International Journal of Advanced Engineering Research and Science (IJAERS)
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-8, Issue-4; Apr, 2021
Journal Home Page Available: https://ijaers.com/
Journal DOI: 10.22161/ijaers
Article DOI: https://dx.doi.org/10.22161/ijaers.84.16

Intrusion Detection System in Software Defined Networks using Machine Learning Approach

Jayasri P, Atchaya A, Sanfeeya Parveen M, Ramprasath J

Department of Information Technology, Dr MCET, India

Received: 08 Jan 2021;

Received in revised form:

25 Feb 2021;

Accepted: 18 Mar 2021;

Available online: 16 Apr 2021

***Keywords*— Naïve Bayes, k-means clustering, Weka, SDN, KDD cupp99**

***Abstract*—** *Now a days, Network Security is becoming the most challenging task. As a result in the growth of internet, the attacks in the network has also been increased. This can be hold back by the intrusion detection system, it identifies the unwanted attacks and unauthorized access in the network. The comprehensive overview of the detailed survey is analyzed with the existing dataset for identifying the unusual attacks in the network. Here machine learning classification algorithms is used to detect several category of attacks. The machine learning techniques can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. In this paper KDD cup99 is used to evaluate the machine learning algorithms for intrusion detection system. Here we have implemented the experiment on intrusion detection system which uses machine learning algorithms like Naïve Bayes and k-means clustering algorithm.*

I. INTRODUCTION

Software Defined Networking (SDN) is a reach to networking that uses software-based controllers or application programming interfaces to meet up with fundamental hardware infrastructure and direct traffic on a network. Software defined networking is a reach via which we take the control plane away from the switch allot it to a centralized unit called SDN controller. Network administrator can outline traffic via a centralized console without having to be in contact with the individual switches. The data plane will still live in the switch and when a packet set foot in a switch, its forwarding activity is clear-cut based on the entries of flow tables, which are pre allotted by the controller.

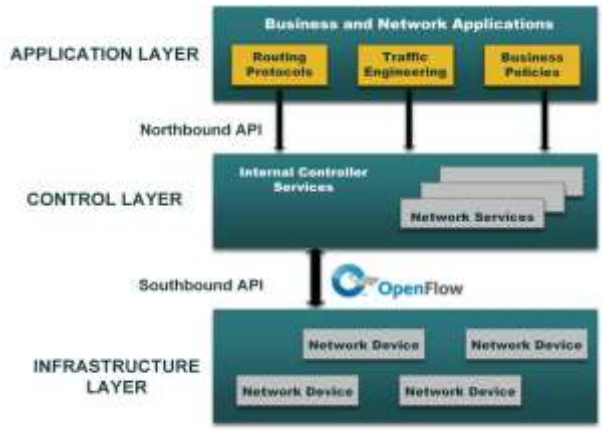

Fig.1: SDN Architecture

Network Virtualization is the process of incorporating hardware and software network assetsand computing into single software software-based entity, that is virtual network and it also helps in incorporating the accessibleassets and splash up the accessible bandwidth to passage, which is unconventional of other and allocated to particular appliance in actual time. Every single channel is unconventionally secured.

Network virtualization is of two types-internal and external.Internal Virtualization refers of using networks by quality in software on single server. It contributes network quality based only on softwares. In networks VMare server is used as common virtualization. However Internal Virtualization is more involute itself and can provide Virtual Switching, Virtual Networking and also Virtual firewall solutions. The advantage of Internal Network Virtualization is it is not hardware dependent and also known as storage virtualization.

External Virtualization is a virtual local area networks and by making use of these systems, they are actually attached to equivalent local networks into various virtual networks and put together by the admin.It utilizesdeviceslike adopters, switches or networks to incorporatesurplus networks into essential units and also uses a CISCO software. The advantages in it is that it has very small footprints due to its devoted nature, so that no other resources can be shared.

Malicious attack can also be called as Malware attacks and it is damage to the device and our cybersecurity.It is provoked by cyber attackers to harm our networks or computer without the victim'sknowledge to gain the personal information. The types of malware attack contains viruses, spyware, and ransomware. This happens on all organized devices and OS together with Windows, macOS, Android and iOS. Malware is even more complex to determine and can get mocked without noticed by the user. There is no interplay needed on the user part other than the looking in on infected webpage.

