Intrusion Detection System using Deep Neural Network and Regularization of Hyper Parameters with Adam Optimizer

R. Sekhar, K. Thangavel.

ABSTRACT ***Intrusion Detection Systems (IDSs) study is unavoidable in the field of network security due to the present target oriented attacks for taking secret data of an organization. Classifying and detecting attacks are highly technical and tedious. In the existing models, the accuracy of intrusion detection in network traffic is different for different algorithms. This paper proposed a better intrusion detection system using Deep Neural Network with regularization of the hyper parameters. Adam optimization is proposed to optimize the weights in the neural network. The proposed system consists of six phases namely data collection, data framing, splitting of data for training and testing, pre-processing/encoding, regularization with Adam Optimizer, training and testing. It produces the better accuracy in detection process than the existing Deep Neural Network model. The bench mark data set NSL_KDD is collected and processed in the suggested system.***

Keywords: Intrusion Detection Systems (IDSs), Deep Neural Network(DNN), Rectified Linear Unit (ReLU), Adaptive moment estimation (Adam) and Stochastic Gradient Decent (SGD).

1. INTRODUCTION

Although the technology of detection of smart network attacks and design of intrusion detection model is improving day by day, the hackers are using complex attack patterns for intrusions and hacking a network [1]. The self and adoptive learning ability of neural network and deep learning network can be used to enhance the detecting capability of the present intrusion detection systems [2]. IDS are classified on the basis of signature,anomaly, host and network. The signature related intrusion detection system draws patterns which can be matched from the stored pattern from database [12].

Revised Manuscript Received on April 25, 2019.
R. Sekhar,Dy Director, DRC, National Intelligence Grid, MHA, Bengaluru-560 063. Email:sekharmca1974@gmail.com
K. Thangavel, Department of Computer Science, Periyar University, Salem, Tamil Nadu.
Email:drkthangavel@periyaruniversity.ac.in

False Positive and False Negative are the two problems due to which IDS could not detect all threats correctly.Network and Host based Intrusion detection system are also deployed in a network to achieve end user strategy to prevent intrusions. Normally, when IDS detect ongoing attack in the computer system, it raises alarm for taking action by the administrator [3]. The attacks are performed by attackers to steal data from a repository or harass a financial institution or resourceful office.

The content of paper is presented in various sections. The related works are studied in section 2. Methodology of the suggested work is presented in section 3. Here data collection, data framing, data splitting for training and testing, pre-processing/encoding, Regularization with Adam Optimizer, training and testing were explained. The Results and discussion including performance analysis of proposed work with existing DNN are elucidated in section 4. The paper is concluded in section 5 by suggestions for future work.

2. RELATED WORK

JingwenTian, MeijuanGao et al[1] had proposed Radial Basic Neural Network based intrusion detection system. RBNN has the capacity of converging cost functions swiftly and better approach of functions. The given intrusion data set like NSL_KDD can be learned and remembered well to find out the test patterns and other information of its kind. This neural network has better detecting capability and is very effective and reliable.

Mohammad RezaNorouzian et al. [3] had proposed Artificial neural networks for detection of attacks and intrusions. Multilayer perceptron were taken for implementation purpose. For propagating the error function, Back Propagation Neural Network algorithm was introduced. An analysis showed that in the three layer neural network 90.78% correct classification happened.

Li Xiangmei et al. [4] proposed a hybrid neural network based intrusion detection system. Here the limitations and strong points of Genetic algorithm and Levenberg-Marquardt

algorithms were studied and modified to hybrid neural network algorithm. Here the output of classification is evaluated for binary, multiple class attacks and repeated for few to many training samples. It is seen that the Hybrid Neural Network algorithm is superior to the improved Genetic algorithm and Levenberg-Marquardt algorithm based intrusion detection system.

