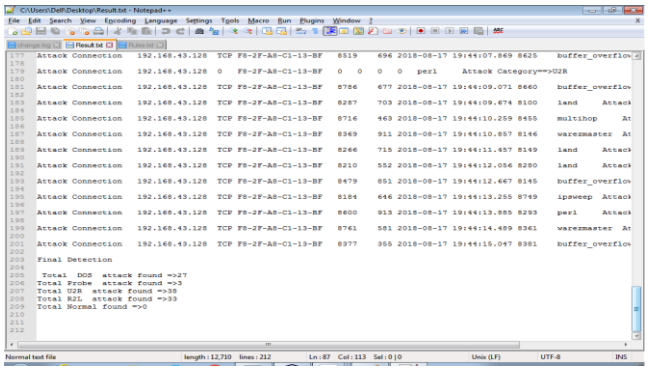


Fig.11. Final Detection Result .txt File

The existing survey basically focus on soft computing and classification based detection approach, basically both methods having the good detection rate but at times it generates more false positive ratio. Some systems are also not applicable in real time environment and some can't be focus on mis-classified anomalies. As observed, most applications still miss the mark as there is no system that at present gives a 100% discovery rate and the sky is the limit.

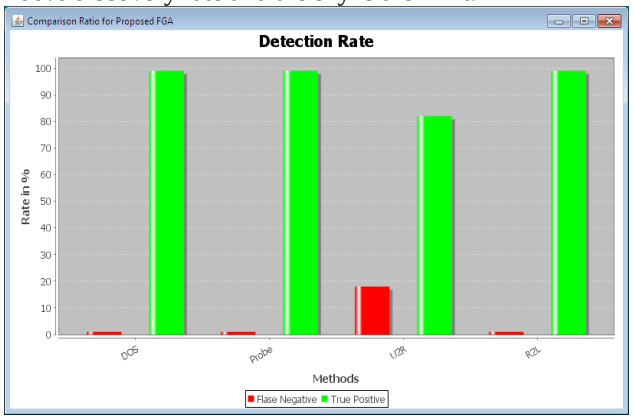


Fig.12. Proposed System Detection Rate

Fig.12. shows graphical representation of proposed system detection ratio using ensemble approach. This result analyzes by passing the particular packet size to the proposed ensemble model. The size of the packet is 10419. In that 8067 packets are attack types and 2352 packets are normal. In graphical representation, the X-axis presents attack types as Dos, Probe, U2R, L2R and Y-axis presents percentage of particular attack types using True Positive (TP) and False Negative (FN) terms.

TABLE I
Training Dataset

Anomaly	Normal	Total
1606	1394	3000

Note: Total Training Packets= 3000

TABLE II
Testing Dataset

Anomaly	Normal	Total
80	20	100
815	185	1000
8172	1828	10,000
81720	18280	1,00000

Note: Total Testing Packets= 100, 1000, 10000, 1,00000

TABLE III
Detection on 100 Packets

	Correct Count	Correct Count(%)	Incorrect Count	Incorrect Count(%)
Normal	20	100	0	0
DoS	54	100	0	0
Probe	9	100	0	0
U2R	0	100	0	0
R2L	12	70.58	5	29.41
Total	95	95	5	5

Note: Training Packets=3000, Testing Packet=100, Accuracy=95

TABLE IV
Detection on 1000 Packets

	Correct Count	Correct Count(%)	Incorrect Count	Incorrect Count(%)
Normal	185	100	0	0
DoS	499	100	0	0
Probe	81	90	9	10
U2R	4	80	1	20
R2L	182	82.35	39	17.64
Total	951	95.1	49	4.9

Note: Training Packets=3000, Testing Packet=1000, Accuracy=95.1

Retrieval Number: D8534118419/2019©BEIESP
DOI:10.35940/ijrte.D8534.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

TABLE V
Detection on 10,000 Packets

	Correct Count	Correct Count (%)	Incorrect Count	Incorrect Count (%)
Normal	1828	100	0	0
DoS	5008	99.96	2	0.03
Probe	785	85.32	135	14.67
U2R	37	58.73	26	41.26
R2L	1842	84.53	337	15.46
Total	9500	95	500	5