It is astrike which meant for closing a network and also making inaccessible to the intended user.It happens when the users are unfit to approach information systems, devices or the network resources due to activity of malicious cyber threat. There are two general methods ofDoS - Floodingservices or crashing services. Flood attacks happens when too much traffic is received for the server causes them to slow down the system and also makes to terminate. Also the popular flood attacks get together with Buffer overflow attacks, ICMP flood, SYN flood. An additional type of DoS attack is Distributed Denial of Service (DDoS).

DDoS is a malicious attack to make an online service inaccessible to users, temporarily breaking the service of its hosting server. It is different from other denial service attacks in it uses single Internet connected device with malicious attack. DoS and DDoS attacks can be classified into three types – Volume based Attack, Protocol attacks, Application layer attacks.Volume based attack are the attacks by engrossing them with a global network of scrubbing centres that scale on request to counter multi gigabyte DDoS attacks. Protocol Attacks are the attack by the bad traffic before stick out the site. Application Layer Attacks are by observingthe visitor behaviour blocking bad bots and demanding the suspicious entities. The best methods of DDoS attacks are UDP Flood, ICMP flood, SYN flood, Ping of deaths, Slowloris, NTP Amplification, HTTP flood. DDoS can be exposed using in-line examination of all packets and out-of-band exposition via traffic flow records.

A firewall is a network security device that observers and filters the incoming and outgoing network traffics and plans whether to allow or block the specific traffic security rules. A firewall can be of both software and hardware. The require of Firewall is to secure the system. Without Firewall the system is open to threats and damage. It works as a filtration system for the data attempting to get in to the computer or networks. Firewall scam packets for malicious attack has been already detected as a threats. Incoming traffic is treated differently. The types of firewall are Host-based firewall-It is installed on each network node which masters each incoming and outgoing packets.Network based firewall- these firewalls filter all incoming and outgoing traffic across the networks. A network firewall might have to or more network interface cards.

II. LITERATURE SURVEY

The Survey confer the related works relevant to using KDD dataset for implementing machine learning algorithms to detect the malicious attack. Studies in SDN security have widelysupervened in the enlargement of system that handle security issues connected with the use of Open-Flow. The classifier selection model proposed by the author [1][2][5] made an evaluation in intrusion detection system using the NSL-KDD dataset and also by implementing number of machine learning techniques like Naïve Bayes, SVM, Decision tree, Neural network, K-nearest neighbour algorithm(K-NN) to find their accuracy in each algorithm.

According to another study, [3,4,6] implemented in Scala programming using the ML lib learning library in Apache Spark. The algorithm proposed by the author was support vector machine algorithm against intrusion detection using machine learning on Big data environment. In this proposed method the author imported the dataset and exported it into RDD dataset in Apache Spark and implemented the pre-processing and feature selection phase. Some researches focus on attribute selection algorithm as they increase the computational cost. The author Chibuzor John Ugochukwu, &E.O Bennett focused on selecting the significant attribute and implemented the

detection system based on Bayes net, J48, Random forest and Random tree algorithm in Weka tool. Dataset used was KDD cup99.

The [5, 7, 9]in addition to random tree classifier, Random forest classifier, J48, Naïve Bayes, Decision table they have also implemented multi-layer perception, and also they propose a methodology to detect different types of intrusion within the KDD. In this paper it is known that there is no single machine learning algorithm which can handle the efficiency of different types of attack.