Chualong Yin et al. [5] had proposed a RNN-IDS and evaluated the model for detection of intrusions in bench mark dataset. Here the performance was studied for binary and multiclass classifications. The learning rate and weight of neurons are optimized in the network. The output of the model was compared with other machine learning and Artificial Neural Network models. From the experimental results, it is proved that the performance of RNN based intrusion detection system is better than other machine learning algorithms and artificial neural networks for both multiclass and binary mode classifications.

SasankaPotluri et al. [14] had proposed accelerated deep neural networks for enhanced Intrusion detection system. It is stated that, identification of attacks and contents of packets are possible through deep learning neural networks. It has the capability of parallel computing for enhancement of performance. Optimum number of training data set is required for classifying the attack patterns correctly. It can classify attack of particular category only and not fit for all attack patterns. Also the classification capability depends on core and processor.

Tuan A Tang et al .[13] had proposed software defined networking and selected deep learning models for detection of intrusions. Here in a SDN environment, deep learning concept was applied for intrusion detection for flow based anomalies. It is proved from their implementation that the suggested system is potentially strong in detecting anomaly based intrusions.

3. PROPOSED METHODOLOGY

This paper proposes a better intrusion detection system using DNN and Adam Optimizer. The hyper parameters such as beginning rate of learning, the count of hidden neurons and the count of hidden layers are changed for getting better accuracy in detection of intrusions in an intrusion detection system. Further ReLU activation function is used in all hidden layers and Softmax layer is used in output layer. ReLU is used as the default activation function in deep learning neural networks.ReLU converges the cost function to a great extent than the tanh and sigmoid activation functions. This proposed system consists of five sections. Initially, data collection phase, in this stage the input data is taken from the NSL_KDD dataset in csv form. Then, the input data is converted into numerical data by data framing .Then the framed data is divided into testing and training data frame where training data comprises to 70 % and that of testing data 30 %. Then label and data are encoded to binary form and sent for training to the deep learning model.

In the last stage preprocessed data is trained and tested using DNN optimized with stochastic gradient decent and DNN optimized with Adam optimizer. Finally, the result is evaluated and compared with existing DNN by the confusion matrix formed by two deep learning network models. The flow chart for the suggested model is as under:

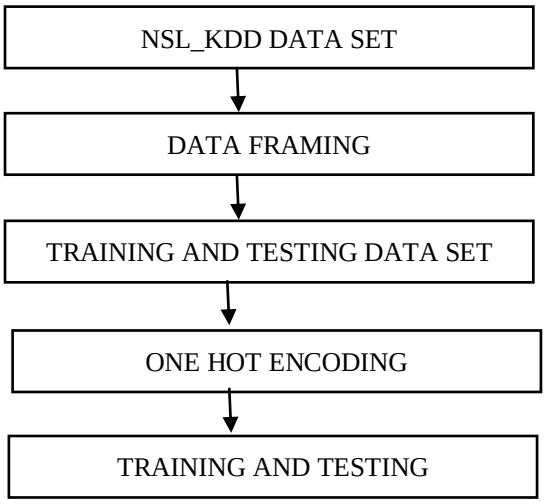


Figure 1: Block diagram of proposed IDS.

3.1 NSL_KDD Dataset

NSL_KDD data set is the benchmark data set for performing research work on intrusion detection system. It contains 41features and the 42nd feature is the decision attribute or class label, where the pattern of attack is mentioned. It has 49,00,0000 rows of data.

In this proposed work, 87153 data are used for training with 41 features. The total training data is divided into 34066 numbers of normal data, 23469 of Dos attack data, 28744 numbers of Probe attack data, 800 numbers of r2l attack data and 44 numbers u2r attack data. Similarly, for testing 37352 data are taken with 41 features. It is divided into 14573 numbers of normal data and 10117 numbers of Dos attack data, 12328 numbers of probe attack data,326 numbers of r2l attack data and 8 numbers of u2r attack data.

