# Detection of Flow Based Anomaly in OpenFlow Controller: Machine Learning Approach in Software Defined Networking

Samrat Kumar Dey[1], Md. Mahbubur Rahman[2] and Md. Raihan Uddin[3]
[1,2]Department of CSE
[3]Department of EEE
[1,2]Military Institute of Science and Technology, Dhaka, Bangladesh
[3]Daffodil International University Dhaka, Bangladesh
sopnil.samrat@gmail.com[1], mahbubcse@yahoo.com[2] and mkraihan13@gmail.com[3]

*Abstract*— **Software Defined Networking (SDN) has come to prominence in recent years and demonstrates an enormous potential in shaping the future of networking by separating control plane from data plane. OpenFlow is the first and most widely used protocol that makes this separation possible in the first place. As a newly emerged technology, SDN has its inherent security threats that can be eliminated or at least mitigated by securing the OpenFlow controller that manages flow control in SDN. SDN provides us a chance to strengthen our network security by decoupling its control plane and data plane. At this level, there also exists some certain risk, which is associated with this technology. In this research, a flow based anomaly detection method in OpenFlow controller have been approached by using machine-learning algorithms in SDN architecture. In order to improve the classifier performance, some feature selection methods namely Info Gain, Gain Ratio, CFS Subset Evaluator, Symmetric Uncertainty, and Chi-square test have been applied as a processing of dataset. A full dataset of 41 features and a reduced dataset has experimented. A dataset with feature selection ensures the highest accuracy of nearly 82% with Random Forest classifier using Gain Ratio feature selection Evaluator. Experimental results ratify that machine-learning approach with feature selection method indices strong potential for the detection of flow based anomaly in OpenFlow controller.**

***Keywords—software defined networking; OpenFlow; machine learning; NSL-KDD dataset; feature selection; intrusion detection;***

## I. INTRODUCTION

Concept of Software Defined Networking (SDN) is considered as a present innovation in computer networking. SDN provides the permission to a network to customize its behavior through a conceptually centralized controller. From over the previous few decades, traditional network architecture has persisted unchanged and considered cumbersome as SDN represent a significant departure in comparing with traditional structure. SDN has recently appeared as a innovative powerful paradigm which is vibrant, convenient, flexible and dynamic in nature. Currently researchers are being drawn by the combination of OpenFlow and SDN from both industry and academic. OpenFlow is used as flow counter to identify the records of network traffic. Google B4 [1] and Carrier network of Huawei [2] have already proven the advantages of SDN by developing a backbone network. Some SDN Controller like NOX [3], Ryu [4] and Floodlight [5] have been widely recommended by some open source organizations. However, SDN is proficient of traffic analysis with logical centralized controller support, but the dissolution of Data and control planes cause new attacks opportunities. Detection of Intrusion in network shields a network from different malicious software attacks. Conventionally, Detection of network attacks has followed two types of approaches namely; Signature based detection and anomaly-based detection. Automatic Network Intrusion Detection System (NIDS) has been a significant investigation point for the last decades and so. Still researchers are developing NIDS with the competency of identifying attacks in numerous existing environments. Machine Learning approaches are latest on the scene. Machine learning are considered as a set of technique, which consists of evolving algorithm that is based on knowledge based learning. As SDN has a nature of flow-based traffic analysis, therefore this paper propose a flow based anomaly detection system using Machine-learning approach. In this assay, we have analyzed the standard NSL-KDD dataset for predicting possible intrusion by using some Machine Learning algorithms. Most effective classification methods namely J48, Random Forest, Projective Adaptive Resonance Theory (PART), Naive-Bayes, Decision Table (DT), Radial Basis Function Network (RBFN) and Bayesian Network have been applied. Finally, for the performance improvement of the classifier, utmost prevalent feature selection methods are applied as data preprocessing.

## II. CONTRIBUTION

According to state of the art research in the field of intrusion detection in context to SDN, we believe that Machine learning (ML) based proper feature selection with a concrete detection model can effectively deliver a better solution for the SDN. A Machine learning based security model is proposed for anomaly detection. In succinct, the prime contributions of this paper are the following:

- According to our knowledge, our proposed model is the first designed architecture which ensures to detect flow based malicious class in OpenFlow Based SDN controller.
- From our experiment, Random forest classifier propagates a detection rate of nearly 82% using an appropriate feature selection mechanism compared to other state-of-the-art approaches. To best of our knowledge, previously no work has been conducted by using such diverse feature selection approach.

