JOURNAL OF BASIC SCIENCE AND ENGINEERING

DESIGN AND APPLICATION OF MACHINE LEARNING METHODS FOR CLOUD SECURITY AND DDOS ATTACK DETECTION

Amarnath J L[1], Dr.Chandramouli H[2], Dr. pritam G Shah[3], Dr. I Manimozhi[4]

[1] Research Scholar, Department of Computer Science and Engineering, East Point College of Engineering and Technology, Bangalore, India
[2] Professor, Department of Computer Science and Engineering, East Point College of Engineering and Technology, Bangalore, India
[3] Chair,SISTMR, Australia.
[4] Professor & Head, Department of Computer Science and Engineering, East Point College of Engineering and Technology, Bangalore, India.

DOI: https://doie.org/10.0126/Jbse.2025695699

ABSTRACT: Cloud computing is a paradigm that delivers software and hardware services over the internet. By adhering to the principles of cloud computing, users can access and manage data and applications from any device connected to the web. In today's digital age, it is almost unimaginable for an individual to be without internet access, even if they could live without a specific gadget. The advantages of cloud computing include scalability, virtualization, user accessibility, lower infrastructure costs, and flexibility. However, one significant drawback is its vulnerability to distributed denial of service (DDoS) attacks. These attacks involve multiple computers simultaneously targeting a specific resource, website, or server to deny service to end users. The onslaught of fraudulent connection requests and an abnormal volume of messages and malformed packets can slow down or shut down the system entirely, preventing legitimate users from accessing the services they need. This article explores the application of machine learning algorithms to detect DDoS attacks. Two primary techniques were employed in this research using datasets from the NSL-KDD repository. On one hand, we utilized the Learning Vector Quantization (LVQ) filter; on the other hand, we applied Principal Component Analysis (PCA), a dimensionality reduction technique. For detecting DDoS attacks, we categorized the characteristics from each approach using Decision Trees (DT), Naïve Bayes (NB), and Support Vector Machines (SVM). We then compared the outcomes of these different categorizations. The results showed that LVQ-based DT outperformed other types of DT in identifying attacks effectively.

Keywords: Distributed Denial of Service, DDoS, Cloud Computing, Security and Machine Learning.

1. INTRODUCTION

The Cloud [1], [2] are two examples of Internet - based platforms that offer servers, databases, and network to consumers and big companies on a large scale, which are perfectly fit in infrastructure reducing models. DDoS attacks are employed by the attackers to prevent the services from being accessed by the rightful users [3]. This kind of attack is so named because the target server is overwhelmed by a flood of requests. The attackers' massive requests overload the target server's bandwidth, making it inaccessible to normal users [4]. By infecting devices on a network with Botnet malware, brute-force distributed denial-of-service attacks can be carried out. Distributed denial of service attacks can be categorised into three types based on their aims and characteristics.

For instance, these attacks couldtarget applications, bandwidth, or traffic. By flooding the targeted server with a flood of TCP or UDP packets, traffic-focused attacks significantly reduce the server's performance. Bandwidth attackers send a flood of sensitive data in an effort to cause congestion. Because application attacks are so common, protecting systems from them is challenging [5]. The application of attack-machine-learning prediction allows for the identification of DDoS attacks.

Artificial intelligence (AI) is an emerging discipline with great promise for solving practical issues in areas like medical image processing[6], sentiment analysis[7], and cloud resource utilisation forecast [80]. Intrusion detection in cloud computing makes use of machine learning. Actually, it's [9]. Multiple approaches to developing intrusion detection systems for the cloud were proposed by researchers. Using autoadaptive evolutionary extreme learning, DDoS attacks can be detected [10]. A Deep Belief Neural Network (DBNN) and a Deep Neural Network (DNN) are used to detect distributed denial of service attacks [11], [12]. The accuracy with which literary methods can be applied to different types of datasets is astounding. Using featureselection and machine learning, this study presents a method for identifying distributed denial of service attacks. When assessing the suggested approach, many metrics are used, including recall, accuracy, precision, and the F measure. There are less miss categorization difficulties and higher accuracy with the proposed technique compared to the existing methods. This research proposes upgrading the machine learning model and choosing the most important variables to decrease the amount of miss categorization mistakes that happen during DDoS attack detection.