3.2 Data Conversion

The input data is converted into data frames. Here some characters such as, tcp, icmp, etc. are replaced by numbers. Example in protocol type feature, tcp is replaced with number 1, icmp is replaced with number 2 and udp with 3 etc. Similarly in service type feature, RSTR is replaced with number 1, SF with 2, S3 with 3, RSTO with 4, SI with 5, REJ with 6 and SO with 7 etc.

3.3 Data Partition :

The attack type is classified into five classes. They are Denial of Service (DoS), User-to-Root (U2R), Remote-to-Locals (R2L), Probe and Normal.[15] Data frame contains 39 attacks which are pod, smurf, teardrop, apahe2,back, land, neptune, udpstorm, processtable, mailbomb, buffer-overflow, httptunnel,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Retrieval Number: E10370585S19/19©BEIESP

loadmodule, perl, rootkit, xterm,sqlattack, worm,ps, ftp-write,snmpguess, guess-password, imap, multihop, phf, spy, warez client, warezmaster, sendmail, ipsweep, nmap, portsweep, satan, saint,snmpgetattack, named, xlock, xsnoop and mscan. In this model, 70 % of dataframe is allotted for training and 30 % for testing.

3.4 Encoding integer data frames to binary

The classified attack patterns is encoded into binary form through one hot encoding. The 41 features are converted into 118 hot codes/features.

3.5 Training and Testing

In training phase 53087 numbers of attack data which belongs to 4 classes and 34066 numbers of normal data with 41 features from every class, total 87153 numbers of data from 5 classes are chosen for training dataset. Here, the training data are trained with deep neural network and the parameters are optimized with stochastic gradient decent. Again the parameters are optimized with Adam optimizer to get better accuracy in detection of attacks.

a) Deep Neural Network

The neural pattern of network takes the input and passes the information to all the hidden layers and compute the probable output [9]. Here the input features are fed into the neural network and trained to identify the intrusion patterns. After training, the network model is asked to find out the similarities and differences without any man intervention [10].

Deep learning and Neural network architectures has many models. Each model works well for some data set. Hence it is not possible to train and study the performance of a deep learning model for all available data sets of different kinds.

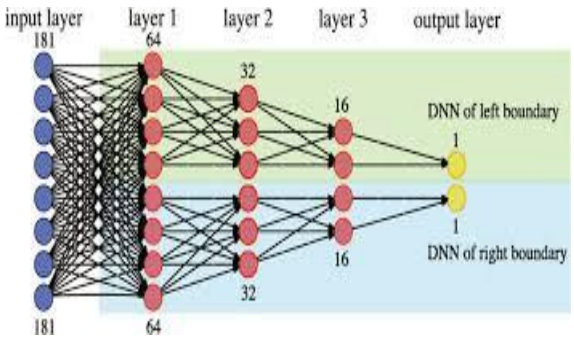


Figure 2: DNN with three hidden layer of the proposed model

The Deep Neural Network model follows feed forward neural network [6]. Here the network passes data from input layer to hidden layer and then to the output layer. The process never return back from output layer to input layer through hidden layers. On reading the data set, the deep neural network forms a pattern of neurons and allot weights to interlink the neurons. The weights are random numerical values only. The weights and the data set (input) are multiplied and creates a value as output between zero and one. Sometimes the network have problem in identifying a pattern. In this case, the neural weight is required to be optimized by proper algorithm till the pattern can be better understood by the network to process the data. in which data flows from the input layer to the output layer and never return back. At first, the DNN creates a map of virtual neurons and assigns random numerical values, or "weights", to connect the neurons. The weights and inputs are multiplied and return an output between 0 and 1. If the network didn't accurately recognize a particular pattern, an algorithm would adjust the weights. That way the algorithm can make modifications in certain parameters until it determines the correct mathematical manipulation to fully process the data. In the proposed model, 3 hidden layers are used.

b) Stochastic gradient decent

Stochastic gradient descent is a type of algorithm, which calculate the cost function for each training sample and update the weight parameter. It does not go for the mean of the cost function results [16]. For larger datasets, gradient decent algorithm, may fail and hence the stochastic gradient decent algorithm is used for converging the cost function.

