

A Design of Intrusion Detection using Modified Bat Algorithm and Deep Autoencoder Network

Soundari D.V M.E., Kuralarasi R, Kalieshwari S, Kanimozhi M, Kanimozhi N.



Abstract: In this world, Intrusion Detection is more popular for preparing the network security systems. In current trend of increasing security system, there is a demand for Intrusion Detection. With these clarifications need to find a huge Data measurement, high speed traffic's and frequent forms of threats. In this work, Intrusion Detection is done by Deep Auto-Encoder network (DAEN) and Modified BAT algorithm (MBA). Our approach improves the Deep Auto Encoder (DAE) classifier by manipulating the benefits of an additional process encourage through the atmosphere of microbats (Bat Procedure). The core aim of this work is to select the features based on Modified Bat Algorithm. Towards examine the model, using the NSL-KDD data's and the survey of Modified Bat Algorithm will be discussed. Moreover, these methods do well to improve DAEN classifier and get reliable performance in standing of accuracy (96.06%), attack detection rate (95.05%).

Keywords: Network Security, Modified BAT Algorithm, Deep Auto Encoder Network, Intrusion Detection, NSL-KDD.

I. INTRODUCTION

Countless activities of peoples indicate consuming the internet's (online outgoings, internet investment, public links and analyses some information's) and just about to stock the grave data's over the internet by the government otherwise private officialdoms. Rising promptness of webs has been regulated the proliferation of miscellaneous hazards. Security schemes became energetic apparatuses from these bases. Even with the recent developments, security happenings be gradually located. IDS take grow into an imperative component of almost individually precautions foundation, for the reason that they arrange for a partition of defense and endure outward attacks technically, anywhere forgoing security modes cannot behave fine. Intrusion Detection Systems (IDS) observer those happenings going on their entire systems and plump whether it is ordinary actions or meddling of the entity. It has two types of detection techniques one is misuse detection and another one is anomaly detection.

The misuse module is more slapdash taking part in spotting realized spells initiated on their signatures that dumped here the databases. The anomaly detection tactic inevitably fabricates a usual manner of the structures. These two techniques can find some unwanted malfunctions but it created more false alarms. In hand there remain lots of dares to be concerned though executing an IDS such as donating the present comebacks through a elevated assault finding speed and a little fake fright pace. Correspondingly, the hefty numeral of features and inconvenient to recognize the multipart relationship among them brews an uncomfortable job. In systematize to lecture these erupts, we have in mind an interruption finding method using Modified BAT algorithm and Deep Auto Encoder (DAE). Our methodology develops two main components: feature selection and enhanced DAEN classifier.

II. RELATED WORK

IDSs take largely explored by way of apologetic habits toward recognize strange or nil-daytime assault. Irregularity-based IDSs the consistent manners of system at that point recognize attacks as unconventionalities since the outdated process. The major challenge in scheming of irregularity-based IDS stays the probable used for elevated fake terror charge as formerly concealed structure activities may be labeled as incongruities. Thus, an efficient IDS is clever to grip a great quantity of statistics with fluctuating forms in actual instance conditions.

Up till now, swarm intelligence (SI) algorithms partaken grossed excessive concentration, largely for the reason that they stand simple, flexible (be capable of pragmatic on the way to a assortment of glitches such as optimization, statistics removal and so forward) and buoyant (the algorithm will hand out yet gamble a few entities flop to complete their tasks). Formerly, in the area a scope of SI course of actions such as: ant colony optimization, atom group optimization, artificial bee settlement algorithm, firefly algorithm, cuckoo look for or strike algorithm. These contain be to be bounded by DAEN to indenture enlarged IDS reproductions and innermost spots be exhausted intended for two optimization innovations: characteristic assortment and picking DAEN parameters. They used BAT algorithm to determine the greatest feature subset and algorithm to search for optimum DAEN parameters. In this paper we using a rather fresh meta heuristic to get stronger the Deep Auto Encoder classifier, the Bat Algorithm (BA), which shows auspicious grades (Yang, 2010a), (Yang and He, 2013). As well, we focus an earlier characteristic assortment that merges Modified Bat Algorithm (MBA) by means of Levee flights and empirically reveal so as to it can beat the average MBA during joined with DAEN.