2. LITERATURE REVIEW

Idhammad M et al. (2018) states that cloud-based HTTP DDoS attacks can now be detected. The suggested detection method is powered by a combination of RF ensemble learning techniques and information theoretic entropy (ITE). In a Cloud setting built on the Open Stack platform, the entropy of the network header characteristic of incoming traffic signals is calculated using a time-based sliding window approach. Reference [13] states that when the pre-processing projected entropy is greater than the usual range, classification jobs are generated.

Using an anomalous intrusion detection approach at the hypervisor layer, Rawashdeh A et al. (2018) increase the performance of distributed denial of service (DDoS) attacks across virtual machines. The discovering mechanism was produced by the autonomously evolving neural network. Evolving neural networks integrate particle swarm optimisation (PSO) with neural networks to detect distributed denial of service attacks and categorise traffic data [14]. Most earlier research focused with the algorithms' performance on the KDD CUP 99 and NSL-KDD datasets. The majority of the data in the dataset pertains to traffic generated by virtual machines (VMs), although future research can expand to include traffic generated by host machines as well. When it comes to cloud computing, Kushwah, G.S., et al. (2020) suggested a way to detect DDoS attacks. [15] Voting ELM, also known as VELM, is used by the novel detection method.

You can use either the NSL-KDD or the ISCX datasets for intrusion detection. The suggested system outperforms the aforementioned artificial neural networks (ANNs) in terms of accuracy, as well as ELM, random forest, Adaboost, and black hole optimisation ANN. This study presents a model for detecting Distributed Denial of Service attacks that is based on ELM (Kushwah et al., 2019). To perform the actual NSL-KDD experiments, the virtual environment is populated with the following sample data set. Presumably, it is possible to have a high detection rate and at the same time the

calculation time will be low with assistance of the suggested detection model. Very good. The writing was done by Hezavehi et al. The DDoS attack identification is incorporated in the Third Party Auditor Notification Generator (TPA). The TPANGNDn architecture which contains the both features of the detection notification and the third-party auditor notification is said to be the way to detect the detection. To guard public cloud settings against DDoS TCP flood attack, Sahi A et al. (2017) proposed a classification based technique. It is now possible to devise new techniques for detecting distributed denial of service attacks, determine the arriving packets and manage to make the decisions necessary for the protection of stored data based on the results of categorization. An example is where there was a flood attack and the use of Wireshark to uncover the attack and stop it was demonstrated. The indicated detection techniques follow up a packet in order to determine if it was produced by an attacker or not in the process of prevention.

The detection approach given by Wani A. R. and colleagues (2019) is of the support vector machine (SVM) kind. When it comes to accuracy, recall, specificity and f measure, it can be brought out that the Support Vector Machine (SVM) is the best among the three. Coming second is Random Forest, with an importance score of 13. Lately it has been established that one can be able to use the Tor Hammer to attack datasets stored within the cloud [16]. A distributed denial of service attack detection system that relies on cloud computing and machine learning was created by Hefz et al. (2017). This particular detection approach prevents network packages from escaping the system by using information related to virtual machines and hypervisors [17].

3. MATERIALS AND PROCEDURES

In this part, we analyse the NSL-KDD dataset [18] to identify DDoS attacks. Figure 1 depicts the procedure of the intrusion detection model. Minimising features is done via two distinct feature selection techniques. The constrained feature set is used by many categorization algorithms. Calculate validation metrics using the confusion matrix as a basis.