c) Adam Optimization:

For each Parameter w_j,

$$c_t = \beta_1 \times c_{t-1} - (1 - \beta_1) \times d_t$$

$$m_t = \beta_2 \times m_{t-1} - (1 - \beta_2) \times d_t^2$$

$$\Delta w_t = - \eta \frac{c_t}{\sqrt{m_t + \epsilon}} \times d_t$$

$$w_{t+1} = w_t + \Delta w_t$$

η :Beginning Learning rate
d_t :Gradient when time= t along w_j
c_t :Exponential average of gradients along w_j
m_t :Exponential average of squares of gradients along w_j
β_1, β_2 :Hyper parameters

Thedefault values of $\eta = 0.001$, $ß_1 = 0.9$, $ß_2 = 0.999$ and $\varepsilon = 10^{-8}$.

The steps involved in Adam Optimizer are as under:

i. The gradient of existing parameters should be calculated. Further the squares of all the gradients with respect to each parameter is also calculated.
ii. The moving average term of first order momentum and second order momentum should be updated.
iii. The unbiased average of first order momentum and second order momentum should be calculated.
iv. Updated weight= Average unbiased first order momentum/ $\sqrt{\text{unbiased}}$ average of second order momentum.
v. The weights should be updated.

Adaptive Moment Estimation (Adam) is a method that calculates adaptive learning rates for each hyper parameter of the deep learning algorithm.[15] Adam keeps

an exponentially decaying average of past gradients similar to momentum and behaves like a heavy ball with friction. Adam can be looked at as a combination of RMS prop, AdaGrad and Stochastic Gradient Descent with momentum to provide an optimization algorithm that can handle sparse gradients on noisy problems.Adam learns the fastest and more stable than the other optimizers, it doesn't suffer any major decreases in accuracy. Adaptive methods tend to converge quickly towards sharper minima. The initial learning rate is fixed as 0.01 and other parameters will be tuned adaptively by adam optimizer to get optimized weight and better accuracy in intrusion detection.It minimizes the cost function with minimum number of epochs and converges swiftly. In this model it converges at epoch no 52.

d) ReLU activation function:

Rectified Linear Activation function (ReLU) is better for activating hidden layers of Deep Neural Networks including MLP. Networks that use the rectifier function for the hidden layers are referred to as rectified networks. Adoption of ReLU may easily be considered one of the few milestones in the deep learning revolution. ReLU is normally used to achieve linear behavior, computational simplicity, and representational scarcity and train MLPs [16].

In the proposed model, three hidden layers with 340, 200, 10 number of neurons in the first, second and third hidden layers are devised and ReLu activation function is used for activation.

e) Softmax activation function

The softmax function squashes the outputs of each unit between 0 and 1, just like a sigmoid function. Here the gross of outputs is equal to one. The output of softmax function follows categorical type of probability distribution. Mathematically the softmax function is shown below [16].$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K}(e^{z_k})}$

Where, z - vector quantities of the input information to the output layer, j- output units(j = 1, 2, ..., K.).In the output layer, the number of neurons are limited to the number of classes and in the proposed model, the neurons are restricted to 5.

4. RESULT AND DISCUSSION

The proposed intrusion detection system is implemented in the working platform of python 2.7 version with system configuration and Processor: Intel core i7, CPU Speed: 3.20 GHz, Operating System: Linux and RAM: 8 GB

4.1 Database Description

Several researchers have used back propagation neural network approach for their experimentation. Mostly KDDCup'99[11] dataset and NSL_KDD dataset [8] are the bench mark datasets for the research work on intrusion detection system. It includes training and testing datasets. The testing dataset includes not only known attacks from the learned/trained data but also new attacks. 87153 data are considered for the training phase, and 37352 data are passed for testing phase with 41 features.

