Machine Learning-based Optimization and Anomaly Detection in Software Defined Networks

Aastik Sharma
Department of Computer Science and Technology SRMIST, Kattankulathur

ABSTRACT
Software Defined Networks (SDN) enhance network programmability by separating the control and data planes, yet challenges remain in performance, traffic optimization, and security. This paper evaluates the integration of machine learning (ML) techniques, including K-Nearest Neighbor, Decision Tree, Support Vector Machine, Bayesian models, and Deep Neural Networks, to improve SDN performance. Experiments across multiple scenarios demonstrate that ML algorithms can enhance traffic prediction, detect anomalies, and mitigate DDoS attacks, achieving up to 100% accuracy in specific configurations. The study highlights the potential of ML to significantly improve SDN efficiency, security, and scalability.

Keywords
Software Defined Network (SDN), Machine Learning (ML), Outliers, Supervised Learning, Unsupervised Learning, Reinforcement Learning, Distributed Denial of Service (DDoS), Anomalies Detection, Traffic Engineering, Elephant Flow, Convolutional Neural Network (CNN), Deep Learning (DL), Support Vector Machine (SVM), Decision Tree, Naïve Bayes, Port Scanning Attacks

1. INTRODUCTION
In simple words, Software defined networks or SDN is a way to make network programmable. In the traditional system the control and data plane were coupled together because of which it was difficult to make any changes which in turn also affected the performance. But this difficulty has been overcome by software defined network (see Figure 1) as the control plane and the data plane are not coupled together. Rather they are separated and it is converted into a centralized network. This provides the advantage of network virtualization that is it can create many virtual networks on physical network infrastructure. Software defined network has been very advantageous as it has reduced the costs and increase the security etc. [2]. Because of various advantages over the traditional system, software defined network is being used in varies technologies like cloud computing.

Machine learning is a subset of artificial Intelligence that is used to predict the future outcomes by analyzing patterns or datasets from past knowledge. Machine learning has the power to evaluate large datasets without any problem and give accurate prediction. Currently there are many different machine learning algorithms available. This paper is going to combine few of the machine learning algorithm with software defined networks and observe if it helps in improving the network.

In the upcoming sections it will discuss regarding the ways in which machine learning can be combined with software defined network and the challenges that would be faced in integrating machine learning algorithms in software defined network along with the future scope of machine learning in networks.

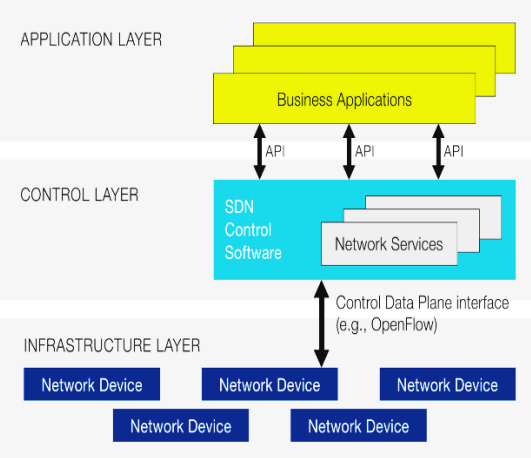

Figure 1: Software Defined Networks(SDN) Architecture

2. BACKGROUND
2.1 Software defined network
In simple words, Software Defined Network (SDN) is a method to make the network programmable. SDN has a centralized software controller that allows network administrators to control the behavior of the network. This technique not only improves the flexibility and scalability but also improves automation and programmability. Software defined network overcomes the drawbacks of the traditional network by separating the data plane and control plane.

In SDN architecture application layer interacts with control layer using northbound APIs. The communication between infrastructure layer and control plane takes place using OpenFlow protocol, and via southbound APIs. New applications and software's can be developed using SDN that are more fast, dynamic, with better traffic management and provide improved efficiency, security, and performance.

2.2 Machine learning
Machine learning is a part of artificial Intelligence that deals with predicting future outcomes by analyzing patterns or datasets from past knowledge. The paper has performed and evaluated various machine learning algorithms. Some of supervised algorithms are K-nearest neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Logistics regression (RL), and Decision tree (DT).

Supervised learning [3] and semi-supervised learning are used for applications such as congestion detection, traffic network prediction, elephant flow detection. Unsupervised learning and Reinforcement Learning are found useful for fault management, flow feature-based traffic classification, and packet loss classification.