Revised Manuscript Received on April 30, 2020.

* Correspondence Author

Soundari D.V M.E.*, Assistant Professor, Department of ECE, Sri Krishna College of Engineering and Technology soundaridv@skcet.ac.in

Kuralarasi R, PG Student, Department of ECE Sri Krishna College of Engineering and Technology 18epcm006@skcet.ac.in

Kalieshwari S, PG Student, Department of ECE, Sri Krishna College of Engineering and Technology. 18epcm003@skcet.ac.in

Kanimozhi M, PG Student Department of ECE Sri Krishna College of Engineering and Technology. 18epcm004@skcet.ac.in

Kanimozhi N, PG Student Department of ECE Sri Krishna College of Engineering and Technology. 18epcm005@skcet.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The overload of this work is outlined as follow: first we need to announce the algorithms that will be make use of to form our IDS pattern. Later, we portray our attempt, define the routine actions and display our analyze consequences. Likewise, we relate our steps with additional method and specify that it can attain improved consequences. Lastly, to finish the conclusion and prospect job.

III. PROPOSED WORK

A. Deep Auto Encoder Network

An auto-encoder include encoder and decoder. Encoder goal is to constrict store input file into a small-dimensional illustration, and decoder recreate input information supported the low-dimension description created through the encoder.

In favor of a agreed training dataset $X = \{x_1, x_2, \dots, x_m\}$ beside m samples, someplace x_i is d -dimensional feature vector, the encoder plans the input vector x_i to an unseen eviction vector h_i concluded a deterministic mapping f_θ as given in (1)

$$h_i = f_\theta(x_i) = s(Wx_i + b), \quad (1)$$

$$s(t) = \frac{1}{1 + \exp^{-t}} \quad (2)$$

Where constraint t disturbs the formula of the role.

The decoder charts bring up the resulting unseen symbol h_i to a reassembled d -dimensional vector y_i in input space as, $y_i = g_\theta(x_i) = s(Wh_i + b)$ (3)

Thus, a failure function is considered by the resulting equation,

$$L(x, y) = \frac{1}{m} \sum_{i=1}^m ||x_i - y_i||^2 \quad (4)$$

Where m is that the whole amount of training dataset.

The focal aim is to find the finest parameters (θ and θ') which can be remarkably reduce the alteration between input and reconstruct output to fulfill the whole training set as,

$$\theta = \{W, b\} = \arg \min_{\theta} L(x, y) \quad (5)$$

B. BAT Algorithm

Bat Algorithm (BA) was put up up by Yang in 2010 (Yang, 2010a) and by the means ecolocation of bats was motivated. These microbats radiate a shrill pulsation and modify their pulse rate since the obstacle is nearer. In sort to explain a sensible bat algorithm, three sweeping statement rubrics are framed:

- Every part of bats aid echolocation to come close the transformation in the middle of an obstacle and sense move away.

- Their steps forward is defined by their station in space (x_i) during aimlessly bats fly and velocity (v_i). These limitations are subtracted based on an erratic wavelength (λ), frequency ($freq_{min}$) and noise (A_o) to look for prey. Besides, bats can change the rate of their emanated pulsations and the amount of pulse emission ($r \in [0, 1]$), depending on the contiguity of their goal.