4.2 Performance Analysis

In this section, the implementation result and its performance are analyzed. By applying the statistical measures, for example, sensitivity, specificity, accuracy, FDR, FPR, PPV, and NPV, the performance of this proposed intrusion detection system is examined.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

4.3 Comparative Analysis

a) The performance analysis of DNN with various optimization algorithms are provided in table 1:

Table1: Comparative study of performance of Deep Neural Network with various Optimizers

Optimiser	Overall Accuracy	Train Accuracy	Test Acy	Epochs
Adam	96.522	96.3	96.5	52
AdaGrad	96.032	95.9	96	138
RMSProp	96.515	96.1	96.3	58
Adadelta	94.480	94.4	94.5	31
Adamax	95.106	95.4	95.5	62
Nadam	95.711	95.5	95.7	62

b) The performance analysis of proposed Deep Neural Network with Adam Optimizer and SGD optimizer when learning rate is 0.01 are illustrated in table 2 and 3.

Table2 : PROPOSED DNN+ ADAM (LR=0.01)

Performance	Dos	Normal	Probe	R2L	U2R
Sensitivity	.9830	.9897	.9459	.4386	0
Specificity	.9762	.9805	.9936	.9994	1
PPV	.9762	.9805	.9936	.9994	1
NPV	.9936	.9933	.9739	.9950	.997
FPR	.0203	.0194	.0063	.0005	0
FNR	.0169	.0102	.0540	.5613	1
FDR	.0169	.0102	.0540	.5613	1
Accuracy	.9805	.9841	.9779	.9945	.997

Table3 : DNN+ SGD (LR=0.01)

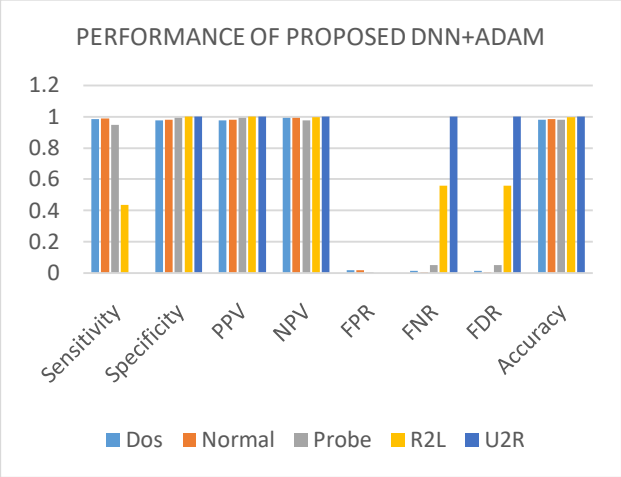
+

Performance	Dos	Normal	Probe	R2L	U2R
Sensitivity	.6530	.6130	.6819	.2091	0
Specificity	.6510	.6210	.6316	.2106	1
PPV	.6312	.6118	.5871	.2346	1

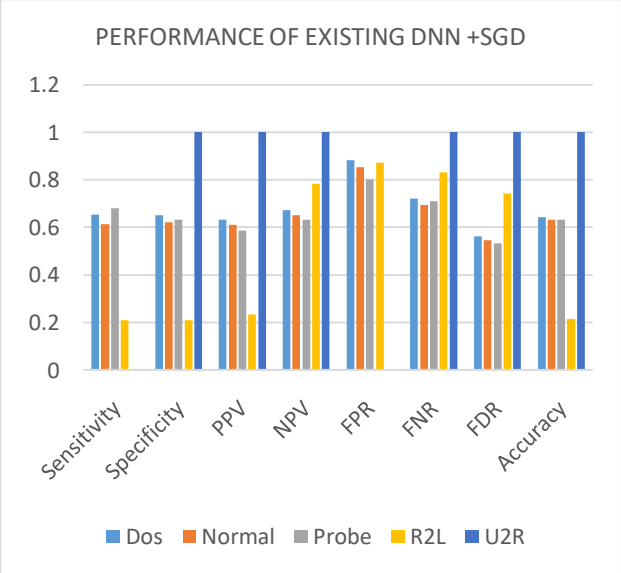
Retrieval Number: E10370585S19/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