3. LITERATURE SURVEY

3.1 A survey conducted on use of Machine Learning in Software-Defined Network

In the research paper of Hamed Mirzaei et al. (2019) [2], the author has given an overall survey/analysis of the current state of machine learning and software defined network (see Figure 2). In the paper the author has discussed about the different types of machine learning algorithms that have already been applied in the field of software defined network and its impact. The author has discussed about the following three algorithms:

- Unsupervised learning
- Supervised learning
- Reinforcement learning

Along with the algorithms the author has also talked over the varies challenges that are faced on integrating machine learning with software defined network. Some of these challenges were as follow:

- Need for machine learning algorithm that are not only accurate but also interpretable
- Machine learning algorithms that can work in real-time
- High quality of data to train the machine learning model

The author [4] has discussed about varies different fields that has the potential of research soon that includes machine learning in different sectors of software defined network like traffic engineering, security, and optimization and the author has become certain that the Machine learning has the potential to improve varies aspects of software defined network environment. It was found that the best algorithm among all was K-Nearest neighbors (KNN) and decision tree (DT) because it gave an accuracy of 90%.

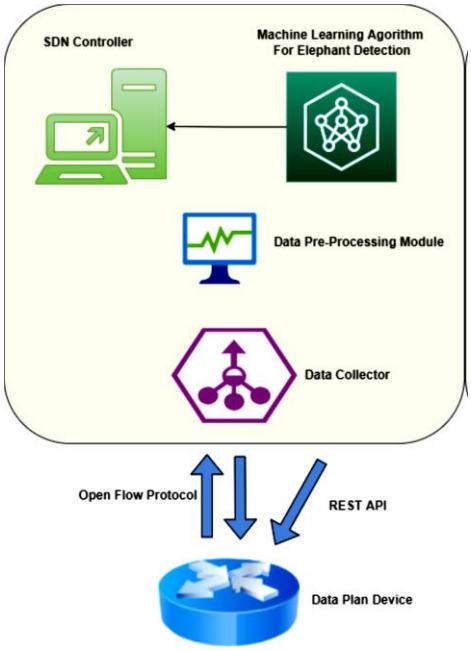

Figure 2: SDN Controller based machine learning model for large flow detection and traffic routing

3.2 Use of Deep Learning in predicting the network traffic

In the research paper of Zhang et al. (2019) [5], the author has introduced a deep learning-based algorithm for distributing resources in software defined network environment (see Figure 3). The author has made use of neural network algorithm to predict the network traffic patterns. The resources were distributed based on the predicted network traffic patterns (see Figure 4). To further the research real world network traffic data was used and it was observed that it performed better than the traditional algorithms [19].

Since the real-world network traffic data was used this algorithm performed exceptionally (see Table 1) well in dynamic environments where the traffic pattern kept on changing (see Figure 5). Although this algorithm has the potential to improve network performance by reducing the latency it had some drawbacks:

- Large amount of training data was not there
- Potential of overfitting

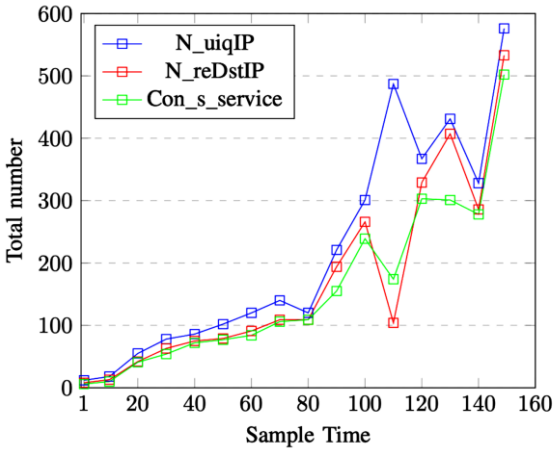

Figure 3: The impact of the added features at attack

Table 1. Accuracy comparison with other classifiers

Classifier	Basic Feature	All Features	11 features
NB	88.957%	97.367%	98.527%
DNN	75.50%	84.77%	85.32%
Logistic Regression	89.61%	93.11%	98.42%
SVM	76%	71%	78%
Our CNN	**96.43%**	**98.94%**	**100%**