- We also compare and evaluate the performance of our RF-Gain Ratio Model with others anomaly detection model and shows that with proper selection of features our approach has significant potential for real time detection.

We have systematized the residuum of this research as follows. In Section III, we discuss some relevant work which was performed previously. Section IV presents the insight description of our Methodology. Experimental procedure and discussion of result analysis are presented in Section V. Utterly, we conclude the paper by presenting future work in Section VI.

## III. Background Study

Nowadays detection of Flow-based anomaly is researched comprehensively. This section presents various review of recent accomplishment related to intrusion detection using Machine learning approach. A flow-based anomaly detection system combining of Gravitational Search Algorithm (GSA) and Multi-Layer Perceptron (MLP) is proposed in [7] where system is capable of classifying malicious flows with a high detection rate. For anomaly detection authors in [8] discussed some machine learning technique like neural networks, Bayesian networks, Fuzzy logic, genetic algorithm and support vector machine that could be used to handle intrusion detection attacks in software-defined networking. Authors in Reference [9] present a simple survey on programmable networks with prominence on SDNs, which presents a discussion on the advancement of programmable networks with the highlights of SDN architecture. A support vector machine of one-class has been proposed for the exploration in [10] which produces a very low false alarm rate while the proposed systems is competent with malicious network dataset rather than a magnanimous dataset. Another detailed survey on SDN has been carried out by [11] which includes a number of various challenges and solutions that have been proposed in the literature to account for network threats. In SDN Environment, different kinds of anomaly detection algorithms are implemented to ensure the security of the OpenFlow network. By using some machine learning algorithms, which trained on historical network attack data to identify the potential malicious connections and potential attack destination have been proposed to predict network attack in SDN by [12]. They have used four widely-known machine learning algorithms: C4.5, Bayesian Network, Decision Table, and Naïve-Bayes. An innovative framework called Atlas presented by authors in Reference [13], which is capable of leverages application awareness and efficient for L2/3/4-based policy enforcement in SDN. It uses a machine learning approach, C5.0 classifier, by collecting data using crowd-sourcing approach to classify the traffic in SDN.Very recent work on flow based anomaly detection in SDN based on Deep Neural Network has been established in [6] where a subset of six features have been used of NSL-KDD dataset. However, their results are not good enough in comparing with available Machine Learning Algorithms.

## IV. Methodology

### A. Proposed ML Based Model of Classification

In this portion, we introduce our suggested model for classifying anomaly class in order to establish an effective intrusion detection scheme, which can provide low false alarm rate and high detection rate. Figure. 1 depicted two layered based hybrid classification model. Immaterial and terminated features are taken out from dataset using some common feature choice methods and then in next layer, the abridged data set is classified using some useful Machine learning algorithm. Further, assessment of the model carried out by using some accurate measures.

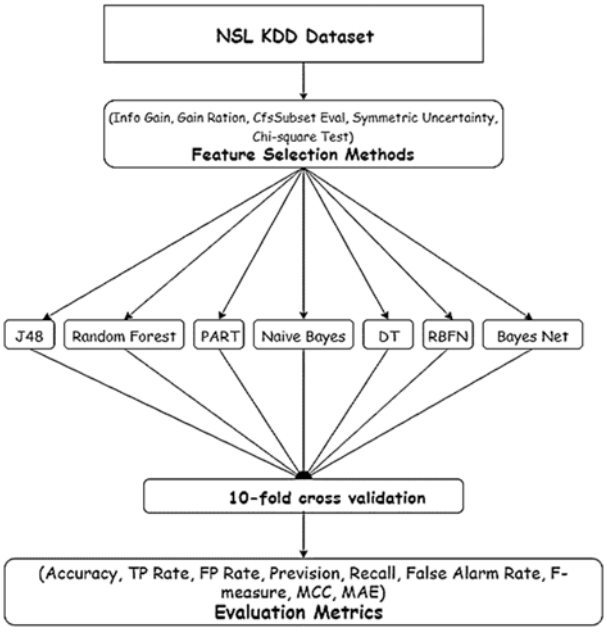


Fig. 1 Machine Learning based Classification Model

### B. Machine Learning Approach

According to [14] the common idea behind most machine learning is that a system learns to accomplish a task by studying a training set of examples. The system of distributed computers and controllers then performs the similar task with data that has not encountered before. Main idea of using ML in flow based anomaly detection to build a predictive model automatically based on the training dataset. Machine learning can be enumerated as the study of algorithms that advance their enactment with experience and the machine takes every required steps accomplished likewise in a mentioned way. Authors in Reference [15] discussed regarding various machine learning algorithms that have been widely used for a number of classification and prediction problems. Below Fig. 2. Showing complete flow chart of anomaly detection mechanism in OpenFlow controller.