- The loudness can be varying in a number of customs, nevertheless we undertake that it revises beginning a bigrate (A_o) to a minutest constant value (A_{min}). BA is cloud intelligence algorithm which accomplishes rummage around via a population of negotiators. In BAT algorithm to each agent i has a present site $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})^t$ and a up-to-date flying velocity, $v_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d})^t$ where d is the trouble dimension. To find the most favorable side, each

bats keep informed its position and velocity conferring to the following calculations:

$$freq_i = freq_{min} + (freq_{max} - freq_{min}) \cdot \beta(1)$$

$$v_{i,j}^t = v_{i,j}^{t-1} + (x_{i,j}^{t-1} - x_{bestj}) \cdot freq_i \quad (2)$$

$$x_{i,j}^t = x_{i,j}^{t-1} + v_{i,j}^t \quad (3)$$

wherever $\beta \in [0, 1]$ remains a casual vector drawn from a static dispersal. As detailed prior, the bat will decreases his noise (A_i) and rises his pulse emission tempo (r_i) when it is closer to the target:

$$A_i^{t+1} = \alpha \cdot A_i^t \quad (4)$$

$$r_i^{t+1} = r_i^0 \cdot [1 - e^{-\gamma t}] \quad (5)$$

where α ($0 < \alpha < 1$) and γ ($\gamma > 0$) are remain stable.

By the side of each one of the iteration, the fitness value be upgraded, while $A \rightarrow 0$ and $r \rightarrow r_0$. In order to upgrade the variation of the identified solutions,

$$x_{new} = x_{old} + \delta \cdot A_t^* \quad (6)$$

where $\delta \in [-1, 1]$ is a random number and A_t^* is the standard noise of all bats at iteration t . The explanation stays symbolized via the site of the unit, to each specific from the flight partakes its own position and velocity which are reorganized affording to their fitness value.

IV. WORKING PROCESS

Our idea takes three basic phases: first we spread over MBA to work out the best solution, afterwards we usage BA to establish the parameters for DAEN. To end, DAEN perceives network spells with the finest limits and sample dimension is reduced.

A. Dataset

We considered using the NSL-KDD data set, which comprises network attack. We picked this dataset because it is unreservedly open and is an upgraded version of KDD-Cup (Tavallae et al., 2009). The effective attacks from NSL-KDD descent into one of the following classes: Denial of Service (DoS), User to Root (U2R), Remote to Local (R2L) or Probing. Besides, here remain two different sets of files: training and testing. It is essential to remark that test data contains attack types not in the training data and hence it will agree us to appraise the classier for mysterious attacks.

B. Feature Selection

Feature selection approaches can be separated keen on: scalar methods and vector methods (Dua and Du, 2011a). In the next we make known to the Modified Bat Algorithm and describe in what manner it can be altered a vector tactic for feature selection.

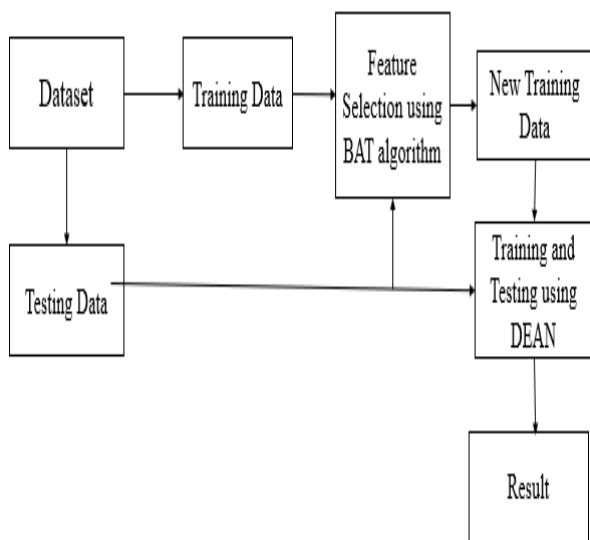


Fig.1. Block Diagram

1. Modified BAT Algorithm

The Modified Bat Algorithm(MBA) is a modified kind of BA that clarifies the bat's movement in a d-dimensional dual cosmos. Therefore, the bat location need to be explained as a route of binary synchronizes and the bat can interchanged diagonally the turns of a hypercube. Certain bat i , its coordinates are totaled by means of a sigmoid function as follows:

$$x_{i,j} = \begin{cases} -1 & \text{if } S(v_{i,j}) > \delta \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Where, the sigmoid function is :

$$S(v_{i,j}) = \frac{1}{1 + e^{-v_{i,j}}} \quad (8)$$

and $\delta \in U(0,1)$. From now, the bat's view can be got by way of binary values. Intended for trait choice process we need to resolve the characteristic detachment that will boost up their concerts of our classifier. In request to become accustomed MBA for feature selection we can deliberate the bat's place as the subclass of features and the bat's manages as the incidence (uncertainty the coordinate is one) or absence (if the coordinate is zero) of a feature. The fitness function intended for MBA will be the exactness of Autoencoder next it has been trained through the features signified via the bat's position.