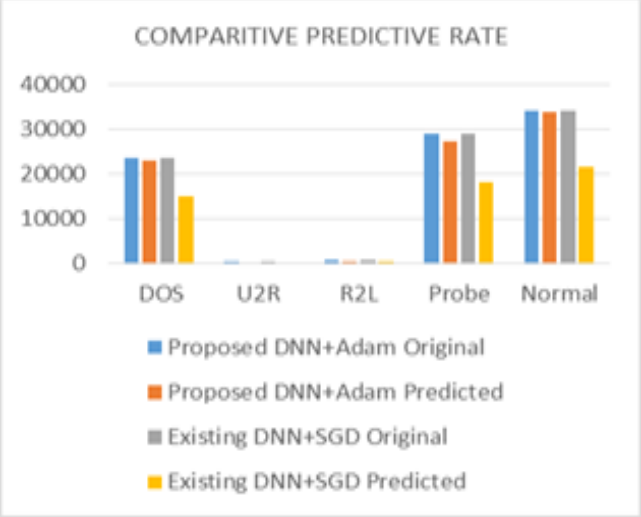
NPV	.6718	.6514	.6312	.7821	.997
FPR	.8812	.8540	.8017	.8718	0
FNR	.7218	.6954	.7113	.8314	1
FDR	.5610	.5471	.5317	.7415	1
Accuracy	.6418	.6310	.6324	.2158	.997



Graph 1: Performance of proposed DNN + ADAM



Graph2: Performance of existingDNN + SGD



Graph 3 : Comparative Predictive Rate

From the above tables and figures, it is evident that the Deep Neural network along with Adam optimizer with learning rate 0.01have produced better accuracy of intrusion detection with 96.52%. The DNN with SGD have an accuracy of 65.2%. The prediction rate of proposed DNN+Adam is better for all attacks except U2R. It is also studied that the accuracy depends on various training parameters such as learning rate, no of hidden layers, no of neurons in hidden layers, etc.

5. CONCLUSION AND FUTURE WORK

In this paper, an intrusion detection system using DNN and Adam optimizer algorithm was implemented. The efficiency of the implemented model was evaluated by NSL_KDD bench mark intrusion detection data set. The performance of the proposed system was analyzed using the data which were taken from the NSL_KDD dataset. In this proposed method for training, DNN + Adam algorithm produced better detection and accuracy rate. The performance analysis has shown that the proposed intrusion detection system has given a better accuracy, sensitivity and specificity. The collation outcome elucidate that the implemented DNN optimized with Adam model has better precision, specificity and sensitivity than the prevailing systems of its kind. The precision level of the proposed method is 96.52% but that of the existing method of DNN+SGD is 65.2%. Also the detection rate varies on change of training parameters such as quantum of hidden layers, quantum of hidden neurons and the beginning rate of learning etc. Hence this proposed intrusion detection system could efficiently detect the attacks than the existing DNN + SGD algorithms. The accuracy in detection rate can be enhanced by using deep learning based models such as Restricted BoltzmannMachine (RBM), Deep BoltzmannMachine (DBM), Deep Belief Neural Network (DBNN), Spiking Neural Networks, Convolution Neural Networks etc as classifiers and optimizing training parameters.

ACKNOWLEDGEMENT

The second author sincerely acknowledges UGC for supporting the partial financial assistance under SAP-DRS(II) grant number F 5-6/2018/DRS-II (SAP-II).

REFERENCES

[1]JingwenTian, MeijuanGao, Fan Zhang, "Network Intrusion Detection Method Based on Radial Basis Function Neural Network", IEEE, 978-1-4244-4589-9/$25@2009.

[2] Luan Qinglin, Lu Huibin, "Research of Intrusion detection based on neural Network optimized by adaptive genetic algorithm", Computer Engineering and Design, Vol. 29, no 12, pp. 3022-3025, 2008.