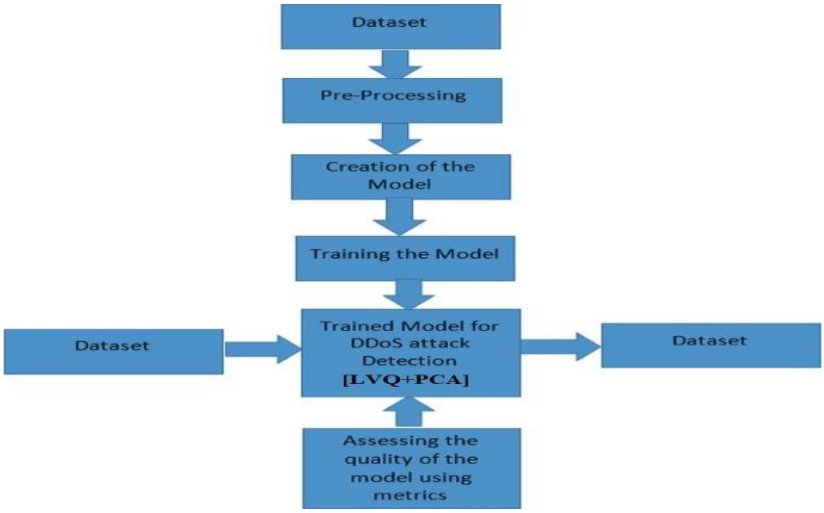

Figure1: Proposed Model for Attack Detection in Cloud Environment

Vol. 22, No. 1, (2025)
ISSN: 1005-0930

Setting up real distributed testing networks is an expensive ordeal. Network researchers can benefit from simulation in their work since it allows for the inexpensive evaluation of issues under different protocols, traffic, and topologies. You can choose to access the public data sets or the direct ones. Public datasets, on the other hand, are user-generated using a variety of online venues, whilst direct data sets are created utilising open source tools. For this study, we use the publicly accessible dataset NSL-KDD [19]. There are 42 attributes, 2, 26,283 occurrences, and four distinct attack categories as per the initial data set. The 15,452 instances of distributed denial of service attacks are the only subject of this research. The training set makes up 70% of the data set, whereas the testing set makes up 30%. R makes use of this idea that has been offered before.

3.1 Feature Selection

The Feature Selection Approach is used to identify the essential components that impact the predicted variable. The prediction data will remain same regardless of whether feature selection data is added or removed. The suggested technique makes use of feature selection methods. It is possible to use both dimensionality reduction and filtering.

3.1.1 Method of Filtering

An example of an algorithm employed by artificial neural networks is the supervised learning LVQ algorithm. The LVQ includes n input units and m output units in its design. All of the weighted layers are linked. According to [20], LVQ employs a k-NN method. The LVQ parameters used for training process are x,T,wj,Cj and j. x is a training vector x(x1,x2……xn). T is the class for training vector x. wj is weight vector for jth output unit. Cj is the class associated with the jth output unit.

parameters used for training process are x,T,wj,Cj and j. x is a training vector x(x1,x2……xn). T is the class for training vector x. wj is weight vector for jth output unit. Cj is the class associated with the jth output unit.

Vol. 22, No. 1, (2025)
ISSN: 1005-0930

Algorithm:

Step 1: *Initialize, determine the initial weight, maximum epoch (number of training processes to be repeated) and the learning rate (alpha) value.*

Step 2: *If repetition conditions are fulfilled, do steps 2- 8.*

Step 3: *Set initial conditions epoch= 0.*

Step 4: *If the condition (epoch<MaxEpoch) then*

epoch= epoch+ 1

Step 5: *Calculate the minimum distance ||xi-wj|| using Euclidean distance.*

Step 6: *Update weight wj with the conditions: If T=Cj, then wj (new) = wj(old) + α(x – wj(old))*

If T≠Cj, then wj (new) = wj(old) –α(x– wj(old))

Step 7: *Reduce learning rate (α)=α – (0, 1* α)*

Step 8: *Stop condition test: the condition where the learning rate (α) and the error reach the specified target value.*

3.1.2 Dimensionality reduction method

Dimensionality reduction is carried out by gradually removing superfluous properties. Power metric analysis reduces dimensions. PCA reduces the number of data variables from a large set to a manageable one [20]. It is easy to see the data's variables when you utilise data set patterns. Principal component analysis (PCA) changes the data variables using an orthogonal statistical variable distribution. Eigen values, Eigen vectors, Eigen values, and Standardization are mathematical notions that are used in principal component analysis (PCA).

3.2 Methods of Classification

The NSL-KDD data is analysed using the classification methods. Classification algorithms are useful tools for data prediction. Classification algorithms use a model to map the values of the predictors to the values of the targets [20]. Data for unknown classes could be represented by a model, depending on the link. In this study, Decision Tree categories, Support Vector Machine, and Naïve Bayes are used as classification algorithms. These categorization methods differentiate between safe and dangerous records in order to foresee data sets.

3.2.1 Naïve Bayes (NB)

A method of machine learning grounded on Bayes' Theorem Naive Bayes is based on probability theory. It has several applications in classification tasks. The feature probabilities should be calculated, and the highest one should be chosen. Naive Bayes classifiers make assumptions about unrelated features. The possibility of malicious records in the dataset is estimated.

