



Solution Model for Intrusion Detection in Software Defined Networking (SDN) Using Machine Learning

Ahmad Ajiya Ahmad¹, Prof. Souley Boukari², Abdullahi Musa Bello¹,
Mustapha Aliyu Muhammad²

¹Department of Computer Science, Federal University Gashua, Bade L.G.A, Yobe State, Nigeria.

²Mathematical Science Department, Abubakar Tafawa Balewa University, Bauchi State, Nigeria.

ABSTRACT: The purpose of this paper is to explicate and show a conceptual solution model for enhancing Software Defined Networking (SDN) with security and performance. SDN is an emerging and prominence network technology usually deployed in large enterprise networks to offer flexibility in overall network and addressing better Quality of Service (QoS) through simplifying and intelligibly interconnect SDN network infrastructures and components. The solution model is an abstraction of a system that consists of application layer, control layer and infrastructure layer which involved enabling the intrusion detection with machine learning. The model encompasses three Machine Learning (ML) algorithms (J48, Random forest and Naïve Bayes) and Feature Selection Methods (IG, GR and Chi-squared) in the SDN application Layer. The ML algorithms are applied for effective classification of dataset for intrusive attacks detection and these could result in low False Alarm Rate and high detection rate. Feature Selection Method to remove redundancy data.

KEYWORDS: SDN, QoS, ML, Intrusion Detection, Feature Selection, WEKA Environment

Received 28 July, 2021; Revised: 10 August, 2021; Accepted 12 August, 2021 © The author(s) 2021.

Published with open access at www.questjournals.org

I. INTRODUCTION

Cyberspace habitual practice has continued to grow in all the areas and the demand of internet is enlarging every day in our daily life [14]. The key factor causative to this demand is the rising number of mobile broadband consumers, number of network connected machines/devices and high use of data based applications [24]. This development introduces new challenges regarding data capacity usage, privacy policy implementation, incremental deployments of newer infrastructure as well as “scalability” rising multiplicity in network infrastructure [21]. Addressing these challenges introduces Software Defined Networking (SDN), a new prominence network technology systems usually deployed in large enterprise networks to offer flexibility in overall network performance [9]. Its ascertainment is to attain ultimate network flexibility and addressing better Quality of Service (QoS) by simplifying and intelligibly interconnect SDN network infrastructures and components [23]. It also improves network management and offers a good quality of service while succeeding the preferred network scalability [25]. SDN is a network organising technique which is innovative over the traditional/legacy network. According to [12], indicated that, the legacy network is integrated with a single aligned controller for the management and maintenance of processing overhead, resource utilization and bandwidth consumption. In large legacy network, networking devices such as routers, packet switches, and LAN switches, embrace both forwarding (data) and control layer together making it hard to adapt or adjust to the network structure and operation to large-scale enterprise network, end systems, virtual machines, and virtual networks [8]. In contrary, SDN fragmented data and control layer into a solitary entity, this Separation of the data and control layers is a defining characteristic of SDN Architecture [12]. One of the applications acknowledged for the split-up of the control and infrastructure layers is precisely considered today is the security features of such framework [23]. The main aim of the paper is to show solution model for intrusion detection in a Software Defined Networking (SDN) network in order to improve security and performance using machine learning approach.

II. MODEL DEFINITION

Developing IT models to support the intrusion detection using machine learning in SDN network, involves a comprehensive design that produces a scheme or framework for acquiring applications that work

together to implement the complete solution model or system. The overall SDN network follows a logical architectural method that simplifies and outlines components or building blocks that structure the complete information model or architecture.

The architecture is detached into application layer, control layer, and infrastructure layer. SDN architecture separates data and control layers with a precise Application Programming Interface (API) among the two [29]. The application layer is the layer that has applications and services that create requests for network functions from the Control Layer and the Infrastructure layer [2]. This layer encompasses several applications that are deployed over the controller, vital for different business essentials and requirements [27]. SDN applications interconnect with the controller by a northbound interface conferring to their network requests and these requests are traffic monitoring, network virtualization, security reinforcement, load balancing, and mobility management [14]. The northbound interface is an interface that separates the application layer and the control layer. The interface that separates the control layer and the infrastructure layer is known as southbound interface. The SDN architecture is characterised in the form of layers, as shown in Figure 1.