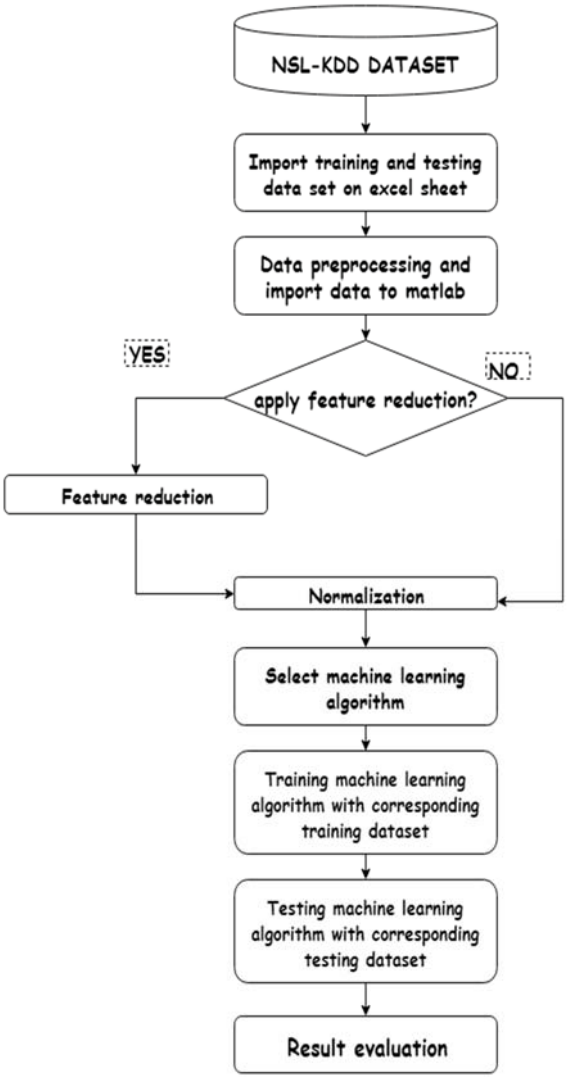


Fig. 2 Flowchart of anomaly detection using Machine Learning

### C. Dataset

It is not an easy task of picking out an accurate data set for model evaluation and establishment. Some of the significant features like extensive, free of noise, consistency and redundancy that we look for. For anomaly detection various dataset have been used, some of the dataset are self-made and some are publicly accessible. Among several publicly accessible datasets, KDD-99 is considered the most extensively accepted standard dataset. According to Reference [16] KDD-99 contains nearly drawbacks as the whole dataset is huge enough that upturns the cost of computation of the Intrusion Detection. Hence, Authors in [17] shows that only 10% of the set is generally used. Again, in training set it contains many redundant data whereas in testing set it shows some identical records. As a result, it might distress the learning process of the systems. A clarified version of KDD-99 is NSL-KDD which is capable of minimizing the redundancies between testing and training. Dataset of NSL-KDD has been suggested by Tavallace et al. [16] that consist of intrusion data which being used by many researchers for experimentation. NSL-KDD contains 41 features that consist both normal and attack patterns where dataset-having 41 features and 5 classes that are normal and 4 different types of attacks: Denial of Service (DoS), User to Root (U2R), Probe and Remote to Local attack (R2L). Following TABLE I shows the features and position of attribute in NSL-KDD Dataset.

