International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 125-132 www.ijaem.net ISSN: 2395-5252

Security Enhancement Model for Intrusion Detection System using Classification Techniques:

Rakhi Shukla, Dr. Aarti Kumar
Research Scholor, Rabindranath Tagore University, Bhopal, Madhya Pradesh
Head-MOOC & Content Rabindranath Tagore University, Bhopal, Madhya Pradesh

Date of Submission: 01-01-2023 Date of Acceptance: 08-01-2023

ABSTRACT
With the growing use of computer networks across different fields and applications, network security is becoming increasingly important. Today, with the rapid growth and the broad application of the Internet and Intranet, computer networks have brought great convenience to people's life and work. Many researchers have introduced more innovative techniques to detect intrusions in recent years, such as machine learning, data mining, evolutionary approaches, and optimization techniques. Intrusion detection is considered one of the emerging research areas nowadays. This paper presents the overview for detecting intrusion using different techniques and also discusses the other intrusion detection techniques.
Keywords:- Intrusion detection, Machine learning, Deep learning, Classification, KDDCUP.

I. INTRODUCTION

An intrusion detection system (IDS) monitors anomalous activities and differentiates between normal and abnormal behaviors (intrusion) in a host system or a network. Intrusion Detection Systems are a mechanism, which protects resources and data from unauthorized access, misuse, and malicious intrusions in a distributed computing environment. Machine learning techniques, such as Neural Networks, Support Vector Machines, Naïve Bayesian Classifiers, etc., are standard techniques for intrusion detection. IDSs constantly monitor and analyze the system, which allows the machine learning model to recognize everyday/normal behavior. This allows the model to detect abnormal/anomalous behavior and react with the appropriate response. The standard dataset used for IDS developments and testing is the KDD99 dataset. The IDS is tasked with monitoring and analyzing network activity to differentiate between normal and anomalous activities. If the anomalous activity goes undetected, this could potentially cause severe damage to the infrastructure and reliability of a computer system. Therefore, the detection rate of anomalous activity must be maximized. Simultaneous to anomalous activity detection, the IDS must minimize the false positive rate to avoid undue hassle and confusion. False positives do not put the system at risk. However, they can become a problem if the rate at which they occur is high, limiting the IDS's ability to provide reliable and precise results. The balance between detection rate and false-positive rate is the key for an effective IDS. The balance between detection and false-positive rates becomes more challenging when regular activity and anomalous activity are not static. The activity on the network can change, and the IDS must be aware of this change and adapt accordingly. If not, the ability of the IDS to provideaccurate and reliable results is greatly diminished. Therefore, an IDS must adapt to different environments, potentially bringing different activities and behavior unseen by the IDS.

II. LITERATURE REVIEW

Here we present the literature survey for the network-based intrusion detection system, as we know that the demand of computer network is increasing day by day, so we have to protect our network from the intruder or attacker, in this section we review the various researcher's paper for the efficient and attack free network or system. Here we present the author review with their respective details is mentioned in reference sections.

Anomaly-based IDS makes the detection of the data packets in the network traffic analyze the packets of data that unfit the typical profile that has been created. Ripon Patgiriet. Al [1]applied machine learning algorithms to detect intrusions effectively. Machine Learning is a statistical

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 125-132 www.ijaem.net ISSN: 2395-5252

method for handling regression and classification tasks. These methods include Support Vector Machines (SVM) for regression and classification, Naive Bayes for classification, and k-Nearest Neighbors(KNN) for regression and classification.

In recent years, one of the main focuses within NIDS research has been the application of machine learning and shallow learning techniques such as Naive Bayes, Decision Trees, and Support Vector Machines. Shone et al. [2] proposed a novel deep learning model to enable NIDS operation within modern networks. The model they propose is a combination of deep and shallow learning, capable of correctly analyzing a widerange of network traffic. More specifically, they combine the power of stacking their proposed non-symmetric deep auto-encoder (NDAE) (deep-learning) and the accuracy and speed of Random Forest (RF) (shallow learning). They have practically evaluated their model using GPU-enabled TensorFlow and obtained promising results from analyzing the KDD Cup '99 and NSL-KDD datasets.