Algorithms, tools and dataset in some of the reference base papers are as follows,

S No	Year	Algorithm used	Tools used	Dataset
1	2018	Naïve Bayes, SVM, Decision Tree, Neural Network, K-Nearest Neighbour Algorithm(K-NN)	Weka	NSL-KDD
2	2018	Spark-Chi-SVM Model	ML lib, Apache Spark	KDD cup99
3	2018	Bayes Net, J48, Random Forest, Random Tree	Weka	KDD cup99
4	2018	Multi-Layer Perceptron, Random Tree Classifier, Random Forest, J48, Naïve Bayes, Decision Tree	Weka	KDD cup99
5	2020	Decision Tree, Random Forest, XG Boost, Support Vector Machine(SVM), Deep Neural Network.	Weka, GNS3	NSL-KDD
6	2019	T-Sne Plot	Weka , hping3	NSL-KDD
7	2019	Naïve Bayes, Decision Tree	Weka	KDD cup99
8	2020	Decision Tree, K-Nearest Neighbour, Support Vector Machine, K-Mean Clustering, Artificial Neural Network	Weka	NSL-KDD
9	2010	Support Vector Machine, Naïve Bayes, K-Nearest Neighbour Algorithm	Weka, WINPCAP	KDD cup99

III. PROPOSED SYSTEM

To detect the malicious attackthe following modules are used,Data Pre-Processing, Attribute Selection, Traffic Grouping and TrafficClassification.

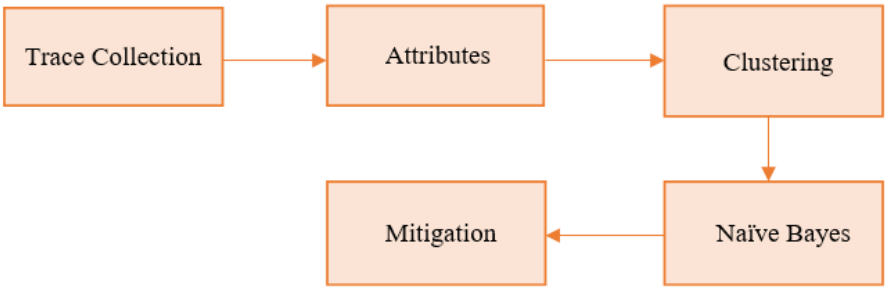

Fig.2: Proposed overall Architecture

3.1 Data Pre-Processing

Data Pre-processing is a data mining technique that converts raw data into an understandable and readable format. Data pre-processing is the beginning of the process. Actual data is frequently insufficient, uncertain, require in obvious behaviours or tendency and is probableofcarrying many errors. Data pre-processing is a demonstrated method tosort out such errors. To make the process simpler data pre-processing is classified into four stages: Data cleaning, Data integration, Data reduction and Data transformation. Data is supposed to be impure if it contains any duplicate or unreal value and noise that interrupt the attribute values and the unfound variables, so

data pre-processing is essential as it is critical in any data mining process as they straightcrash theachievement of the project. It is the conversion applied to the data before suckle to the algorithm.

3.1.1 Steps in data Pre-processing in machine learning

- Acquire the dataset
- Import libraries
- Import the dataset
- Identifying and handling the missing values
- Splitting the dataset into train and test set
- Feature scaling

3.2 Attribute Selection

The mandatory attributes used in Naïve Bayes algorithm to detect the malicious attacks are

- Source mac address
- Source Ip address
- Destination mac address
- Destination Ip address
- Time

3.3 Traffic Groping

To detect the malicious attack here the algorithm used is K-means Clustering Algorithm.K-means clustering is one of the simplest and well-liked unsupervised machine learning algorithms. K-means algorithm determines K number of centroids, and then assigns every data point to the neighbouring cluster, while caring the centroids as small as possible. K clarifies the number of pre-defined clusters that have to be developed in the process, as if K=2, then there will be 2 clusters and for K=3, there will be 3 clusters. It is a centroid-based algorithm. The motive of this algorithm is to keep down the sum of distances between the data point and their matching clusters. The algorithm grasp the unlabelled dataset as input, classifies the dataset into k-number of clusters, and recurrent the process until it does not find the finest clusters. The value of k should be pre-arranged in this algorithm. The k-means clustering algorithm mainly perform two tasks

- Determines the finest value for k centre points or centroids by anrepetition process.
- Assigns each data point to its neighbouring k-centre. Those data points which are neighbour to the particular k-centre, create a cluster.