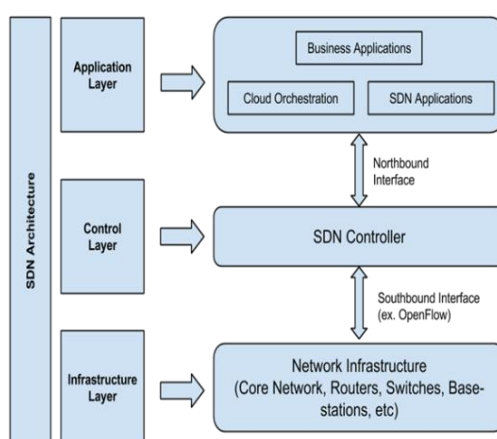


Figure 1: Overview of SDN Architecture

Infrastructure layer comprises a number of routers and switches. They are linked with each other via a wireless network or physical wired. An SDN switch is merely involved to forward the traffic data on the foundation of the controller's command. Each switch holds a flow table that comprises the records of traffic packets to make forwarding decisions [16]. The switch embraces a particular kind of high-speed memory called Ternary content-addressable memory (TCAM) and this kind of memory is used in switches in order to perform data lookup in an entire content of flow table in a single clock cycle direction [23]. Each single record in the flow table indicated with three features action, counter and rule. The rule identifies the values of the packet header [14]. Every time the switch accepts a new traffic packet, the value of the counter increases when it endorses the flow table to discover the rule. If the field values are corresponding, then, the relevant decision is engaged by the switch. Equally, if the field values do not counterpart, then the switch will notify the controller by given instruction to drop the particular incoming packet, forward or adding up fresh rules to the switches [27].

However, the SDN controller is positioned in control layer and takes application layer requests and controls the SDN data paths [19]. It implements all composite functions, including routing, naming, policy declaration, and security authorizations [7]. The control layer portrays an abstracted form of all the physical features to the application layer and it may call for several applications having diverse functionalities [21]. The SDN Controller specifies the traffic flows that happen in the SDN Infrastructure layer. Every traffic flow over the network must first get authorization from the controller, which validates that the communication is accepted by the network policy [31]. If the controller agrees to a traffic flow, it computes a path for the traffic flow to proceed and adds an entry for that traffic flow in each of the switches along the path on a flow table. Figure 2 shows a generic block diagram of an SDN connecting switch block and controller block and this scenario modelled as a queue.

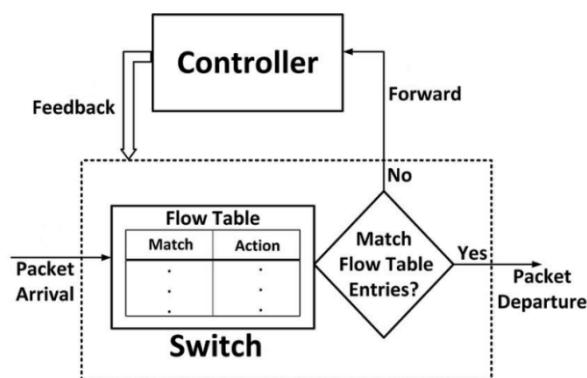


Figure 2: Generic block diagram of an SDN

The communication surrounded by the control and infrastructure layers uses a consistent protocol known as OpenFlow [10] and this protocol is standard southbound interface used in SDN implementations [27]. Hence, the protocol has turned into the standard protocol for employing software-defined networks and it's used to provide consistency in traffic management by controlling and managing how the traffic data are forwarded through the OpenFlow network among routers and switches [14].

III. SDN CHALLENGES

The separation of data and control layer also leaves the SDN network open to numerous vulnerabilities [10]. These vulnerabilities cause a threat to the extensive operation and adoption of SDN [27]. OpenFlow protocol is the most commonly used protocol in the SDN architecture and it has numerous security threats that can be misused to compromise the network [13]. Therefore, it is essential to address these security defects in OpenFlow network to safeguard the SDN development to obtain and improve network security [25].