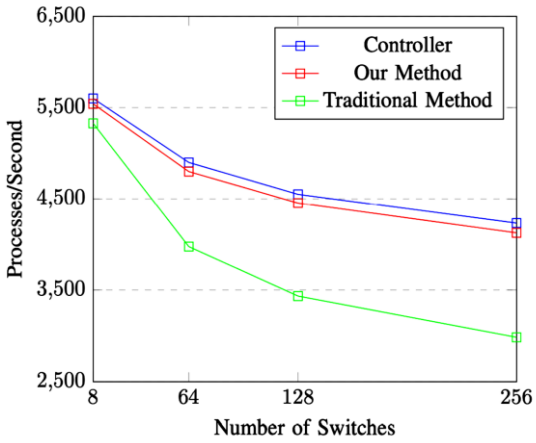

Figure 4: Throughput Evaluation

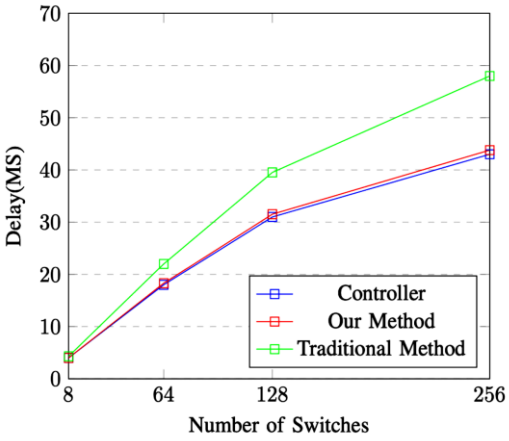

Figure 5: Latency Evaluation

3.3 Machine Learning and Traffic Engineering

In another research paper of Zhang et al. (2018) [6], the author gave an algorithm for traffic engineering in software defined network. This algorithm was primarily based on reinforcement learning algorithm and aimed towards improving traffic routing. On experiment it showed that the algorithm performed better than the traditional algorithm [17].

The author claimed that the algorithm would be effective in difficult/complex network environments where there were minimizing latency along with maximizing throughput (see Figure 6) [7]. It was also simple to implement it for any topology and thus it was considered flexible. However, there were some limitations that were faced:

- Necessity of careful parameter tuning
- Potential for instability

In the paper [12], there are two different scenarios given, regular data delivery over network (Scenario A) and a malicious network (Scenario B). In scenario A the accuracy of RF and LDA was found to be 95% and 98% respectively, and DNN showed 69% accuracy [8]. In scenario B the accuracy of RF was 42%, accuracy of LDA was 76% and accuracy of DNN was 74% (see Table 2).

Table 2. Sender node classification accuracy in scenario A and b

		RF	LDA	DNN
ACCURACY (%)	SCENARIO A	95%	98%	69%
	SCENARIO B	42%	76%	74%

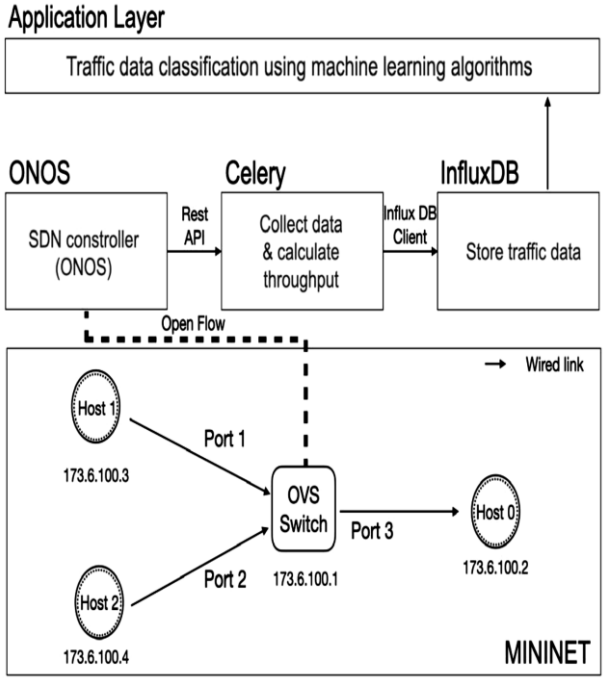

Figure 6: Experiment network model

3.4 Using Machine Learning to find DDoS attacks in Software Defined Network

In the research paper by Amini et al. (2018) [1] used a machine learning algorithm to identify the denial-of-service or DDoS attacks. Denial-of-service [9] attack is basically a cybercrime in which hacker sends a large amount of malicious traffic and as a result it is unable to operate properly. This overload of resources on any website could lead the website to crash. So, it is important to detect the DDoS attacks [10].

In the algorithm the traffic was divided into two groups normal or malicious and used random forest algorithm to do so (see Table 3). When this algorithm was executed against the real-world traffic data than it showed high accuracy in identifying the DDoS attacks (see Figure 7). It could identify any level of attack irrespective of number of hosts or protocols involved.