3.4 Traffic classification

To detect the malicious attack here the algorithm used is Naïve Bayes Classifier.Naïve Bayes algorithm is a supervised learning algorithm, which is depended on Bayes Theorem. It is generally used in text classification that contains a high-dimensional training dataset. Naïve Bayes Classifier is one of the easier and most successful Classification algorithms which helps in defining the fast machine learning modules that can make quick forecasting. It is a probabilistic classifier, which means it forecast on the basis of the probability of an object. A Naïve Bayes classifier suppose that the presence or absence of a specific feature of a class is unrelated to the presence or absence of any other feature, it's naïve because it makes supposition that may or may not turn out to be true. Bayes Theorem is used to determine the probability of a hypothesis with earlier knowledge. It depends on the conditional probability. The formula for Bayes Theorem is given as

$$P(A\backslash B)=P(B\backslash A)P(A) / P(B)$$

IV. RESULT & ANALYSIS

By using Weka tool the malicious attack have been detected. Weka (Waikato Environment for Knowledge Analysis) is a group of machine learning algorithms for data mining tasks. The algorithms can either be applied straight to a dataset or called from our own java code. Weka contains tools for data pre-processing, Classification, Clustering, association rules and visualization. Weka hold up a large number of file formats for the data, and the default file type is ARFF. This tool gets the data file format in comma separated value (csv) or attribute-relation file format (arff). As Weka is written in java which is well documented and allocates integration into our own application. It has the feature of command line interface as all software features can be used from the command line. The KDD 99 dataset is used for the experiments. It is the most used dataset for Intrusion Detection System. As the size of the KDD 99 dataset is very large and has approximately 490000 records with 41 features it is difficult to extract all the data. So the dataset is reduced to meet requirement.

4.1 Result of K-means Clustering algorithm

Final cluster centroids:

		Cluster#	
Attribute	Full Data	0	1
	(549.0)	(279.0)	(270.0)
a1	7.1548	14.0789	0
a2	tcp	tcp	udp
a3	private	http	private
a4	SF	SF	SF
a5	732.2441	1343.2258	100.8963
a6	2128.0874	4052.5233	139.5037
a7	0	0	0
a8	0	0	0
a9	0	0	0
a10	0.071	0.1398	0
a11	0	0	0
a12	0.4353	0.8566	0
a13	0.0036	0.0072	0
a14	0	0	0
a15	0	0	0
a16	0	0	0
a17	0	0	0
a18	0	0	0

Fig.3: Traffic Groping

4.2 Result of Naïve Bayes Clustering

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.966	0.014	0.985	0.966	0.975	0.953	0.992	0.994	udp
	0.986	0.030	0.972	0.986	0.979	0.956	0.987	0.975	tcp
	1.000	0.002	0.800	1.000	0.889	0.894	0.999	0.888	icmp
Weighted Avg.	0.976	0.022	0.977	0.976	0.976	0.954	0.989	0.983	