Note: Training Packets=3000, Testing Packet=10000, Accuracy=95

TABLE VI
Detection on 1,00000 Packets

	Correct Count	Correct Count(%)	Incorrect Count	Incorrect Count(%)
Normal	18280	100	0	0
DoS	50090	99.98	10	0.019
Probe	8954	97.32	246	2.67
U2R	626	99.36	4	0.6
R2L	21382	98.12	408	1.87
Total	99332	99.32	668	0.66

Note: Training Packets=3000, Testing Packet=100000, Accuracy=99.32

NSL KDD Total packets
Train= 125973
Test= 22544

Used
NSLKDDTrain+20% ==total packets=25192
NSLKDDTest+21%== total packets=11850
Fig.12. The accuracy of testing of proposed system is shown in Figure 13 for normal, DoS Probe, U2R, and R2L. Accuracy for normal, DoS Probe, U2R, and R2L is 100%, 99.98%, 97.32%, 99.36% and 98.12% respectively.

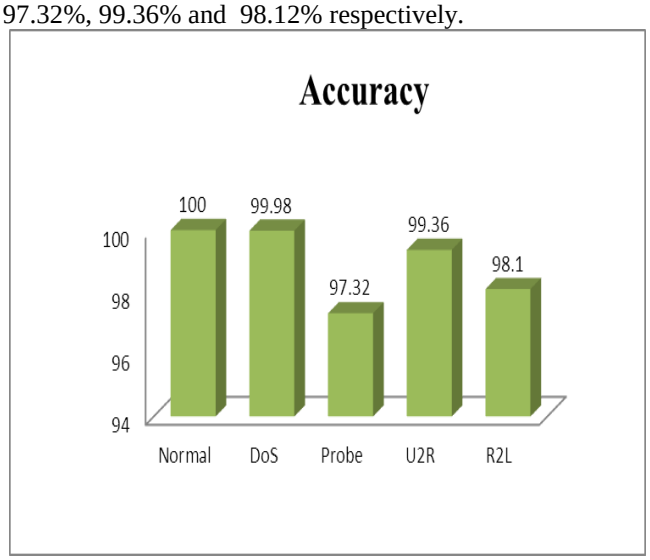


Fig.13. Accuracy of Proposed System

TABLE VII
OVERALL PERFORMANCE OF PROPOSED SYSTEM COMPARED WITH EXISTING SYSTEM

Detection Rate (in %)	DoS	Probe	U2R	R2L
Existing	99.9	96.2	99.1	97.9
Proposed	99.9	97.3	99.3	98.1

Table VII shows the result of proposed system compared with existing system[15]. Here finding the detection rate is done. Detection rate is finding on correctly classified records upon total number of records. Detection rate shows the accuracy of the system. Fig.14 shows the graphical representation of overall performance of proposed system compared with existing system[15] based on Table VII. Here X-axis shows anomaly attacks detection ratio of proposed system compared with existing system and Y-axis shows detection rate in percentage.