TABLE I. FEATURES AND ATTRIBUTE POSITION OF NSL-KDD DATASET

| Type | Features | Attributes position |
|---|---|---|
| Nominal | Protocol_type, Service and Flag | 2,3,4 |
| Numeric | Duration, src_bytes, st_bytes, wrong_fragment, urgent, hot, num_failed_logins, num_compromised, num_root, num_file_creations, num_shells, num_access_files, num_outbound_cmds, count srv_count, serror_rate, srv_serror_rate, rerror_rate, srv_rerror_rate, same_srv_rate diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate and dst_host_srv_rerror_rate | 1,5,6,8,9,10,11, 13,16,17,18,19, 20,23,24,25,26,27, 28,29,30,31,31 32,33,34,35 36,37,38,39 40,41 |
| Binary | Land, logged_in, root_shell, su_attempted, is_host_login and is_guest_login | 7,12,14,15,21,22 |

### D. Feature Selection

The existence of extraneous attributes in the intrusion dataset often deterred the detection of accuracy. For that reason, there exist a research challenge of suitable feature selection methods, which can eliminate particular attributes that may not contribute to the intrusion detection process. According to Reference [18], feature selection is the technique of eliminating features from the dataset that are not relevant to the task that is to be accomplished. By reducing the number of irrelevant, redundant, and noisy features feature selection methods speeds up a data-mining algorithm with an improvement in learning accuracy which leading to comprehend a better model [19]. In this exploration, the dataset features are diminished by using Info Gain, Gain Ratio, CFS Subset Evaluation, Symmetric Uncertainty and Chi-Square Test. TABLE II shows the attribute selection procedures with different evaluator with their search strategies. For the experiment, the following selected attribute has been used for accuracy calculation of our proposed model.

TABLE II. FEATURES SELECTION WITH DIFFERENT EVALUATOR AND SEARCH METHOD

| Evaluator | Search | Selected Attributes |
|---|---|---|
| Info Gain | Ranker | 5,6,3,4,33,35,34,40,41,23,30,29,12,27,28: 15 |
| Gain Ratio | Ranker | 28,12,41,27,4,6,5,30,29,40,3,25,26,39,34: 15 |

| | | |
|---|---|---|
| CFS Subset Evaluator | Best First | 5,6,12,25,28,30,31,37,41 : 9 |
| Symmetric Uncertainty | Ranker | 6,5,4,41,28,12,27,30,3,40,29,34,35,33,37: 15 |
| Chi-Square Test | Ranker | 5,6,3,33,35,34,4,40,23,12,41,30,29,27,37: 15 |

### *E. Evaluation Metrics*

Performance of intrusion detection rate is estimated in terms of accuracy (AC), precision (P), recall(R), F-measure (F), False Alarm Rate (Far) and Mathews correlation coefficient (MCC). Some performance metrics derived from confusion matrix are taken into consideration for calculating the value of mentioned techniques. In general, confusion matrix visualize the performance of the algorithm according to following Fig. 3. True positive is the total number of samples predicted as normal while they were actually normal meanwhile false negative predict total sample as normal while they were actually attack. Moreover, false positive is opposite to True Positive as it predicts total sample as attack where they were truly normal and True negative predict total number of sample as attack while they were actually attack.

| | Predicted as Normal | Predicted as Attack |
|---|---|---|
| Normal Class (Actually) | True Positive (TP) | False Positive (FP) |
| Attack Class (Actually) | False Negative (FN) | True Negative (TN) |

Fig. 3 Tabular form of a confusion matrix

A good intrusion detection scheme entails high rate of accuracy and high detection rate with a very low false alarm rate. The relation between false alarm rates with misclassification rate that they are directly proportional to each other. The following metrics are used for evaluating a model. A brief discussion and calculating formula are showing below for each metric.

**Accuracy (AC):** shows the proportion of the classification from overall N Examples that were correct.

$$AC = \frac{TP + TN}{TP + FP + TN + FN}$$

**Precision (P):** shows the proportion of intrusion anticipated by a Network intrusion detection systems are a real intrusion. As the value of P is higher, then the probability of lower false alarm rate is:

$$P = \frac{TP}{TP + FP}$$

**Recall (R):** shows the proportion of positive examples that were correctly classified. We are in search of a high value of R

$$R = \frac{TP}{TP + FN}$$

**F-measure (F):** By conveying balance between accuracy and recall, it gives a better measure of accuracy. We are in quest of a high F-measure value.

$$F = \frac{2}{\frac{1}{P}+\frac{1}{R}}$$

**Mathew's correlation coefficient (MCC):** It returns a binary value between -1 and 1 by showing the value correlation coefficient between the predicted and observed binary classifications.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN}}$$

### *F. SDN-based flow-based anomaly detection architecture*

SDN controller is capable of monitoring all the OpenFlow switches and send the request to all network data whenever it is necessary. Therefore, our proposed and designed intrusion detection segment is implemented in SDN controller, which is depicted in Fig. 5. Following Algorithm 1 encapsulates our suggested approach.