However, [1] has reported that, SDN lacked a mechanism to detect abnormal traffic behaviour and to separate legitimate traffic from attack traffic while improving and maintaining system performance [11]. This remains as a big issue in an SDN network. Therefore, this solution model tends to improve accuracy to detect abnormal behaviour and achieve high security by proposing machine learning methods for detecting and separating intrusive attacks from normal attacks in the SDN network. At the same time, it involves monitoring the network traffic behaviour for abnormality.

The prospect of security threat is high when a single controller is deployed in an SDN network, because when the single controller fails it will increase the burden of network overhead and bad resource utilization [27] and [25]. A single centralized controller could fail under bombardment of packets [13] and [8]. Thus, using a single controller is an unreliable method to detect intrusive attacks [1]. To address this issue accordingly, this model introduces three multiple controllers to tackle new incoming packets.

According to [18] High Quality training datasets is required because reducing the large number of false alerts and increasing accuracy during the process of detecting unknown attack patterns remains unresolved problem. Unfortunately, the dataset available have deficiencies, correlated dataset and applicable to only DoS attack [20]. This system, will provide the solution to this problem by producing a qualitative and comprehensive SDN dataset with numerous type of attack and applied feature selection methods to produce redundancy-free and reduced irrelevancy dataset for anomaly detection in SDN environment using machine learning method on WEKA on environment.

IV. MODEL DEVELOPMENT

After clear model definition will result in creating a system model capable of addressing the SDN challenges identified. The main aim of this paper is to develop an enhanced and secured SDN to produce high accuracy and low false alarm rate. To effectively achieve the aim of this paper, a methodology flow diagram has been planned. Figure 3 depicts the Methodology flow diagram for developing the solution model. It indicates the procedures to follow to accomplish this development.