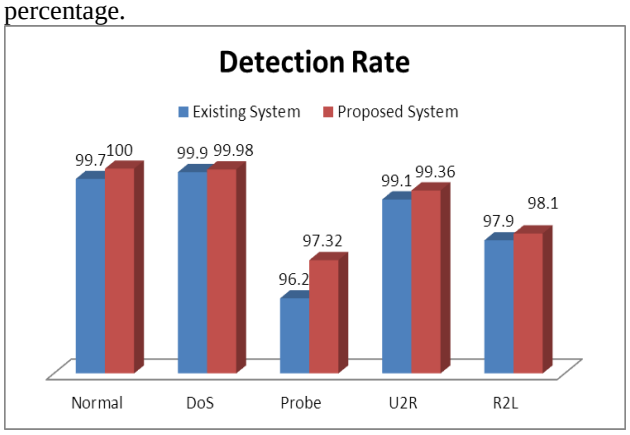


Fig.14. Overall Performance of Proposed System compared with existing system

VI. CONCLUSION

In this research work we proposed ensemble method for real time network traffic anomaly detection. Our approach concentrated on building normal traffic profile of the anomaly detection model. For normal profile experimental result showed that the features of NSL-KDD is efficient. The experimental result showed that system has excellent performance with small training dataset and detection accuracy. Also our proposed system worked better than existing system on real time network traffic. Result showed that the proposed system gives higher detection ratio on real time network traffic. We also proposed a new model integrates anomaly detection system with Rule-based detection system along with some enhancements of building quality normal profile. Ensemble approach contains three algorithms as Naive Bayes, Artificial Neural Network and J48. Ensemble classifiers apply on network packets mapping with GA rule set and generate the result. Proposed research work also perform the better detection, On the basis ensemble approach implementation we got a IDS system can achieve better detection rate for all attacks as well as unknown attacks. In future work we will plan to minimize the computation time of Genetic algorithm with parallel execution. Also used different algorithms for ensemble approach and analyze the detection rates.