Chuanlong et al. [3]explored how to model an intrusion detection system based on deep learning, and we propose a deep learning approach for intrusion detection using recurrent neural networks (RNN-IDS).
Moreover,They analyze the performance of the model in binary classification and multiclass classification. The number of neurons and different learning rate impacts the performance of the proposed model. Here, they compare it with J48, artificial neural network, random forest, support vector machine, and other machine learning methods proposed by previous researchers on the benchmark data set.

An intrusion detection system (IDS) monitors anomalous activities and differentiates between normal and abnormal behaviors (intrusion) in a host system or a network. The IDS must maintain a high intrusion detection rate (DR) while simultaneously maintain a low false alarm rate (FAR).James Brownet al.[4] focus on detecting anomalous network packet instances using an Evolutionary General Regression Neural Network (E-GRNN). They use simulated network data obtained from the UNB ISCX Intrusion Detection Evaluation Dataset. They extracted features from the application layer protocols (e.g., HTTP, FTP, SMTP, etc.) used in network activities. The E-GRNN takes the standard GRNN by evolving the sigma value used for training and a feature mask, which extracts salient features from the dataset by removing irrelevant and redundant features. The E-GRNN model reduces the computational complexity of the network anomaly detection and increases the accuracy as well. The E-GRNN reduced the feature set by an average of 60% while maintaining an average detection rate of 93.63% and a false positive rate of 2.82%. This shows the efficacy of the EGRNN model for network anomaly.

Longjing Liet al. [5] proposed the GINI GBDT-PSO method, a novel hybrid intrusion detection model to improve the performance of network intrusion detection systems. The proposed model first extracts the optimal subset of features from the whole dataset using the Gini index. Then, the GBDT algorithm, a gradient boosting approach, is adopted to detect abnormal connections. In addition, the PSO algorithm is employed to optimize the parameters of the GBDT algorithm in the proposed model. This model can enhance the overall performance for network intrusion detection effectively and improve the detection performance for each type of attack.

Machine learning plays an essential role in building intrusion detection systems. However, with the increase of data capacity and data dimension, shallow machine learning is becoming more limited. Yanqing Yanget al. [6] proposed a fuzzy aggregation approach using the modified density peak clustering algorithm (MDPCA) and deep belief networks (DBNs). To reduce the size of the training set and the imbalance of the samples, MDPCA is used to divide the training set into several subsets with similar sets of attributes. Each subset is used to train its sub-DBNs classifier. These sub-DBN classifiers can learn and explore high-level abstract features, automatically reduce data dimensions, and perform classification well. According to the nearest neighbor criterion, the fuzzy membership weights of each test sample in each sub-DBNs classifier are calculated. The output of all sub-DBNs classifiers is aggregated based on fuzzy membership weights. Experimental results on the NSL-KDD and UNSW-NB15 datasets show that our proposed model has higher overall accuracy, recall, precision, and F1-score than other well-known classification methods. Furthermore, the proposed model achieves better accuracy, detection rate, and false positive rate than state-of-the-art intrusion detection methods.

Network Intrusion Detection System (NIDS) constitutes an essential security tool for monitoring network traffic and identifying network attacks. NIDS can be categorized into three main categories based on the detection method they use in identifying potential attacks as signature-based, anomaly-based, or specification-based NIDS. Malek Al-Zewairiet. al[7], they explore a deep

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 125-132 www.ijaem.net ISSN: 2395-5252

learning binomial classifier for Network Intrusion Detection System is proposed and experimentally evaluated using the UNSW-NB15 dataset. Three different experiments were executed to determine the optimal activation function, select the essential features, and test the proposed model on unseen data.

The evaluation results demonstrate that the proposed classifier outperforms other models in the literature with 98.99% accuracy and 0.56% false alarm rate on unseen data.