3.2.2 Support Vector Machine (SVM)

To partition large datasets with distinct class members, a decision plane is necessary. Using support vector machines (SVM), decision planes with specified bounds are feasible. Assuming it's feasible, it

creates a line that divides the classes after receiving the data. What we call "support vectors" are the pairs of data points from each class that are positioned closest to the line. Margins are defined as the distance between hyperplanes and support vectors. The ideal hyperplane is the one with the largest margin [20]. The SVM method yielded LVQ and PCA values of 0.9288 and 0.9847, respectively, using the provided data set.

3.2.3 Decision Tree (DT)

Decision trees, or DTs, are useful for making predictions and classifications. The decision tree algorithm uses the current state of affairs to generate potential future outcomes. When developing this algorithm, we considered every possible outcome of our decisions. A large number of subsets and nodes make up DT, which represents every choice and its consequences [20]. Because it uses a tree-like structure to account for all potential final judgements, the DT technique more correctly represents the data. The implementation of this approach relies heavily on recursive decision-making. Even when working with data that has several dimensions, it maintains accuracy. Applying the DT technique to the provided data set yielded LVQ accuracy of 0.9874 and PCA accuracy of 0.9860.

1. OUTCOMES

The findings have been obtained. An R-based prediction of the malicious record is made using the NSL-KDD data. Feature selection methods allow one to acquire a collection of categorization features. The validation metrics that were achieved by applying the f-measure, recall, specificity, accuracy, and precision are shown in Tables 1 and 2. The tables display these metrics.

4.1 Evaluation Measures

Using assessment metrics, we can see how well the prediction model is doing. This research used precision, accuracy, recall, and F score to evaluate machine learning's capacity to detect distributed denial of service (DDoS) attacks.

4.4.1 Accuracy

Accuracy, or the percentage of observations that were properly predicted, is the most essential performance statistic. To be a practical metric for assessment, accuracy requires datasets to be uniform and to have an equal amount of false positives and false negatives. In Equation (1), we can see that the classifier is good at predicting future data points.

$$Accuracy = \frac{TP}{TP + TN + FP + FN} \quad (1)$$

4.4.2 Precision

Vol. 22, No. 1, (2025)
ISSN: 1005-0930

Accuracy is defined as the ratio of anticipated positive observations to the number of positively predicted observations. Reducing false-positives is a key component of high accuracy. The accuracy with which classifiers predict positive classifications is called precision. Equation (2) is used to compute the accuracy.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

4.4.3 Recall

Recall is defined as the percentage of positive observations that were correctly predicted relative to the total number of observations in a class. According to Equation (3), the accuracy with which the classifier anticipates the positive class is measured by its precision.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4.4.4 F1 Score

Both recall and accuracy are normalised as components of the F1 Score. Consequently, both false positives and false negatives are included in this score. Despite its seeming simplicity, F1 score outperforms accuracy in cases when the distribution of classes is uncertain. The F1 value, as shown in Equation (4), is the harmonic mean of the accuracy and recall scores.

$$FMeasure = 2 \times \frac{PR}{(P + R)} \tag{4}$$

4.2 Classification on features using Learning Vector Quantization method

Table 1 and image 2 show the outcomes of the tests performed on the dataset. Next, DT, SVM, and NB are used for classification after the use of LVQ. When compared to NB and SVM, the DT classifier performs better at detecting malicious data.

Table1: Results of Learning Vector Quantization Method

Parameter	Naïve Bayes	SVM	Decision Tree
Accuracy	0.9197	0.9288	0.9874
Precision	0.9085	0.9060	0.9887
Recall	0.9727	0.9897	0.9914
Specificity	0.8250	0.8249	0.9808
F-Measure	0.9395	0.9460	0.9901