**Algorithm 1** Machine Learning based anomaly class detector for SDN attacks

1: **procedure** FLOW BASED ANOMALY DETECTOR BASED ON ML MODEL
2: Selection of appropriate machine learning algorithm
3: Train the ML-based model using benchmark dataset NSL-KDD
4: Nomination of appropriate feature after using different feature selection methods
5: **if** The trained model predicts an anomaly class on a OpenFlow Controller by the ML based Intrusion Detection Model
**then**
6: Update the SDN OpenFlow controller rules to block that class attack type
7: **else** Allow the normal class to pass through SDN controller and access the available resources.

For requesting network data, on OpenFlow stats request message will be sent from the controller to all OpenFlow switches. As controller request for all the available statistics, an OpenFlow stats reply message with all available data send back to the controller by OpenFlow switch. Following Fig. 4. Describes the architecture of how OpenFlow switch handles the incoming packet and responds according to the availability of data in Flow table by using OF protocol.

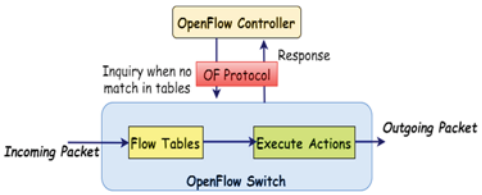


Fig. 4 Diagram of Handling Incoming Packets in OpenFlow Switch

The centralized controller of SDN can take Opportunities of the complete network to evaluate and associate feedback from the network. For analyzing and detecting any real time network intrusion will then be sent to Intrusion detection segment according to Fig. 4. OF protocol can effectually alleviate an intrusion via flow table adjustment if once a network anomaly is discovered and recognized.

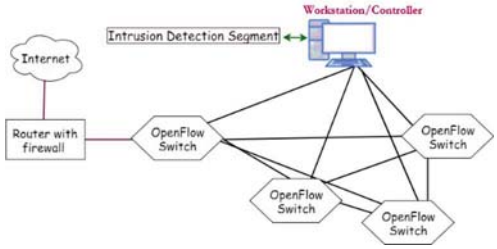


Fig. 5 Proposed flow based anomaly detection architecture in SDN

## V. Experimental Result and Analysis

Experiments are carried out in WEKA [20] environment using NSL-KDD Dataset on an objective machine having a memory of 6 GB and a Processor Intel(R) Core(TM) i5-2410M CPU @ 2.30 GHz, 2301 MHz, Dual Core(s), and 4 Logical Processor(s). The NSL-KDD dataset is used for training and testing each elected features. For performing experiment 10 fold cross validation technique was used. In cross-validation, the training set is distributed into 10 subsets and apiece subset is tested when the model is trained on the other 9 subsets. Every subset is used as a test data only once and the process repeats 10 times. Each fold is combined to produce a distinct result obtained from testing dataset. For the simplicity of our research, we have only shown the results of highest accuracy achieved by different classifier in terms of feature selection in following TABLE III. As we aim to have high Accuracy, Recall and MCC and while low in False alarm rate, we have successfully achieved that in our experiment. Among the experimental data, Random Forest shows the highest accuracy with Gain Ratio feature selection method which depicted in Fig. 6 with an accuracy of 81.946 %. Initially, we implemented the network intrusion detection for 2-class based classification namely normal and anomaly. In addition, we assessed our work by assimilate the results and proposed a model security architecture for detecting flow based anomaly in OpenFlow based Controller. From the above discussion and also from Fig. 6 its clearly shows that with the feature selection methods of Gain Ratio, Random Forest classifier provides highest Accuracy of 81.95% with a very low false alarm rate of 0.287%. Moreover, our results is generated based on the selection of features from complete dataset of NSL-KDD. Very few approaches has been presented from different authors in order to show the accuracy of machine learning algorithm for NSL-KDD Dataset. However, there also exists some lacking of preprocessing of dataset and appropriate feature selection for testing and training. For comparing with others work we will only consider that approaches which is only based on NSL-KDD dataset for the detection of flow based anomaly in Software Defined Networking. Authors in Reference [19] shows some approaches for performance analysis of NSL-KDD dataset using Artificial Neural Network. It shows an accuracy of 81.2% for intrusion detection but suitable way of feature selection is not described clearly. Moreover, no architecture is provided for intrusion detection in SDN. Another work in reference [21] present a technique of intrusion detection in SDN with an accuracy of 97% using pattern recognition of neural network but the dataset contains only 7 features for testing. For handling intrusion and DDoS attack in SDN, ML-based techniques is presented in Reference [22]. It discussed thoroughly some ML approaches but no performance evaluation using dataset is carried out. An experiment is carried out in Reference [23] using some classification algorithm like J48, SVM, and Naïve Bayes showing a high percentage of accuracy but dataset contains only 6 features for testing. Another approach of Deep learning based intrusion detection in SDN is presented in Reference [6] where their proposed DNN Model capable of achieving 75.75% accuracy. However, the procedure of feature selection methods is not evidently discussed whereas they have used only 6 features for experiment design. Moreover, there exist some scatter research work on the analysis of NSL-KDD dataset for intrusion detection but no fitting method of feature selection for training and testing has been provided. From all the above, our demonstrated proposed model for detection of an anomaly in SDN architecture with a fitting feature selection procedure can be generalized and provides us an auspicious accuracy.