Ref. No.	Publication Details	Methods	Dataset	Performance Parameters	Limitations
[1]	IEEE Symposium Series on Computational Intelligence, 2018.	Random forest & Support vector machines	NSL-KDD	Accuracy, Precision, Recall	May used some more classification techniques
[2]	IEEE Transactions On Emerging Topics In Computational Intelligence, 2017.	Deep learning	KDDCUP 99, NSL-KDD	Accuracy, Precision, Recall, F-Score, False alarm	Real-world network traffic Datasets
[3]	IEEE Access, 2017.	Recurrent neural networks	NSL-KDD	Actual positive rate, False positive rate, the Detection rate	To reduce the training time using GPU acceleration
[4]	IEEE, 2016.	Evolutionary General Regression Neural Network	UNB ISCX Intrusion Detection Evaluation Dataset.	Accuracy, True positive rate, False positive rate, Detection rate, False-negative rate	Need more trained dataset with feature reduction techniques
[5]	Journal of Sensors, 2018.	Gini index with Particle swarm optimization	NSL-KDD	Accuracy, Detection rate, Precision, F1-Score, and False alarm	Performance may increase using some machine learning model
[6]	Applied Science Journal, 2018.	Modified Density Peak Clustering Algorithm and Deep Belief Networks	NSL-KDD	Accuracy, Detection rate and False positive rate	Plan to use the adversarial learning method to synthesize U2R and R2L attacks
[7]	International Conference on New Trends in Computing Sciences, IEEE 2017.	Multilayer feed-forward artificial neural network using back-propagation	UNSW-NB15 dataset	Accuracy, False positive rate, Precision, F1-Score, and Recall	The performance will also measure in some other data types
[8]	International Conference on Soft-computing and Network Security, IEEE 2018	Decision Tree model	KDD99 intrusion dataset	Accurate positive, False positive, and Processing time	The performance will also measure in some other performance parameters value.

Table 1: Comparative study for intrusion detection techniques.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 125-132 www.ijaem.net ISSN: 2395-5252

KDD Cup99

The KDD Cup '99 dataset was used in DARPA's IDS evaluation program. The data consists of 4 gigabytes-worth of compressed tcpdump data resulting from 7 weeks of network traffic. This can be processed into about 5 million connection records, each with about 100 bytes. It consists of approximately 4,900,000 single connection vectors, each of which contains 41 features. These include Basic features (e.g., protocol type, packet size), Domain knowledge features (e.g., number of failed logins), and timed observation features (e.g.%, of connections with SYN errors). Each vector is labeled as either normal or as an attack.

It is good to note that the KDD CUP dataset has been widelyused by the researchers, especially for IDS studies. This database contains a standard set of data to be audited, including a wide variety of intrusions simulated in a military network environment. Moreover, it is prettytricky to collectsuch a vast amount of data with a Lab's set-up environment to obtain long-term rawTCP dump data for a network. Thenetwork was operated as if it were a natural environment butsprinkled with multiple attacks.

Sr. No.	Name of Features	Types
1	duration	continuous
2	protocol_type	symbolic
3	service	symbolic
4	flag	symbolic
5	src_bytes	continuous
6	dst_bytes	continuous
7	land	symbolic
8	wrong_fragment	continuous
9	urgent	continuous
10	hot	continuous
11	num_failed_logins	continuous
12	logged_in	symbolic
13	num_compromised	continuous
14	root_shell	continuous
15	su_attempted	continuous
16	num_root	continuous
17	num_file_creations	continuous
18	num_shells	continuous
19	num_access_files	continuous
20	num_outbound_cmds	continuous
21	is_host_login	symbolic
22	is_guest_login	symbolic
23	count	continuous
24	srv_count	continuous
25	serror_rat	continuous
26	srv_serror_rat	continuous
27	rerror_rate	continuous
28	srv_rerror_rate	continuous
29	same_srv_rate	continuous
30	diff_srv_rate	continuous
31	srv_diff_host_rate	continuous
32	dst_host_count	continuous
33	dst_host_srv_count	continuous
34	dst_host_same_srv_rate	continuous
35	dst_host_diff_srv_rate	continuous
36	dst_host_same_src_port_rate	continuous
37	dst_host_srv_diff_host_rate	continuous
38	dst_host_serror_rate	continuous

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 125-132 www.ijaem.net ISSN: 2395-5252

39	dst_host_srv_serror_rate	continuous
40	dst_host_rerror_rate	continuous
41	dst_host_srv_rerror_rate	continuous

Table 2: List of KDDCUP features.

DATASET DESCRIPTION

To get to know about the data and find relations between data, it is necessary to discuss data objects, data attributes, and data attributes. All the features present in KDD datasets have two types, either **symbolic or continuous**.