MBA efforts to enhance the characteristic division at every iteration, such that the classifier's accuracy improves. The core chapters of this method are specified below:

- All bat require to explain their feature selection and it has several locations. The bat trains and estimates the Auto Encoder classifier based on the selection.
- Subsequently all bats have been assess the total optimum strength worth is formed.
- Pulse rate and loudness can be updated during improve their positions of bats, then need to find the best solution.
- The frequency and velocity of each and every bat will update though their unimprovement of fitness fuction.
- lastly, if the novel fitness value is greatest than the worldwide worth, the strike will add to his beat speed and reduce his volume. The worldwide best be changed.

Individually dataset is presented to feature selection technique and more than a few features are designated. Unlike features are chosen for each detached

dataset. 12 features are carefully chosen after the perfect dataset. Then, these are given to following module.

2. Flow Chart

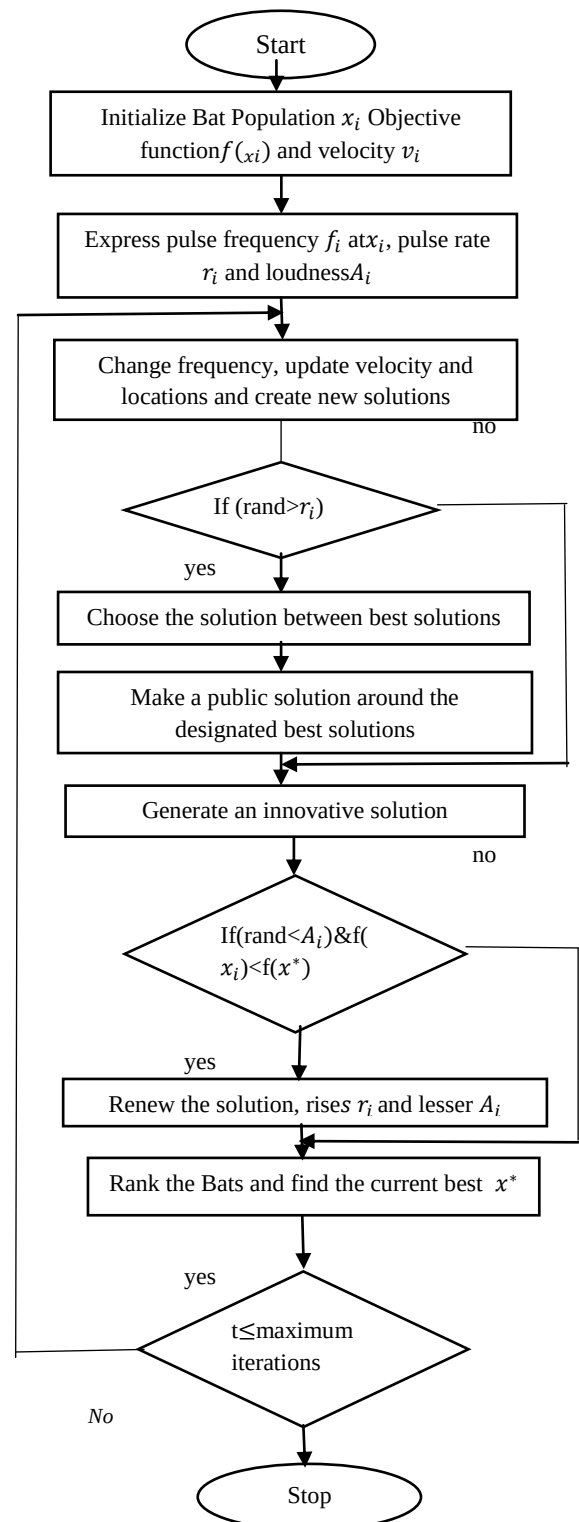


Fig.2 Flow chart for Bat algorithm

C. Classification using Autoencoder Network

bottomless automobile-program base interruption discovery scheme turns in half stages: preparation and taxing.