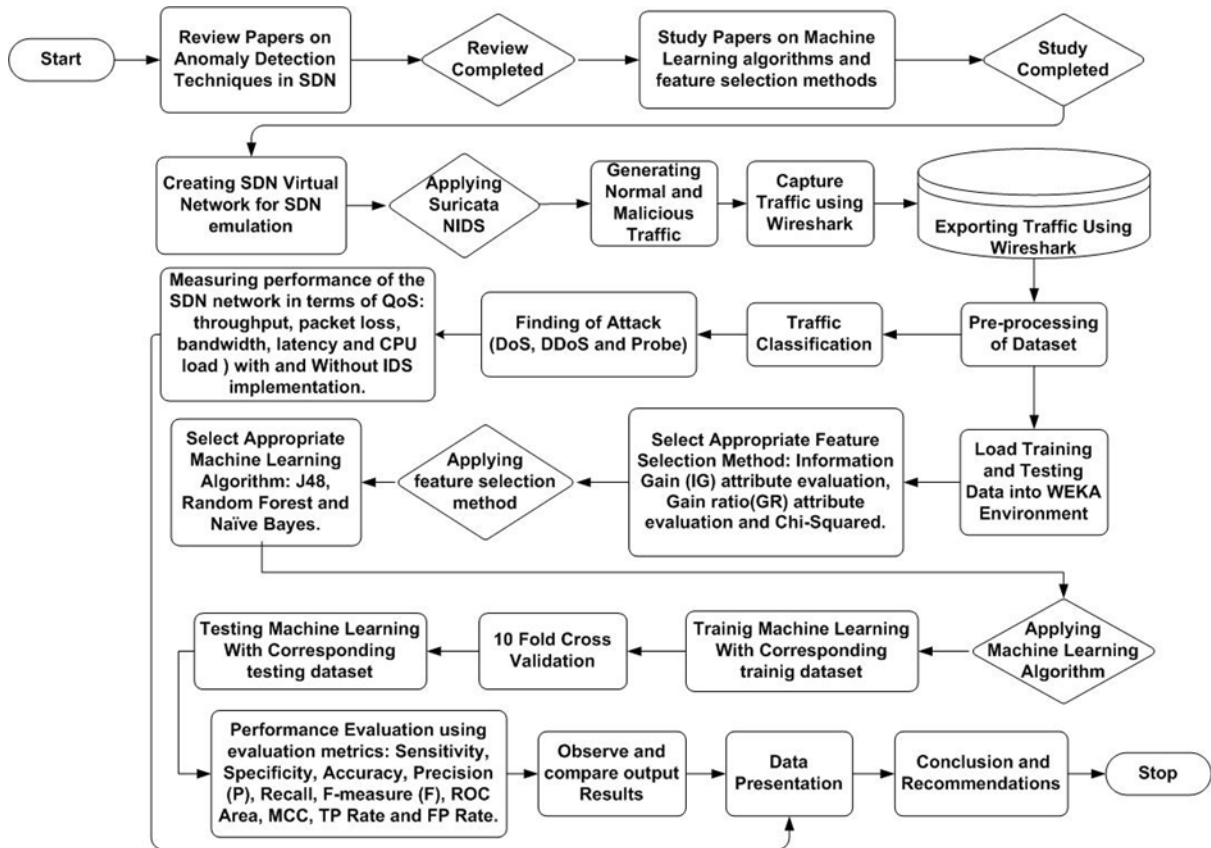


Figure 3: Methodology flow diagram for developing the solution model

The model embraces of virtual box mounted on a Linux operating system which will be used as a hypervisor and Virtual System for the simulation environment. The system consists of three core components, application layer, control layer and infrastructure layer which are involved to support the intrusion detection with machine learning. SDN Open flow protocol is used as a southbound interface between the Infrastructure layer and the control Layer of the SDN OpenFlow network [6]. The model includes an OpenFlow controller at the control layer.

The application layer is capable of governing of applications and services that create requests for network functions from the controller. At this point, this system presents a Machine Learning Based Scheme that establishes an operational intrusion detection model for classifying intrusion data. This involves applying different valuable machine learning algorithm to classify produced dataset for abnormality. The system will also apply feature selection techniques to eliminate unrelated and irrelevant data obtained from the generated dataset. The system will be trained using k-fold cross-validation technique in WEKA environment to estimate accuracy.

According to [12] agrees that, the common idea behind most machine learning is that a system studies to achieve a task by learning a training set of instances. The key indication of applying ML in flow-based anomaly detection is to form a predictive model automatically based on the training dataset. The main role of machine learning based NIDS is to automatically observe network activities and detect the existence of a cyberpunk on a network [28]. NIDS has developed to be the most important factor of the computer system security [17]. As the number of malicious attach is growing in our day-to-day life, IDS are developed with improved techniques [3]. Considering the control layer, which is the main brain of the system, the system focus on the know network based intrusion detection system Suricata on Linus operating system. According to [5] most of NIDS are false alerts generated, which must then be analysed and classified by system administrators. This increases the intrusion detection time. According to [30] mentioned that, most of the IDS System they cannot identify the source of an intrusion and in any issue of attack, they just lock the whole network. Also, indicated that Suricata IDS is one of the widely used NIDS and has advantage of producing low false alert. Suricata NIDS will be employed in this system to illustrate the effectiveness on how NIDS detect malicious attacks in SDN network. The NIDS will be capable of automatically detecting of intrusive attacks at its origin on the network and notifies the SDN controller by ensuring normal working of the SDN network.

The model also comprises three categories of users, legitimate user, attacking user and target user. These hosts are capable of sending data to each other going through OpenVSwitch. Link A, link the SDN controller and the SDN switch. Link B, link the OpenFlow switch and the Suricata IDS which help the SDN

switch to send all the scanned flows to Suricata NIDS for deep investigation. Link C, allow Suricata NIDS to notify the controller about ongoing attacks. The switch contain flow table that save all packets entries for both malicious and normal. When and NIDS indicate attack it will halt the packets and pass the details to controller through link C while the controller will notify switch to save the information of the attacker to prevent the attacker from sending packets onto the switch. Figure 4 shows the Experimental set-up of the solution model.