Continuous data have an infinite number of states. Continuous data is of float type. There can be many values between any numbers. For example, the height of any person may vary between 5.2 to 6.2.

Symbolic data are distinctive in their own right on any sized data sets, small or large. For example, it is not unreasonable to have data consisting of variables, each recorded in a range. Likewise, we can formalize a computer security-based engineering company as having a knowledge base consisting of the files and data. Such data and files are more aptly described as concepts rather than standard data, and as such, are also examples of symbolic data.

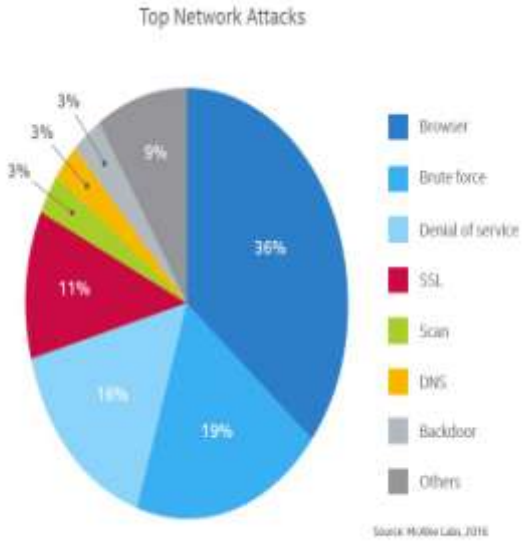

Figure 1: Top network Attacks [McAfee Labs, 2016].

Intrusion detection techniques used by the researchers are:

RANDOM FOREST

Random Forest is a supervised classification algorithm.A random forest can be used for regression and classificationtasks [1]. Random forest classifier forms a bunch of several decision trees from randomly selected features. Then, itcalculates votes from the different decision trees for eachpredicted target, and the highest voted class is considered thefinal prediction. Let training set is provided as: [A1, A2, A3,A4] with their corresponding label as [B1, B2, B3, B4] randomforest can generate three decision tree taking a subset of input,for example

1.[A1, A2, A3]
2.[A1, A2, A4]
3.[A2, A3 A4]

Finally, it predicts based on the majority of votes from eachdecision made by the decision trees. The random-forest algorithmbrings extra randomness into the model while growing thetrees. Instead of searching best feature while splitting the node, itsearches for the best features among a random subset of features.

SUPPORT VECTOR MACHINE

Support vector machine is a supervised classification algorithm.SVM is a discriminative classifier that separates definedby separating hyper-plane. More clearly, SVM takes training data and separates data into categories divided by a clear gapcalled the hyper-plane. SVM tries to find the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 125-132 www.ijaem.net ISSN: 2395-5252

best or optimalhyperplane, which has the most significant distance from the nearestpoint, in high dimensions, separatingthe training setinto categories. Support vectors are those vectors that arenearest to the hyper-plane. The goal is to select a hyperplane with a margin as much as possible between hyperplane and any vector within the training set, giving a greater chance of newdata being classified correctly.

RECURRENT NEURAL NETWORKS

Recurrent neural networks include input units, output units,and hidden units, and the hidden unit completes the mostimportant work. The RNN model essentially has a one-wayflow of information from the input units to the hidden units. The synthesis of the one-way information flow from the previous temporal concealment unit to the current timing hiding unit is shown in the figure below. Here can regard hidden units asthe storage of the whole network, which remembers the end-to-end information. When we unfold the RNN, we can findthat it embodies deep learning. A RNNs approach can beused for supervised classification learning.Recurrent neural networks have introduced a directionalloop that can memorize the previous information and apply it to the current output, which is the essential differencefrom traditional Feed-forward Neural Networks (FNNs). Thepreceding output is also related to the current output of asequence, and the nodes between the hidden layers are nolonger connectionless; instead, they have connections. The output of the input layer and the output of the last hidden layer act on the input of the hidden layer.