Inside the preparation stage, the scheme practices a preparation dataset and create a typical model continued the future Deep Auto-Encoder (DAE) model. At that moment, the technique employs for recognizing the tag of indiscernible data (test dataset) in the testing phase to evaluation the presentation of the scheme whether stipulation it is employed on top of-row.

The planned bottomless automobile-encoder structural design intended for fashion the replica in the preparation stage is made known in Fig.2. DAE involves of three sorts of layer: contribution, concealed and production layer.

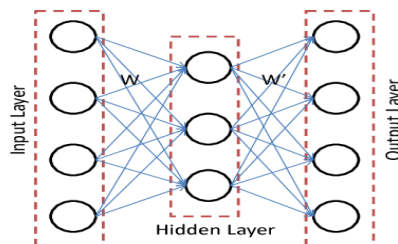


Fig.3. Autoencoder Architecture

The area of classification is to exactly envisage the objective session for every single occasion in the records. A classification mission instigates along with a data set in which the class labels are recognized. Unusual classification algorithms custom dissimilar methods designed for detecting relations. These affairs are briefly explained during a model, which could then be beneficial to a dissimilar data set in which the class labels are mysterious. Classification models are established by relating the anticipated values to identified target values in an array of test data.

D. Performance Measures

Table 1: Formulas for confusion matrix

Confusion matrix	Formulas
True positive	$TPR = \frac{TP}{(TP + FN)}$
True Negative	$TNR = \frac{TN}{(FP + TN)}$
False Positive	$FPR = \frac{FP}{(FP + TN)}$
False Negative	$FNR = \frac{FN}{(TN + FN)}$

The following parameters measured for estimating the performance of the projected idea.

Table 2: Formulas for different parameters

Detection Rate	Accuracy	False Positive Rate
$DR = \frac{TP}{(TP + FN)}$	$AC = \frac{TP + TN}{(TP + TN + FP + FN)}$	$FA = \frac{FP}{(FP + TN)}$

V. RESULT AND DISCUSSION

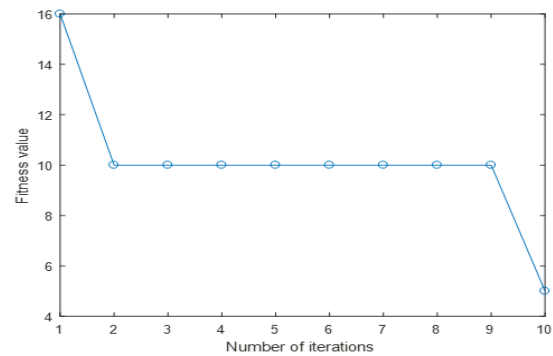


Fig.4 Fitness Function

In this iterative result, when the iteration starts initially, we can get maximum fitness value which is equal to feature selection on further when the process reaches 2nd iteration the fitness value decreases linearly and remains stable for remaining iterations up to 10.

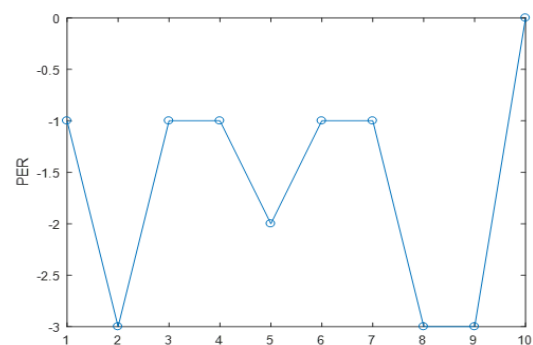


Fig.5 Pulse Emission Rate

In this graph, we can see that the Pulse Emission Rate for each microbats that indicate the position and velocity of each bats in sample dimension.