The system will examine four different classes of traffic flow, three attacks traffic and normal traffic as well as the tools used for generating the traffic in the proposed virtual network environment. The main idea is to generate normal traffic from the legitimate user (traffic generator) to target user, then from attacking user to generate malicious traffic to target user and evaluate the performance of SDN OpenFlow network Based on quality of service parameters. NIDS will be used to detect and eliminate these attacks. At the same time, a tcpdump tool will be used to capture traffic flows.

This system will study the benchmark of generated dataset from experimental testbed for predicting possible attacks types including DoS attacks, DDoS attacks and probe attack. Also it will apply the utmost resourceful and efficient classification models including random forest, J48 and naïve Bayes. The random forest classifier is recognized as an ensemble machine learning technique and it is act as supervised learning responsibilities. Random forest comprises of different classes of training samples with corresponding decision tree, every distinct decision tree indicates a class prediction and class containing the greatest votes turn out to be the prediction of the system. In [11]the author indicated that, it yields greater accuracy in model classification. It steps involves, picking random samples from a specified dataset. It then generates a decision tree for each test and acquires a predictive result and then holds a vote for each predicted result.

The average values of the quality of service parameters will be obtain from the following tools Wireshark, iperf and ping parametersrespectivelyonto the simulation environment. Also, the system will implement Low Orbit Ion Canon (LOIC) tool to execute out numerous DoS attacks such as HTTP flood attacks, UDP and TCP. Different DDoS attacks such as ICMP Flood attacks, UDP Flood and TCP-SYN Flood will be executed using the Hping3 tool, which considered as the well-known widely DDos tool [22]. Considering probe or probing attack, an open source Nmap tool will be used to execute the respective type of attack. In the instance of normal traffic, we consider using various internet protocols such as, SSH, HTTP, DNS and FTP, this protocols are access to generate various samples for normal traffic. Additional tools considers for the development of the system are: - (i) Mininetresponsible as SDN emulator. (ii) Floodlightresponsible as SDN controller. (iii) Suricata NIDS. (iv) OpenVSwitch (OVS). (v) Wireshark as packet tracing software.

The system will implements WEKA Framework as an environment to implement the Machine Learning experiment. According to [26] WEKA is a machine learning (ML) developed by the University of Waikato in New Zealand that also implements data mining algorithms. WEKA is a state of the art facility for developing machine learning (ML) techniques and their application to real-world problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied straight to a dataset. The OpenVSwitch is a software implementation of a virtual multilayer network switch that will be installed on OpenFlow switch machine to enable effective data packets forwarding and checking data packets before moving them to a destination. Also, the Suricata NIDS software will be installed on the OpenFlow switch machine to observe network activities.