Although the algorithm was scalable it had a few drawbacks:

- Constant need to update the model
- Potential of false positives

It was observed that the Support Vector Machine (SVM) and Decision Tree (DT) provide the best accuracy and detection rate.

Table 3. Accuracy Rate

ALGORITHM	ACCURACY RATE
DECISION TREE	0.78
SVM	0.85

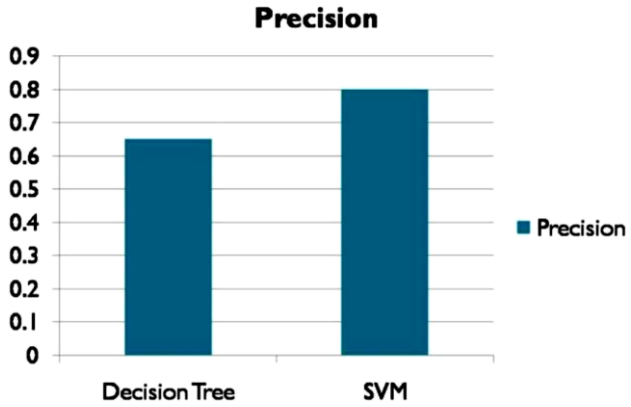

Figure 7: Precision of Machine Learning algorithms

3.5 Intrusion Detection and Machine Learning

Anomaly detection is basically detection of the outliers. Outliers are the data points that are rare or suspicious and different from rest of the cluster [11]. Detection of outliers is important to identify the frauds etc.

In the research paper by Guo et al (2018) [18] introduced a machine learning algorithm that can identify such anomalies in software defined network traffic. In the algorithm it used autoencoder neural network. It used the autoencoder neural network to train the data regarding the low dimensional representation of network traffic. This trained model was then used to detect any outliers (see Figure 8).

The algorithm [20] was also able to detect even the small anomalies that were generally not identified by traditional algorithms. On comparison with traditional algorithm, it performed better against the real-world traffic data [13]. It was observed that Bayes machine learning model had accuracy of 86.9% for DoS attacks and for probe attacks it showed an accuracy of 93.5% (see Figure 9).

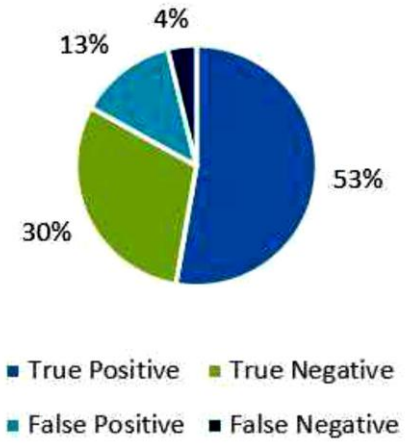

Figure 8: Confusion matrix: Dos Attack

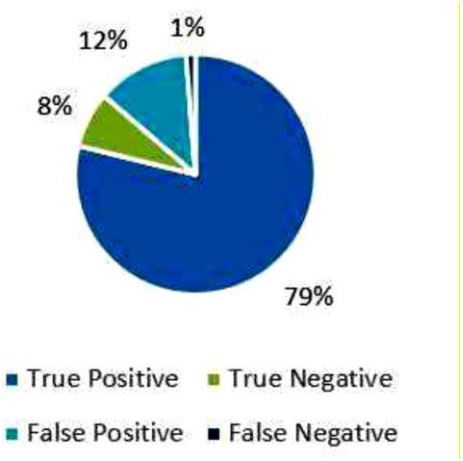

Figure 9: Confusion matrix: Probe Attack

4. INFERENCE

If researched more in the field related to integration of machine learning in software defined network, the machine learning can show great results in terms of scalability, efficiency, security, performance, and flexibility. Machine learning can improve the SDN and give great results when presented with real world data. It can complete complex tasks using classification, clustering, and regression. It is found very useful in various aspects such as traffic engineering [16]. In SDN network, K-Nearest neighbor and decision tree showed the maximum accuracy of 90% [14]. In traffic engineering the algorithms RF, LDA, DNN outperformed all the other algorithms. In terms of DDoS attacks decision tree and Support vector machine was found most useful [15]. In detection of intrusion and probe attacks Bayes algorithm was found to give maximum accuracy.