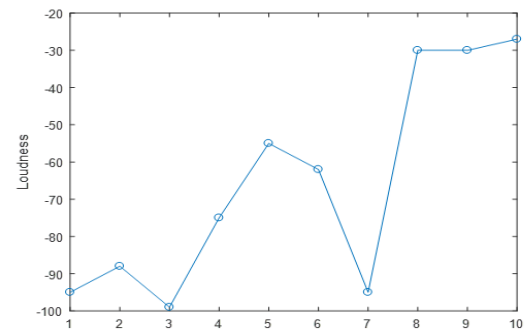


Fig.6 Loudness

In this graph indicate the loudness for each bat. The loudness and Pulse Emission value changes with increase in number of iterations.

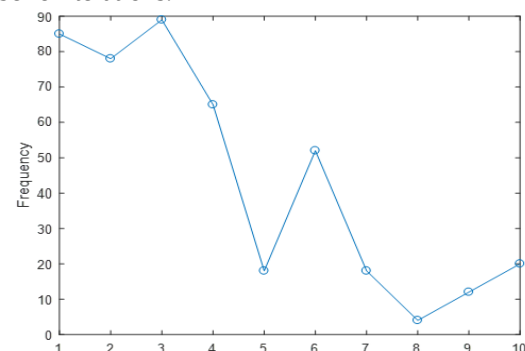


Fig.7 Frequency

In this graph, frequency of each bat is updated during number of iterations.

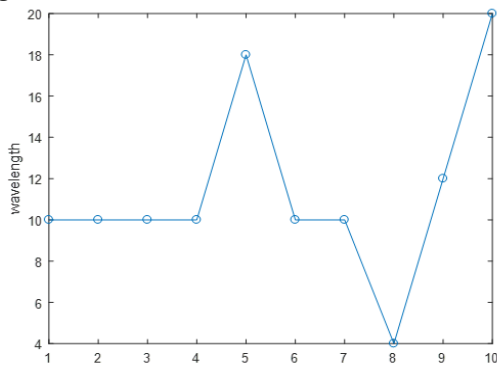


Fig.8 Wavelength

In this graph, it be able to mechanically regulate the wavelength of their emit pulse depending on the target values through iterations.

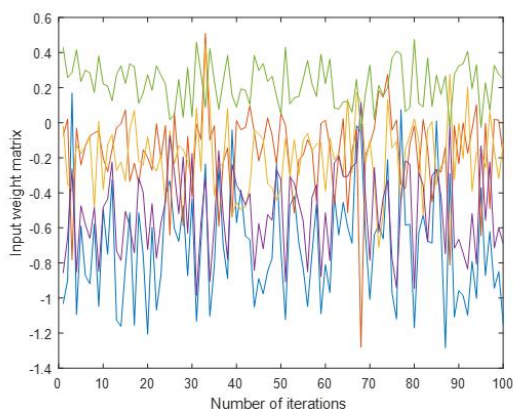


Fig.9 Different Number of weight matrix

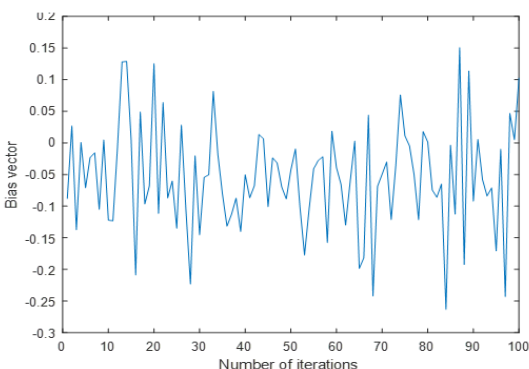


Fig.11 Bias vector

In this input weight matrix and bias vector, different number of inputs are influenced in the several iterations and biasing between inputs and weight up to these 100 iterations.