Fig.4: Traffic Classification

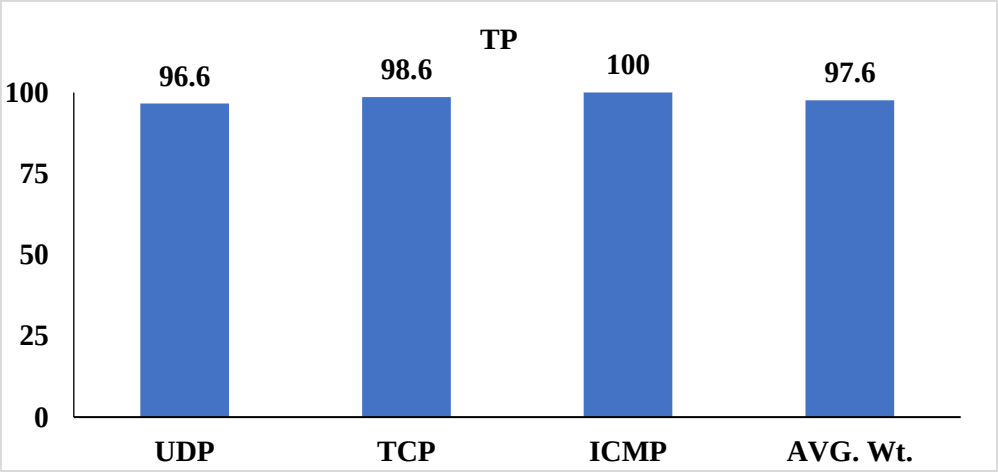

Fig.5: TP rate

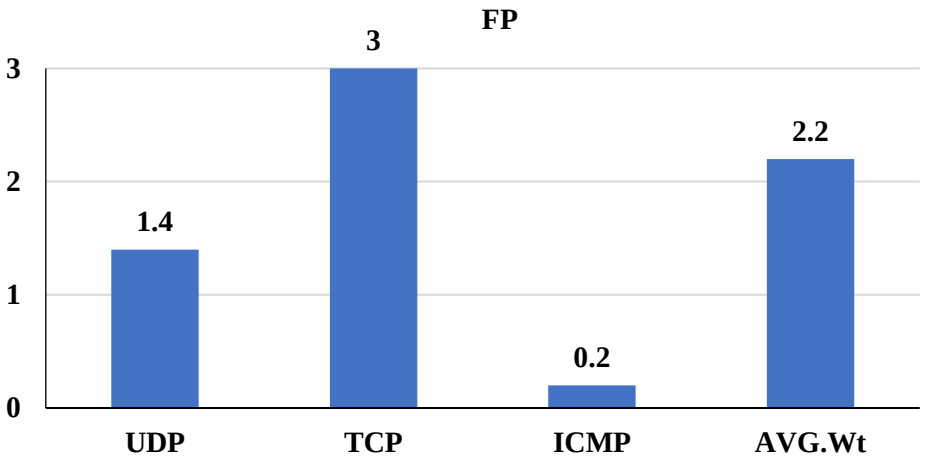

Fig.6: FP Rate

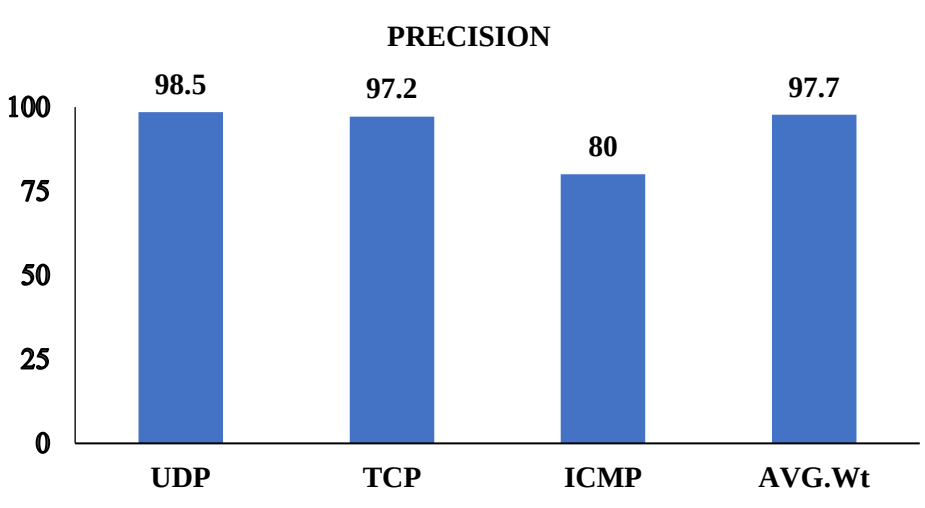

Fig.7: Precision

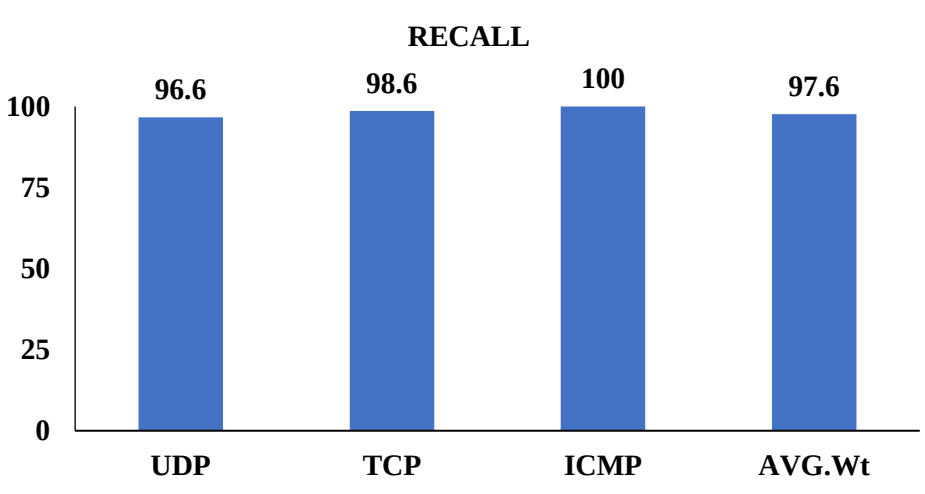

Fig.8: Recall

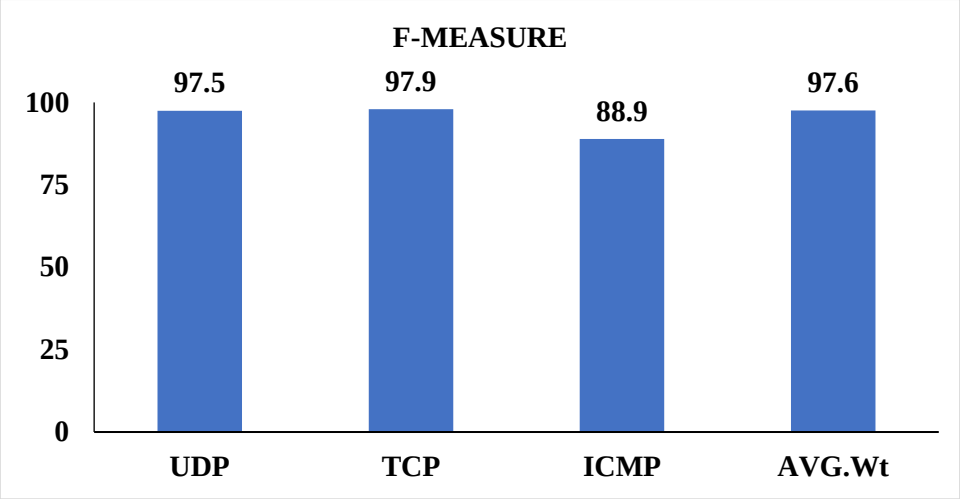

Fig.9: F-Measure

In this paper, the proposed system has 99% of UDP, 98% of TCP, 99% of ICMP efficiency. Comparing to other algorithms, naïve Bayes algorithm proposes a little high efficiency as shown in the above figure. True Positive rate, False Positive rate, precision, recall, f-measure values are calculated using this algorithm, and the graph of all those were figured above.

V. CONCLUSION

As there were several Algorithms in machine learning, in this paper, experiments were performed and tested to evaluate the efficiency and the performance of the following algorithms: Naïve Bayes algorithm and K-means clustering algorithm. The main objective of this paper is to detect the malicious attack by using those two algorithms and hence it was done successfully. Both the algorithms performed were based on the KDD intrusion detection dataset. The rate of the different attacks like DOS, R2l, U2R and PROBE can be found using the KDD dataset. 549 instances of records have been extracted as training data to define the training models for the selected machine learning algorithms. Several performance metrics were computed which are accuracy rate, precision, false negative, false positive, true negative and true positive. Further work will be based on some data mining algorithms applied to Intrusion Detection System to detect the attack.

REFERENCES

[1] Balasamy K, Ramakrishnan S, An intelligent reversible watermarking system for authenticating medical images using wavelet and PSO, Clust. Comput., 22(2), 4431–4442 (2019). https://doi.org/10.1007/s10586-018-1991-8

[2] Balasamy K, Suganyadevi S, A fuzzy based ROI selection for encryption and watermarking in medical image using DWT and SVD, Multimed Tools Appl 80, 7167–7186 (2021), https://doi.org/10.1007/s11042-020-09981-5

[3] K. Balasamy, D. Shamia, Feature Extraction-based Medical Image Watermarking Using Fuzzy-based Median Filter, IETE Journal of Research, (2021) DOI: 10.1080/03772063.2021.1893231.