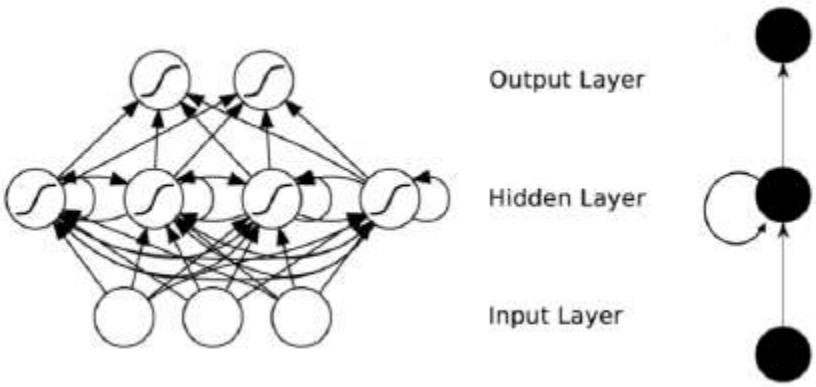

Figure 2. Recurrent Neural Networks (RNNs) [3].

GINI INDEX

Usually, the dataset for network intrusiondetection contains many features. However, not every featurecontributes to the task of detecting intrusion. Feature selection,which can remove redundant or irrelevant features, isa crucial step for intrusion detection. Based on the optimalfeature space here, we can enhance the speed of traininga classifier for network intrusion detection and improveits detection performance. The goal of feature selection is to get a group of significantfeatures from the whole dataset, such that theseselected features are essential for training a classificationmodel. Gini index toundertake the mission of feature selection. The Gini index,which Corrado Gini, an Italian statistician, developed, and sociologist, in 1912, was used initially to measure thestatistical dispersion of income distribution across differentpopulation sectors [5].

DEEP LEARNING

Deep Learning algorithms are a modern update to artificialneural networks that exploit abundant, affordable computation. Deep learning permits an algorithm to learn a representationof data with various levels of generalization. Thesemethods have been applied to visual object recognition, objectdetection, detecting network intrusion, and many other domains. A deep learning algorithm can be trained in asupervised and unsupervised way.Deep Learning algorithm in a supervised way:Convolutional neural network (CNN) usually is trainedin a supervised way. CNN is now the benchmark model for computer vision purposes. The CNN architecture is used tostructure 2D images, and the most crucial acknowledgmentof CNN is face recognition and intrusion detection [13].

C4.5 DECISION TREE MODEL

A decision-making tree is a decision advocate systemrepresented as a tree graph.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 125-132 www.ijaem.net ISSN: 2395-5252

Decision tree learning takes it as a predictive model to observe an item (represented in the branches)to conclude the item's target value (represented in the leaves).Target variable helps to classify the tree models; treestructure contains leaves and branches representing class labelsand junctions. A decision tree specifically performs indecision analysis, decisions, anddecision-making. Togenerate a decision tree, an algorithm is being used in C4.5 evolved, and It also alluded asa statistical classifier.C4.5 construct decision trees by getting data from training set as
in ID3, using the approach of information entropy [8].

PROBLEM IDENTIFICATION

Among these mechanisms, intrusion detection systems(IDSs) play a vital role in protecting computing infrastructuresfrom attackers and intruders. The balance between detection and false-positive rates becomes more challenging when regular activity andanomalous activity are Few things are explored for system and systematize standard practices in the area ofevaluation of such systems.

1. Researchers In this article, the author applied the different machine learning techniques for intrusion detection techniques like the support vector machine as a classification and random forest classification technique [1]. In future work, we may also use some other classification techniques and optimization techniques for optimal results. We also used feature selection techniques, reduced the feature value, and enhanced the existing system's performance.

2. The author presents the deep learning model [2] for the intrusion detection system using the KDD datasets, and these datasets are divided into two categories normal and abnormal. Abnormal categories include Dos, probe, U2R, and R2L; with the deep learning model author used here, the number of hidden layers is 8, 16, and 24; in the future, we may increase the number of hidden layers and also reduce the computation time.

3. Here[4],the authors present the regression techniques for the intrusion detection system and compute the value of performance parameters like accuracy, false-negative rate, actual positive rate, and detection rate. The balance between detection and false-positive rates becomes more challenging when regular activity and anomalous activity are not static. not static. The activity on the networkcan change, and the IDS must be aware of this change andadapt accordingly. If not, the ability of the IDS to provideaccurate and reliable results is greatly diminished. Therefore,an IDS must adapt to different environments, potentially bringing different activities and behavior unseen by theIDS.