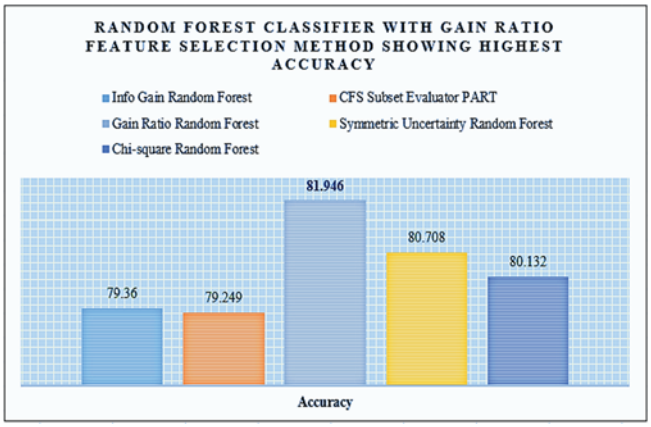


Fig. 6 Random forest shows highest accuracy with Gain Ratio feature selection method

TABLE III. Random Forest Classifier Showing The Highest Accuracy With Gain Ratio Feature Selection Methods

| Feature Selection Method | Classifier Techniques | Evaluation Criteria | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **Accuracy** | **TP Rate** | **FP Rate** | **Precision** | **Recall** | **FAR** | **F-Measure** | **MCC** | **MAE** |
| Info Gain | Random Forest | 79.360 | 0.794 | 0.163 | 0.846 | 0.794 | 0.341 | 0.792 | 0.641 | 0.229 |
| CFS Subset Evaluator | PART | 79.249 | 0.792 | 0.167 | 0.839 | 0.792 | 0.333 | 0.791 | 0.633 | 0.264 |

| **Gain Ratio** | **Random Forest** | **81.946** | **0.819** | **0.143** | **0.860** | **0.819** | **0.297** | **0.819** | **0.681** | **0.232** |
|---|---|---|---|---|---|---|---|---|---|---|
| Symmetric Uncertainty | Random Forest | 80.708 | 0.807 | 0.153 | 0.853 | 0.807 | 0.317 | 0.806 | 0.661 | 0.221 |
| Chi-square | Random Forest | 80.132 | 0.801 | 0.157 | 0.850 | 0.801 | 0.328 | 0.800 | 0.653 | 0.222 |

## VI. CONCLUSION

In this paper, we have presented a machine learning based classification model for detecting network intrusion and shows the best classifier in terms of different evaluation metrics with different feature selection mechanism. Five feature selection methods with different machine learning algorithm have been analyzed with the support of NSL-KDD dataset. Although our experimental results are not yet high enough comparing with others but it still has significant contribution in the field of appropriate feature selection from a dataset. In SDN environment, machine learning approach has enormous potential to detect malicious activity. SDN supports the nature of centralized controller and a very flexible structure. Our proposed intrusion detection module is capable of easily extract the information about network traffic due to its centralized controller and flexible nature. From the experiment, it's clear that RF shows a high test accuracy comparing with all other algorithms therefore for flow based anomaly detection use of RF is very essential in order to achieve high accuracy and speeding up the process of intrusion detection in SDN. To improve the accuracy we will analyze the traffic flow and propose other kinds of feature selection Evaluator. In near future, we plan to implement this proposed model in a real Environment of SDN with real traffic of Network.