5. CONCLUSION AND FUTURE WORK

We can conclude that the machine learning has potential to improve software defined network in various fields. Although it stills requires some research to be conducted but it can change the networks in future in terms of efficiency, security, and performance. Algorithms like K-nearest neighbor, decision tree, RF, LDA, DNN, Support vector machine and Bayes were found to be most useful.

Extensive research can solve the problem of integrating machine learning with software defined network and other limitations. There is need to explore this field even further to improve the working of software defined network.

6. REFERENCES

[1] Amini, m., fathi, h., & shahriari, h. R. (2018). A machine learning approach for ddos attack detection in software-defined networking. Journal of ambient intelligence and humanized computing, 9(5), 1455-1467.

[2] M. Mirzaei, M. Mozaffari, and M. Movahedi, “Machine Learning in Software-Defined Networking: A Comprehensive Survey,” *Journal of Network and Computer Applications*, vol. 126, pp. 61–87, 2019.

[3] Guo, y., wu, y., wang, l., & liu, x. (2018). An anomaly detection scheme based on deep autoencoder in software defined network. Ieee access, 6, 21004-21014.

[4] Mirzaei, m., mozaffari, m., & movahedi, m. (2019). Machine learning in software-defined networking: a comprehensive survey. Journal of netowrk and computer applications, 126, 61-87.

[5] Zhang, j., hu, w., & jiang, j. (2018). Traffic engineering in software-defined networks using reinforcement learning. Ieee network, 32(2), 154-161.

[6] Zhang, y., liu, z., liu, y., &fang, y. (2019). Dynamic resource allocation in software defined network using deep learning. Computer networks, 151, 127-138.

[7] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," Journal of Internet Services and Applications, vol. 9, no. 1, pp. 1–99, 2018.

[8] K. Bhushan and B. B. Gupta, ``Distributed denial of service (DDoS) attack mitigation in software dened network (SDN)-based cloud computing environment," J. Ambient Intell. Hum. Comput., vol. 10, no. 5, pp. 19851997, May 2019.

[9] H. Wang and W. Li, ``DDosTC: A transformer-based network attack detection hybrid mechanism in SDN," Sensors, vol. 21, no. 15, p. 5047, Jul. 2021.

[10] J. Kim, H. Kim, M. Shim, and E. Choi, ``CNN-based network intrusion detection against denial-of-service attacks," *Electronics*, vol. 9, p. 916, Jun. 2020.

[11] R. Palanikumar and K. Ramasamy, ``Software de_ned network based self- diagnosing faulty node detection scheme for surveillance applications," *Comput. Commun.*, vol. 152, pp. 333_337, Feb. 2020.

[12] Y. Goto, B. Ng, W. K. G. Seah, and Y. Takahashi, ``Queueing analysis of software de_ned network with realistic OpenFlow_based switch model," *Comput. Netw.*, vol. 164, Dec. 2019, Art. no. 106892.

[13] H. Kim, M. Shin, B. Ahn, J. Lee, S. Lee, S. Lee, J. Ham, and S. Hyeon, "Network intelligence technologies," *ETRI Insight*, 2018.

[14] H. Huang and S. Guo, "Proactive failure recovery for NFV in distributed edge computing," IEEE Communications Magazine, vol. 57, no. 5, pp. 131–137, 2019.

[15] L. Fern´andez Maim´o, A. L. Perales G´omez, F. J. Garc´ıa Clemente, M. Gil P´erez, and G. Mart´ınez P´erez, "A self-adaptive deep learning-based system for anomaly detection in 5G networks," IEEE Access, vol. 6, pp. 7700–7712, 2018.

[16] Dong, S., & Sarem, M. (2019). DDoS Attack Detection Method Based on Improved KNN With the Degree of DDoS Attack in Software-Defined Networks. IEEE Access, 8, 5039-5048.

[17] Dong, S., Abbas, K., & Jain, R. (2019). A survey on distributed denial of service (DDoS) attacks in SDN and cloud computing environments. IEEE Access, 7, 80813-80828.

[18] Gu, Y., Li, K., Guo, Z., & Wang, Y. (2019). Semisupervised K-means DDoS detection method using hybrid feature selection algorithm. IEEE Access, 7, 64351-64365.

[19] Q. A. Al-Haija, "On the Security of Cyber-Physical Systems Against Stochastic Cyber-Attacks Models," 2021 IEEE International IoT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1-6, 2021.

[20] Q. A. Al-Haija, C. D. McCurry and S. Zein-Sabatto, "A Real Time Node Connectivity Algorithm for Synchronous Cyber Physical and IoT Network Systems," 2020 SoutheastCon, pp. 1-8, 2020.