REFERENCES

1. Kagan Tumer a, Adrian K. Agogino, "Ensemble clustering with voting active clusters," Elsevier, 2008.
2. Li Zheng Tao Li, Chris Ding, "Hierarchical Ensemble Clustering", IEEE International Conference on Data Mining, 2011.
3. SANDRO VEGA-PONS, JOS RUIZ-SHULCLOPERy, "A Survey of clustering Ensemble Algorithms", International Journal of Pattern Recognition and Artificial Intelligence Vol. 25, 2011.
4. Wei Li , "Using Genetic algorithm for network intrusion detection", 2012.
5. Kaur, P.," Adaptive Intrusion Detection Based on K-SVMeans Algorithm", (Doctoral dissertation, THAPAR UNIVERSITY PATIALA), 2013.
6. P. Jongsuebsuk and N. Wattanapongsakorn , "Network intrusion detection with fuzzy genetic algorithm for unknown attacks", 2013.
7. Wagh SK, Kolhe SR., "Effective semi-supervised approach towards intrusion detection system using machine learning techniques", International Journal of Electronic Security and Digital Forensics, 7(3):290-304, 2015.
8. Naila Belhadj Aissa, Mohamed Guerroumi, " A Genetic Clustering Technique for Anomaly-Based Intrusion Detection Systems", IEEE, 2015.
9. Basant Subba , Santosh Biswas, Sushanta Karmakar , "A Neural Network Based System for Intrusion Detection and Attack Classification", IEEE, 2016.
10. Wagh SK, Pachghare VK, Kolhe SR, "Survey on intrusion detection system using machine learning techniques", International Journal of Computer Applications. 1;78(16). Jan 2013.
11. Mohammed A. Ambusaidi et. al., "Building an intrusion detection system using a filter-based feature selection algorithm" , IEEE TRANSACTIONS ON COMPUTERS, VOL., NO , NOVEMBER 2014.
12. Fatemeh Barani , "A Hybrid Approach for Dynamic Intrusion Detection in Ad Hoc Networks Using Genetic Algorithm and Artificial Immune System", IEEE , 2014.
13. Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis, "Intrusion Detection System Using Genetic Algorithm", Science and Information Conference , 2014.
14. Alka Chaudhary, Vivekananda Tiwari, Anil Kumar , "A Novel Intrusion Detection System for Ad Hoc Flooding Attack(Using Fuzzy Logic in Mobile AdHoc Networks)", IEEE, 2014.
15. Shadi Aljawarneha, Monther Aldwairi, Muneer Bani Yassein, "Anomaly based intrusion detection system through feature selection analysis and building hybrid efficient model", Elsevier ,2017.
16. Eduardo K. Viegas, Altair O. Santin , Luiz S. Oliveira , "Toward a reliable anomaly-based intrusion detection in real-world environments", Elsevier , 2017.
17. Amira SayedA. Aziz, Sanaa EL-OlaHanafi, Aboul EllaHassanien, "Comparison of classification techniques applied for network intrusion detection and classification", Elsevier , 2016.
18. Weiwei Chen, Fangang Kong, "A Novel Unsupervised Anomaly Detection Approach for Intrusion Detection System", IEEE, 2017.
19. Kumari VV, Varma PR., "A semi-supervised intrusion detection system using active learning SVM and fuzzy c-means clustering", In I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), IEEE, International Conference on 2017 Feb 10 (pp. 481-485), 2017.
20. Zhang X, Zhu P, Tian J, Zhang J., "An effective semi-supervised model for intrusion detection using feature selection based Lap SVM", In Computer, Information and Telecommunication Systems (CITS), IEEE, International Conference on 2017 Jul 21 (pp. 283-286), 2017.
21. N. H. Duong, H. D. Hai, "A semi-supervised model for network traffic anomaly detection", IEEE 2015 17th International Conference on Advanced Communication Technology (ICACT), Phoenix Park, 2015: 70-75, 2015.
22. Gao, G., Miao, G., Sun, J. and Han, Y., " Improved semi-supervised fuzzy clustering algorithm and application in effective intrusion detection system", International Journal of Advancements in Computing Technology (IJACT), Vol. 15, No. 4, pp.689696, 2013.
23. Meng, Y. and Kwok, L-F. , "Intrusion detection using disagreement based semi-supervised learning: detection enhancement and false alarm reduction", 4th International Symposium, Lecture Notes in Computer Science, Vol. 7672, pp.483497, Springer, Berlin Heidelberg, 2012
24. Giorgio Giacinto, Fabio Roli, and Luca Didaci, "Fusion of Multiple Classifiers for Intrusion Detection in Computer Networks," Journal of Pattern Recognition Letters, vol.24, pp.1795-1803, 2003.
25. Suseela T. Sarasamma, Qiuming A. Zhu, and Julie Huff "Hierarchical Kohonenen Net for Anomaly Detection in Network Security," IEEE Transactions on Systems, Man, And Cybernetics-Part B: Cybernetics, Vol. 35, No. 2, April 2005.
26. Adel Nadjaran Toosi, Mohsen Kahani, "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers," Computer Communications vol.30, pp.2201–2212, 2007.
27. M. Bahrololum, E. Salahi and M. Khaleghi, "Anomaly intrusion detection design Using Hybrid of Unsupervised and supervised neural Network," International Journal of Computer Networks & Communications (IJCNC), Vol.1, No.2, July 2009.
28. Iftikhar Ahmad,Azween Abdullah,Abdullah Alghamdi,Muhammad Hussain,"Optimized intrusion detection mechanism using soft computing techniques,"Telecommun System,2011.
29. K. M. Faraoun and A. Boukelif,"Neural Networks Learning Improvement using the K-Means Clustering Algorithm to Detect Network Intrusions," International Journal of Computational Intelligence, Vol.3, no.2, 2005.
30. Ms. Anuradha S. Varal, Dr.S.K.Wagh, "Anomaly Based Intrusion Detection System Using Soft Computing and Classification Approach" ", International Journal of Scientific Research Computer Science, Engineering and Information Technology (IJSRCSEIT) 2017, Volume 2, Issue 6, ISSN : 2456-3307
31. Ms. Anuradha S. Varal,Dr.S.K.Wagh, "Missuse and anomaly detection Using Ensemble Learning Network Traffic Model",Journal for Advanced Research in Applied Sciences [IAETSD-IARAS] April 2018, Volume 5, Issue 4, ISSN NO.: 2394-8442.

AUTHORS PROFILE

Dr.Sharmila Kishor Wagh, completed her Ph.D in computer engineering in 2016 from North Maharashtra University. Her areas of research are cyber security and machine learning. She published research papers in various national and international journals. She is life member of CSI,IETE and ISTE.

Ms.Anuradha Varal, completed her ME (computer engineering) in 2018 from SPPU, Pune. Her areas of research are cyber security and machine learning. She published research papers in various national and international journals.