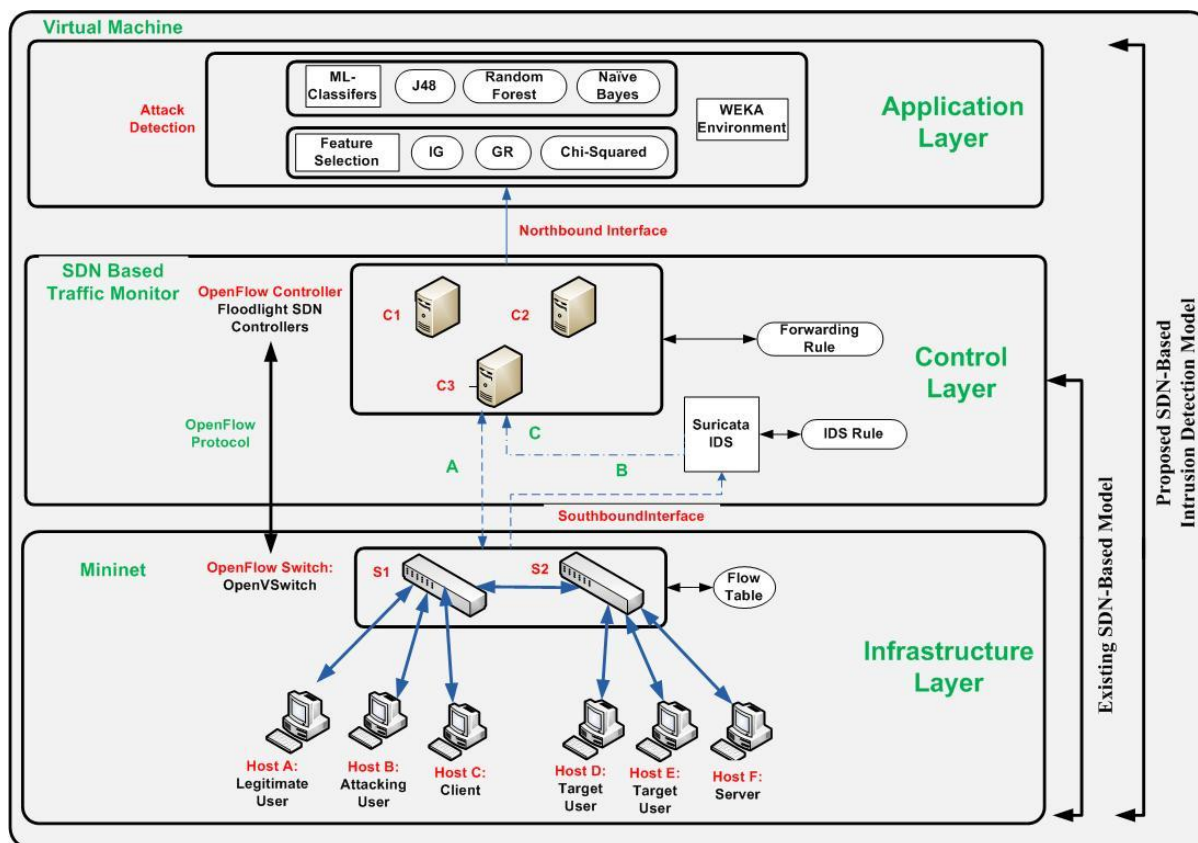


Figure 4: Overview of the solution model

4.1 Evaluation Metrics

This section illustrates and explains the performance parameters to be applied in the solution to measure the performance the model. The evaluation involves two stages, the first stage is an evaluation based on Quality of Service parameters, this involves evaluation the performance of the model experimental testbed in terms of the following parameters, throughput, packet loss, bandwidth, latency and CPU load. The second stage involves evaluating the performance of the machine learning based algorithms using confusion matrix in WEKA environment. Mostly, performance of any Machine learning techniques is evaluated in terms of accuracy, specificity and sensitivity [15]. To compute accuracy, specificity and sensitivity the following parameters need to be calculate, false positive (FP), true negative (TN), false negative (FN) and true positive (TP). An efficient machine learning model requires, low false alarm rate, high accuracy and high detection rate. A confusion matrix is applied to define these parameters.

Furthermore, the confusion matrix donate True positive as the total number of intrusion correctly recognized as intrusions [1]. False positive is the total number of ordered data wrongly recognized as intrusion. And, True negative is total number of ordered data correctly recognized as normal records. False negative is total number of intrusions wrongly recognized as ordered records.

4.2 Feature Selection

This method is responsible of removing redundant data from the dataset. According to [4] described Feature selection as an involuntary method of selection of variable data or attribute from a specified dataset that are completely relevant to the predictive modelling problem to predict output. In [11] the author emphasizes that, in the existence of redundant variable and attributes in the intrusion dataset, the dependability and consistency of anomaly detection decreases. Applying Feature selection, will contribute to an improved strategy by fast-tracking a machine learning algorithm with an enhancement in accuracy of learning [12]. Consequently, it can be an exploration or research opportunity to establish approaches for the good selection of features selection methods that can remove out unlike features effectively.

4.3 Cross validation Technique

Cross-validation will work in the system to evaluate the predictive machine learning methods by partitioning the original dataset into training and a testing dataset. It involves estimating the accuracy but not increasing the value of the accuracy [16]. The process has a solitary parameter called k that denotes to the total

number of collections that a particular dataset sample is to be divided and proceeds with every one collection individually as a test dataset [12]. In this solution model, the generated dataset would be split in to 80% training dataset and 20% testing dataset.