## REFERENCES

[1] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu et al., "B4: Experience with a globally-deployed software defined wan," ACM SIGCOMM Computer Communication Review, vol. 43, no. 4, pp. 3–14, 2013.

[2] C. T. Huawei Press Centre and H. unveil world's first commercial deployment of SDN in carrier networks, "[online]. available: pr.huawei.com/en/news/hw-332209-sdn.htm."

[3] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, "Nox: towards an operating system for networks," ACM SIGCOMM Computer Communication Review, vol. 38, no. 3, pp. 105–110, 2008.

[4] "Ryu," Available: http://osrg.github.io/ryu/.

[5] "Floodlight," Available: http://www.projectfloodlight.org/.

[6] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi and M. Ghogho, "Deep learning approach for Network Intrusion Detection in Software Defined Networking," 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM), Fez, 2016, pp. 258-263.doi: 10.1109/WINCOM.2016.7777224Y.

[7] Z. Jadidi, V. Muthukkumarasamy, E. Sithirasenan, and M. Sheikhan, "Flow-based anomaly detection using neural network optimized with gsa algorithm," in 2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops, 2013, pp. 76–81.

[8] J. Ashraf and S. Latif, "Handling intrusion and DDoS attacks in Software Defined Networks using machine learning techniques," in Software Engineering Conference (NSEC), 2014 National, pp. 55–60, Nov. 2014.

[9] B. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," Communications Surveys & Tutorials, IEEE, vol. 16, no. 3, pp. 1617–1634, 2014.

[10] P. Winter, E. Hermann, and M. Zeilinger, "Inductive intrusion detection in flow-based network data using one-class support vector machines," in New Technologies, Mobility and Security (NTMS), 2011 4th IFIP International Conference on. IEEE, 2011, pp. 1–5.

[11] S. T. Ali, V. Sivaraman, A. Radford, and S. Jha, "A survey of securing networks using software defined networking," Reliability, IEEE Transactions on, vol. 64, no. 3, pp. 1086–1097, 2015.

[12] S. Nanda, F. Zafari, C. DeCusatis, E. Wedaa and B. Yang, "Predicting network attack patterns in SDN using machine learning approach, " 2016 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Palo Alto, CA, 2016, pp. 167-172.doi: 10.1109/NFV-SDN.2016.7919493.

[13] Z. A. Qazi, J. Lee, T. Jin, G. Bellala, M. Arndt, and G. Noubir, "Application-awareness in SDN," in ACM SIGCOMM Computer Communication Review, vol. 43, pp. 487–488, ACM, 2013.

[14] P. Louridas and C. Ebert, "Machine Learning," in IEEE Software, vol. 33, no. 5, pp. 110-115, Sept.-Oct. 2016.

[15] G. M. Khan, S. Khan, and F. Ullah, "Short-term daily peak load forecasting using fast learning neural network," in Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on, pp. 843–848, IEEE, 2011.

[16] M. Tavallaee, E. Bagheri, W. Lu, and A.-A. Ghorbani, "A Detailed analysis of the kdd cup 99 data set," in Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications 2009, 2009.

[17] Y.-X. Meng, "The practice on using machine learning for network anomaly intrusion detection," in International Conference on Machine Learning and Cybernetics (ICMLC), vol. 2, pp. 576–581, IEEE, 2011.

[18] Y. Yang, and Jan O. Pedersen, "A comparative study on feature selection in text categorization," in Proceedings of the ICML '97 Fourteenth International Conference on Machine Learning, 1997, pp. 412-420.

[19] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," 2015 International Conference on Signal Processing and Communication Engineering Systems, Guntur, 2015, pp. 92-96.doi: 10.1109/SPACES.2015.7058223.

[20] E. Frank, M. A. Hall, and I. H. Witten, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[21] A. Abubakar and B. Pranggono, "Machine learning based intrusion detection system for software defined networks," 2017 Seventh International Conference on Emerging Security Technologies (EST), Canterbury, 2017, pp. 138-143.

[22] J. Ashraf and S. Latif, " Handling intrusion and DDoS attacks in Software Defined Networks using machine learning techniques," 2014 National Software Engineering Conference, Rawalpindi, 2014, pp. 55-60.

[23] L. Dhanabal and P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms," International Journal of Advanced Research in Computer and Communication Engineering, pp. 446-452, 2015.