[3] Mohammad Reza Norouzian, SobhanMerati"Classifying Attacks in a Network Intrusion Detection System Based on Artificial Neural Networks", ICACT 2011, ISBN 978-89-5519-155-4, Feb 13-16, 2011.

[4] Li Xiangmei,QinZhi, "The Application of Hybrid Neural Network Algorithms in Intrusion Detection System" IEEE, 978-1-4244-8694-6/11/$26@2011

[5]Chuanlong Yin, Yuefei Zhu, JinlongFei and Xinzheng He, "A Deep Learning Approach for IntrusionDetection Using Recurrent Neural Networks" IEEE Access, Vol.5, pp. 21954-21961, 2017.

[6] Lei Xiao, Xiaohui Chen, and Xinghui Zhang, "A Joint Optimization of Momentum Item and Levenberg-Marquardt Algorithm to Level Up the BPNN's Generalization Ability" Research Article, , Mathematical Problems in Engineering Volume 2014, Article ID 653072, 10 pages http://dx.doi.org/10.1155/2014/653072.

[7] Kloft, M. Lakshov,P " Security analysis of online centroid anomaly detection and Detection Using Recurrent Neural Networks" IEEE Access, Vol.5, pp. 21954-21961, 2017.

[8] MajdLatah ,LeventToker "Towards an Efficient Anomaly-Based Intrusion Detection for Software-Defined Networks. Publication in IET Networks,

[9] P Aggarwal and S K Sharma " Analysis of KDD Dataset attributes-classwise for Intrusion detection", Procediacomput.Sci, vol,57,pp 842-851, 2015

[10] PriyankaAlekar " Cloud based intrusion detection system using BPN classifier" International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 11, November 2018, Pages 1-7.

[11] RoshaniGaidhane C. Vaidya Dr. M. Raghuwanshi "Intrusion Detection and Attack Classification using Back-propagation Neural Network" International Journal of Engineering Research & Technology (IJERT),ISSN: 2278-0181 Vol. 3 Issue 3, March – 2014, pages 1112-1115.

[12] Ahmed Elsherif, "Automatic Intrusion Detection System Using Deep Recurrent Neural Network Paradigm" Journal of Information security and cybercrimes research, Vol 1, Issue 1, June 2018, Pages 28 -41.

[13] Tuan A Tang, LotfiMhamdi, Des McLernon, " Deep Learning Approach for Network Intrusion detection in Software Defined Networking" IEEE 978-1-5090-3837-4/16@2016 Pages 1-6

[14] SasankaPotluri, "Accelerated Deep Neural Networks for enhanced Intrusion Detection System" IEEE, 978-1-5090-1314-2/16 @2016.

[15] S Devaraju, S Ramakrishnan, "Performance comparision of Intrusion detection system using neural network with kdd dataset", ICTACT journal on soft computing, 2014

[16] Duchi.JHazan E and Singer,Y, "Adaptive subgradient methods for online learning and stochastic optimization". Journal of Machine learning research, 12, 2121-2159, 2011.

ABOUT AUTHORS

Dr. K. Thangavelis currently working as the professor of Computer Science at Periyar University, Salem, Tamil Nadu, India. His Research Areas are Data Mining, Bioinformatics, Image Processing, Soft Computing, Mobile Computing, Bio-informatics, Optimization Algorithms, Pattern recognition, Machine Learning and Deep Learning. Under his research supervision, 29 Scholars have been graduated. He has edited 3 books and published 191 research papers in reputed International Journals. He is the recipient of Tamil Nadu State Science Scientist Award (TANSA) for the year 2009 and Sir C.V.Raman award for the year 2013 instituted by Periyar University.

R Sekhar, is presently working as a Deputy Director in NATGRID, MHA, GoI. He has 8 years of experience in IT and Data Centre. He has been graduated from University. His research area is Intrusion detection system using deep learning. Currently he is pursuing research in Periyar University, Salem, Tamil Nadu, India.