V. CONCLUSIONS

In conclusion, the solution model will be of great significant to any business commerce orbusiness enterprises as that are on the cyberspace.Subsequently it will support the security inspection of an intrusion attacks in the cyber systems and define the dimensions where the intrudersarecoming from in the network.This will provide an assistance to block such dimensions or areas accordingly to safeguard the SDN network to improve network security. Furthermore, the model will monitor abnormal behaviour and provide effective solution and improve accuracy for anomaly detection in SDN environment using machine learning method.Applying Machine Learningto the model can result in low False Alarm Rate and high detection rate.Another importance aspect of this model is the inclusion of the multi-controller system to increase the burden of network overhead, resource utilization and regulate bandwidth consumption.

REFERENCES

- [1]. Aladaileh, M. A., Mohammed, A.,Iznan, H. yung-weychong, H., and Yousef,K. S. (2020). Detection Techniques of Distributed Denial of Service Attacks on Software-Defined Networking Controller–A Review. DOI: 10.1109/ACCESS.2020.3013998.
- [2]. Asadollahi, S., Bhargavi, G., Ahmad, S. R. and Hedmilson, G. J. D. (2017).Scalability of Software Defined Network on Floodlight Controller using OFNet. *International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques*.557-561.
- [3]. Brian, R. G., Bob, A. and Angelos, K. M. (2015). SDN-PANDA: Software-Defined Network platform for Anomaly Detection Applications. *PhD Forum of ICNP*.
- [4]. Budugutta, S.and Nithya, S. (2017).Intrusion Detection using fuzzy logic in Software Defined Networking. *International Conference on Intelligent Computing Systems (ICICS)*.Pp.102 – 108.
- [5]. David, J. D. and Benjamin, M. B. (2011).A Performance Analysis of Snort and Suricata Network Intrusion Detection and Prevention Engines.*The Fifth International Conference on Digital Society*.187-192.
- [6]. Emil, H. S. and John D. L. (2018). Hands-on Labs and Tools for Teaching Software Defined Network (SDN) to Undergraduates. *annual conference and exposition*.
- [7]. Faris, K. and Shavan, A. (2015). Emulation of Software Defined Networks Using Mininet in Different Simulation Environments. *IEEE*. 10.1109/ISMS.2015.46.
- [8]. Husham, B. A. A. (2017). Detection of DDoS Attacks against the SDN Controller using Statistical Approaches.(Master’s Theses, Wright State University, 2017).*Theses and dissertations CORE Scholar*.
- [9]. Jayamagarajothi, M. and Murugeswari, P. (2015).A Survey on Data Mining and Digital Forensics Techniques for Intrusion Detection and Protection System.*International Journal of Advanced Research in Computer and Communication Engineering*. 4, 1, 159-164.
- [10]. Kumar, G. (2014). Evaluation Metrics for Intrusion Detection Systems - A Study.*International Journal of Computer Science and Mobile Applications*. 2, 11, 11-17.
- [11]. Kumar, S. D. and Mahbubur, M. R. (2019).Effects of Machine Learning Approach in Flow-Based Anomaly Detection on Software-Defined Networking.MDPI: *Symmetry*. 12, 7.1-21.
- [12]. Kumar, S. D., Raihan, U. and Mahbubur, R. (2020).Performance Analysis of SDN-Based Intrusion Detection Model with Feature Selection Approach.*International Joint Conference on Computational Intelligence, Algorithms for Intelligent*.pp.483-494.
- [13]. Kumar, S., Tarun, K., Ganesh, S., Maninder, S. N. (2012). Open Flow Switch with Intrusion Detection System. *International Journal of Scientific Research Engineering & Technology (IJSRET)*. 1, 7, 001-004.