[4] Ramprasath J, Seethalakshmi V, Secure access of resources in software-defined networks using dynamic access control list, International Journal of Communication Systems, 2020. e4607, https://doi.org/10.1002/dac.4607

[5] Ramprasath J, Seethalakshmi V, Improved Network Monitoring Using Software-Defined Networking for DDoS Detection and Mitigation Evaluation, Wireless Personal Communications, 116, 2743–2757 (2021), https://doi.org/10.1007/s11277-020-08042-2

[6] J Ramprasath, Dr S Ramakrishnan, P Saravana Perumal, M Sivaprakasam, U ManokaranVishnuraj, Secure Network Implementation using VLAN and ACL, International Journal of Advanced Engineering Research and Science, Vol-3, Issue-1, 2349-6495, Jan-2016.

[7] N Krishnaraj, S Smys, A multihoming ACO-MDV routing for maximum power efficiency in an IoT environment Wireless Personal Communications 109 (1), 243-256, 2019

[8] M Sakthivadivel, N Krishnaraj, P Ramprakash, Utilization of big data in oil and gas industries using Hadoop MapReduce technology and HiveQL, Global Journal of Multidisciplinary and Applied Sciences 1 (2), 52-57, 2013

[9] N Krishnaraj, RB Kumar, D Rajeshwar, TS Kumar, Implementation of Energy Aware Modified Distance Vector Routing Protocol for Energy Efficiency in Wireless Sensor Networks, International Conference on Inventive Computation Technologies, 2020

[10] P Ramprakash, R Sarumathi, R Mowriya, S Nithyavishnupriya, Heart Disease Prediction Using Deep Neural Network, International Conference on Inventive Computation Technologies, IEEE, 666-670, 2020/2/26

[11] P Ramprakash, M Sakthivadivel, N Krishnaraj, J Ramprasath, Host-based Intrusion Detection System using Sequence of System Calls, International Journal of Engineering and Management Research, Vandana Publications, Volume 4, Issue 2, 241-247, 2014.

[12] Saroj Kr. Biswas, Intrusion Detection using Machine Learning a comparison study, International Journal of Pure and Applied Mathematics, Volume 118, No.19, 101-114, 2018.

[13] Othman, S.M., Ba-Alwi, F.M., Alsohybe, N.T. et al. Intrusion detection model using machine learning algorithm on Big Data environment. J Big Data 5, 34 (2018). https://doi.org/10.1186/s40537-018-0145-4.

[14] Chibuzor John Ugochukwu, & E.O Bennett, An Intrusion Detection System Using Machine Learning Algorithm, International Journal of Computer Science and Mathematical Theory, Volume 4, No.1, 2018 www.iiardpub.org.

[15] Oqbah Ghassan Abbas, KhaldounKhorzom , Mohammed Assora, 2020, Machine Learning based Intrusion Detection System for Software Defined Networks, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 09 (September 2020)

[16] Shivam Tiwari, VanshikaPandita, Samarth Sharma, Vishal Dhande, Shailesh Bendale, Survey on SDN based Network Intrusion Detection System, International Research Journal of Engineering and Technology(IRJET), Volume 6, 2019, www.irjet.net

[17] Celyn Birkinshaw, Elpida Rouka, Vassilios G.Vassilakis, Journal of Network and Computer Application 136, 2019, https://doi.org/10.1016/j.jnca.2019.03.005

[18] Nivedita S Naganhali, Dr Sujata Terdal, Network Intrusion Detection using Supervised Machine Learning Technique, International Journal of Scientific &Technology Research Volume8, 2019.

[19] Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad, Network intrusion detection system: A systematic study of machine learning and deep learning approaches, Volume32, 2020, https://doi.org/10.1002/ett.4150

[20] Vipin Das, Vijaya Pathak, Sattvik Sharma, Sreevathsam, MVVNS.Srikanth, Gireesh Kumar T, Network Intrusion Detection System Based on Machine Learning Algorithms, International Journal of Computer Science& Information Technology (IJCSIT), Vol2, No.6, 2010.