III. CONCLUSION

Network Intrusion Detection Systems (NIDS) have been developedrapidly in academia and industry in response to theincreasing cyber-attacks against governments and commercialenterprises globally. The annual cost of cybercrime is continuouslyrising. The most devastating cyber crimes are caused by malicious insiders, denial of services, and web-basedattacks.With the increasing variety and complexity of IDSes, the development of IDS evaluationmethodologies, techniques, and tools has become a key research topic.

> The activity on the network can change, and the intrusion detection system must be aware of this change and adapt accordingly. If not, the ability of the intrusion detection system to provide accurate and reliable results is greatly diminished. Therefore, an intrusion detection system must adapt to different environments, which potentially bring different activity and behavior unseen by the intrusion detection system; in future work, we may increase the performance of an existing system to reduce the false alarm rate.

In the future, we plan to implement an intrusion detection model based on machine learning algorithms and improve the existing system's performance.

REFERENCES

[1]. Ripon Patgiri, Udit Varshney, Tanya Akutota, and Rakesh Kunde, "An Investigation on Intrusion Detection System Using Machine Learning," IEEE 2018, pp 1684-1691.

[2]. Shone, N, Tran Nguyen, N, Vu Dinh, P and Shi, "A Deep Learning Approach to Network Intrusion Detection," IEEE Transactions on Emerging Topics in Computational Intelligence, 2017, pp 1-11.

[3]. Chuanlong Yin, Yuefei Zhu, Jinlong Fei, Xinzheng He, "A Deep Learning Approach for Intrusion Detection Using

Recurrent Neural Networks," IEEE 2017, pp 21954-21961.

[4]. James Brown, Mohd Anwar, Gerry Dozier, "An Evolutionary General Regression Neural Network Classifier for Intrusion Detection," IEEE 2016, Pp 1-5.

[5]. Longjing Li, Yang Yu, ShenshenBai, Jianjun Cheng, Xiaoyun Chen, "Towards Effective Network Intrusion Detection: A Hybrid Model Integrating Gini Index and GBDT with PSO,"Journal of Sensors, 2018, Pp 1-10.

[6]. Yanqing Yang, Kangfeng Zheng, Chunhua Wu, XinxinNiu, Yixian Yang, "Building an Effective Intrusion Detection System Using the Modified Density Peak Clustering Algorithm Deep Belief Networks," Applied Science Journal, 2019, Pp 1-25.

[7]. Malek Al-Zewairi, Sufyan Almajali, Arafat Awajan, "Experimental Evaluation of a Multi-Layer Feed-Forward Artificial Neural Network Classifier for Network Intrusion Detection System," International Conference on New Trends in Computing Sciences, IEEE 2017, pp 167-173.

[8]. M. Mazhar Rathore, Faisal Saeed, Abdul Rehman, Anand Paul, Alfred Daniel, "Intrusion Detection using Decision Tree Model in High-Speed Environment," International Conference on Soft-computing and Network Security, IEEE 2018, Pp 1-5.

[9]. L. Khalvati, M. Keshtgary, N.Rikhtegar, "Intrusion Detection based on a Novel Hybrid Learning Approach,"Journal of AI and Data Mining, 2018, Pp 157-162.

[10]. Ali Safaa Sadiq, BasemAlkazemi, SeyedaliMirjaliliNoraziah Ahmed, Suleman Khan, Ihsan Ali, Al-Sakib Khan Pathan, KayhanZrar Ghafoor, "An Efficient IDS Using Hybrid Magnetic Swarm Optimization in WANETs," IEEE Access 2018, Pp 29041-29052.

[11]. Aleksandar Milenkoski, Marco Vieira, Samuel Kounev, Alberto Avritzer, Bryan D. Payne "Evaluating Computer Intrusion Detection Systems: A Survey of Common Practices," ACM Computing Surveys, Vol. 48, September 2015, pp 1-41.

[12]. Rafah Samrin, D Vasumathi," Review on Anomaly-based Network Intrusion Detection System," ICEECCOT 2017, pp141-147.

[13]. Nasrin Sultana,Naveen Chilamkurti,Wei Peng,RabeiAlhadad," Survey on SDN based network intrusion detection system using machine learning approaches," Springer Nature 2018, pp 1-9.