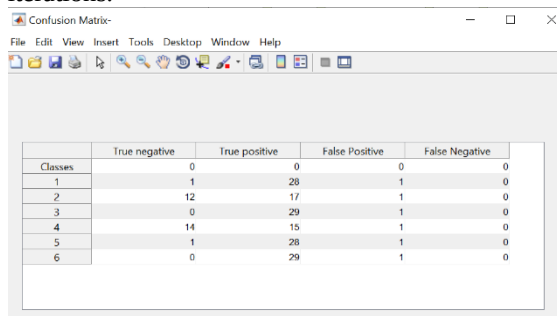


Fig.12 Evaluation of Detection Rate

These are the performance evaluation of our planned work, with this table matrix true positive value is higher than the true negative values.

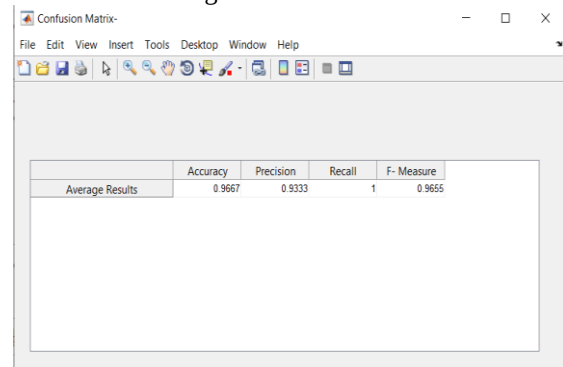


Fig.13 Average results for attacks

The traffic distribution of NSL-KDD is attained. It gives the accuracy which can reach up to 83%.

Method	Ipsweep	Neptune	Nmap	PortswEEP	Teardrop
GA-DBN	83.04%	80.32%	82.66%	93.00%	90.32%
FC-ANN	83.70%	48.12%	83.33%	93.11%	86.07%
BPNN	80.35%	89.12%	89.12%	25.58%	80%
MBA-DEAN	97.88%	99.09%	96.32%	93.54%	92.43%

Fig.14 Performance comparison using different algorithms with various numbers of attacks

VI. CONCLUSION

In this work, we planned an innovative NIDS model that associates DAEN as classifier through the modified Bat Algorithm. The leading participation of this algorithm is it selects the features and to classify different type of attacks. we aid the NSL-KDD dataset for intrusion standard and it possess the testing and training datasets. By implementing the dataset, we can predict Accuracy, Attack detection rate using DAEN classifier and also pulse emission rate, frequency, wavelength of bat motion. The overall system arrangement is made such that it categories the faults and reach its correctness than further system structure, which can arrive at 96%.

REFERENCES

1. A. C. Enache and V. Sgrciu, "A feature selection approach implemented with the binary bat algorithm applied for intrusion detection," in International Conference on Telecommunications and Signal Processing, 2015, pp. 11–15.
2. Md. Zahangir Aalom, Paheding Sidike, Tarek M. Taha, Vijayan K. Asar State preserving extreme learning machine: A monotonically increasing learning approach Neural processing letters, vol. 45, issue 2, pp. 703–725, April, 2017.
3. E. Hodo, X. Bellekens, A. Hamilton, P.-L. Dubouilh, E. Iorkyase, C. Tachtatzis, and R. Atkinson, "Threat analysis of IoT networks using artificial neural network intrusion detection system," Tetrahedron Lett., vol. 42, no. 39, pp. 6865–6867, 2017.
4. E. Viegas, A. O. Santin, A. França, R. Jasinski, V. A. Pedroni, and L. S. Oliveira, "Towards an energy-efficient anomaly-based intrusion detection engine for embedded systems," IEEE Trans. Comput., vol. 66, no. 1, pp. 163–177, Jan. 2017.