- [14]. Kunal, S., Gandhi, P., Sutariya, R. and Tarpara, H.(2020). A secure software defined networking for distributed environment. Security and Privacy.<https://doi.org/10.1002/spy2.130>.
- [15]. Li, W., Yu, W., Zhiping, J., Keping, Y., Jin, L. and Yang, X. (2020).Challenge-based collaborative intrusion detection in software-definednetworking: *An evaluation*. *Digital Communications and Networks*. <https://doi.org/10.1016/j.dcan.2020.09.003>.
- [16]. Muthamil, K. S. and Deepalakshmi, P. (2020).A two level security mechanism to detect a DDoS flooding attack in software-defined networking using entropy-based and C4.5 technique.*Journal of High Speed Networks*. No. 26, pp 55–76. DOI 10.3233/JHS-200630.
- [17]. Nalavade, K.C. and Meshram, B.B. (2011).Comparative Study of IDS and IPS.*BIOINFO Computer Engineering*. 1, 1, 01-04.
- [18]. Omar, S., Asri, N.and Hamid, H. J.(2013). Machine Learning Techniques for Anomaly Detection: An Overview. *InternationalJournal of Computer Applications* (0975 – 8887). 79, 2, 33-35.
- [19]. OpenAirInterface (2021). *Cloud RAN (C-RAN)*. <https://openairinterface.org/use-cases/cloud-ran-c-ran>.
- [20]. Ramkumar, M. P., Emil, S. and Bavani, K. (2020).Statistical Approach Based Detection of Distributed Denial of Service Attack in a Software Defined.*International Conference on Advanced Computing & Communication Systems (ICACCS)*.No. 6.pp. 380 -385.
- [21]. Rao, S. P. and Shakthipriya, P. (2015). Construction of a Simplified Software Defined Networking (SDN) Test-Bed. *International Journal of Applied Engineering Research*.10, 18, 39030-39033.
- [22]. Said, M. E., Nhien-An, L. and Anca, J. (2020). InSDN: A Novel SDN Intrusion Dataset. DOI10.1109/ACCESS.2020.3022633, IEEE Acces
- [23]. Scott-Hayward, S., Natarajan, S., and Sezer, S. (2016). A Survey of Security in Software Defined Networks. *IEEE Communications Surveys and Tutorials*, 18(1), 623-654. <https://doi.org/10.1109/COMST.2015.2453114>
- [24]. Sonal, P., Rajesh, S. S. and Mandoria, H.L. (2015).Analytical Study on Intrusion Detection and Prevention System.*International Journal of Emerging Trends and Technology in Computer Science (IJETTCS)*.4.6.177-182.
- [25]. Suresh K.,Tarun K.,Ganesh S.and Maninder S. N.(2012). Open Flow Switch with Intrusion Detection System.*International Journal of Scientific Research Engineering & Technology (IJSRET)*, 1, 7. pp 001-004.
- [26]. Svetlana S. A. (2004). Machine Learning with WEKA.*School of Engineering and Computer Science Department of Computer Science California State University, Sacramento California*, 95819.
- [27]. Swami,R., Mayank,D. and Virender,R.(2019).Software-defined Networking-based DDoSDefense Mechanisms. *ACM Computing Surveys*.Vol.52, No. 2, Article 28, pp. 28 -36.

- [28]. Vinutha, H.P. and Poornima, B. (2015).A Survey - Comparative Study on Intrusion Detection System, *International Journal of Advanced Research in Computer and Communication Engineering* 4, 7.410-414.
- [29]. William, S. (2013). Software-Defined Networks and OpenFlow.*The internet protocol journal*.16, 1, 1-6.
- [30]. Xing, T., Huang, D., Xu, L., Chung, C.J., and Khatkar, P. (2013).“Snortflow: Aopenflow-based intrusion prevention system in cloud environment,” Research and Educational Experiment Workshop (GREE), 2013 *Second GENI. IEEE*, pp. 89–92.
- [31]. Zhu, L., M. Karim, Kashif, S., Fan, L., Xiaojiang, D. et al. (2019). SDN Controllers: Benchmarking & Performance Evaluation. *IEEE Journal on Selected Areas in Communication*.