5. L. Junsong, "Research on BP algorithm based on neural network and its application in network intrusion detection," *Mod. Electron. Technique*, vol. 40, no. 11, pp. 91–94, 2017.
6. A. Mehmood, A. Khanan, M. M. Umar, S. Abdullah, and K. A. Zainol, "Secure knowledge and cluster-based intrusion detection mechanism for smart wireless sensor networks," *IEEE Access*, vol. 6, pp. 5688–5694, 2017.
7. Zhida Li, Ana Laura Gonzalez Rios, Guangyu Xu, and Ljiljana Trajković "Machine Learning Techniques for Classifying Network Anomalies and Intrusions" *IEEE* 2019.
8. Parisa Lotfallahtabrizi, Yasser Morgan "A Novel Host Intrusion Detection System Using Neural Network" *IEEE* 2018.
9. Feng Chen, Zhiwei Ye, Chunzhi Wang, Lingyu Yan, Ruoxi Wang "A Feature Selection Approach for Network Intrusion Detection Based on Tree-Seed Algorithm and K-Nearest Neighbor" *IEEE* 2018.
10. Mostafa darkaie , Reza Tavoli "Providing a way to scale back the warning rate in network intrusion detection systems using the multilayer perceptron technique and backpropagation algorithm" 5th Conference on Knowledge-Based Engineering and Innovation, Iran University of Science and Technology, Tehran, Iran, IEEE2019.



second International Conference on "Recent Advances in Engineering and Technology-ICRAET 2019" in February 2019.

Kanimozhi. M., currently pursuing a degree in Engineering in the communication system stream under ECE branch at Sri Krishna College of Engineering and Technology. Have published a paper entitled "Cognitive Technology in Various Applications" in e-journal Volume 7, Issue 1 On March 2019 in International Journal of Scientific Research & development. Additionally, have presented paper entitled "Application of Selective Region Growing Algorithm in Lung Nodule Segmentation" in 2018th fourth International Conference on Devices, Circuits and Systems (ICDCS) 2018.

Kanimozhi. N., currently pursuing a Master degree in Engineering in the communication system stream under ECE branch at Sri Krishna College of Engineering and Technology. Received B.E degree in ECE from Sri Krishna College of Engineering and Technology. Have published a paper entitled "A patient Monitoring in Ambulance by using Internet Protocol" in e-journal Volume 7, Issue 1 On March 2019 and "Animal footprint Identification for safety purpose" in the second International Conference on "Recent Advances in Engineering and Technology-ICRAET 2019" in February 2019.

AUTHORS PROFILE



Soundari.D.V., Assistant Professor, ECE Department, Sri Krishna College of Engineering and Technology. Received M.E degree in VLSI Design from Bannari Amman Institute of Technology. Received B.E degree in ECE from Bannari Amman Institute of Technology Published 10 papers in Scopus indexed journals. She is having 8 years of experience in teaching and research and published 9 papers in various reputed journals. Also, she delivered guest lectures and acted as notable keynote speaker in various national conference and international conference, and also her meticulous effect on research leads the students of SKCET to obtain funds from the industry as a part of the consultancy work. She completed nearly 10 consultancy projects in the field of electronics & VLSI



Kuralarasi.R., currently pursuing a Master degree in Engineering in the communication system stream under ECE branch at Sri Krishna College of Engineering and Technology. Received B.E degree in ECE from Dhanalakshmi Srinivasan Engineering College. Have published a paper entitled "A patient Monitoring in Ambulance by using Internet Protocol" in e-journal Volume 7, Issue 1 On March 2019 in International Journal of Scientific Research & development. Additionally, have presented the paper entitled "Animal footprint Identification for safety purpose" in the second International Conference on "Recent Advances in Engineering and Technology-ICRAET 2019" in February 2019.



Kalieswari.S., currently pursuing a Master degree in Engineering in the communication system stream under ECE branch at Sri Krishna College of Engineering and Technology. Received B.E degree in ECE from Dhanalakshmi Srinivasan Engineering College Have published a paper entitled "A patient Monitoring in Ambulance by using Internet Protocol" in e-journal Volume 7, Issue 1 On March 2019 in International Journal of Scientific Research & development. Additionally, have presented the paper entitled "Animal footprint Identification for safety purpose" in the