Vol. 22, No. 1, (2025)
ISSN: 1005-0930

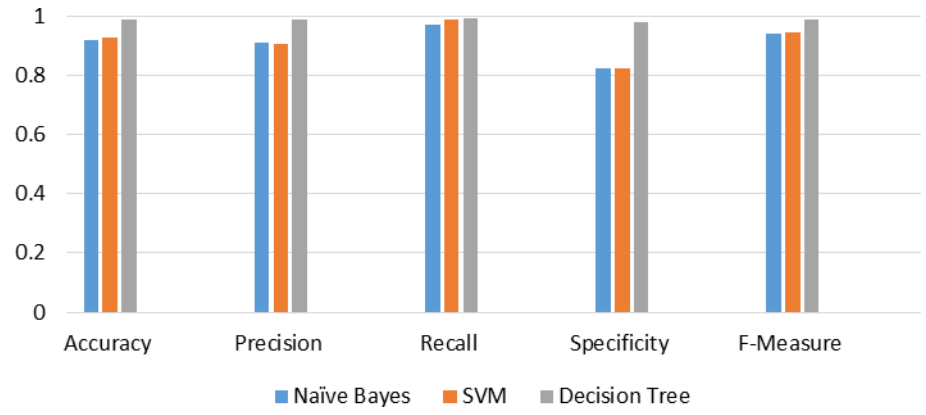

Figure2: Learning Vector Quantization Method

4.3 PCA Method for Classification

Image 3 displays the outcomes of dimensionality reduction using principal component analysis (PCA). The DT technique outperforms the NB and SVM approaches with a detection accuracy of 0.9860, presented in the mentioned table 2. Out of 42 attributes, 22 are considered in this feature selection technique.

Table2: Results of Principal Component Analysis Method

Parameter	Naïve Bayes	SVM	Decision Tree
Accuracy	0.8721	0.9847	0.9860
Precision	0.9562	0.9904	0.9983
Recall	0.8561	0.9851	0.9781
Specificity	0.9358	0.9839	0.9972
F-Measure	0.9034	0.9878	0.9881

Vol. 22, No. 1, (2025)
ISSN: 1005-0930

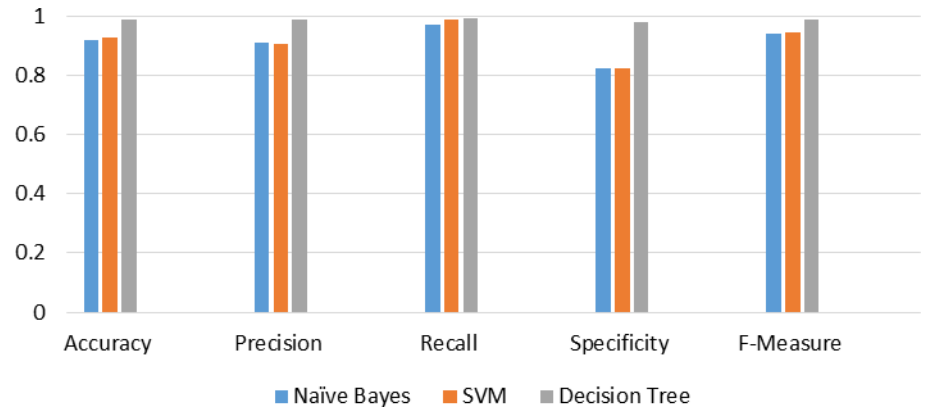

Figure3: Learning Vector Quantization Method

5. CONCLUSION

Identifying distributed denial of service (DDoS) attacks is a complex challenge that requires robust solutions. Given the disruptive impact of these attacks on cloud services, accurate detection is paramount. Machine learning models have emerged as effective tools for detecting such attacks, and this study is driven by the need to enhance the accuracy of DDoS attack detection. One notable example of a data intrusion detection system is the one based on the NSL-KDD benchmark dataset. This research focuses exclusively on data related to DDoS attacks. To classify these attacks, we employed a combination of feature selection methods, including Principal Component Analysis (PCA) and Learning Vector Quantization (LVQ), along with machine learning algorithms such as Decision Trees (DT), Support Vector Machines (SVM), and Naïve Bayes (NB).The classification of DDoS attacks was evaluated based on the performance of these algorithms. From a total of 42 attributes, LVQ selected 20 features and PCA selected 21 features. The results indicate that the DT model using LVQ-based feature selection outperforms other techniques in terms of accuracy, recall, specificity, and F-score.In summary, this study demonstrates that the integration of LVQ with DT significantly improves the detection accuracy of DDoS attacks compared to other methodologies.

REFERENCES

[1]. Alahmadi, Amal A., et al. "DDoS Attack Detection in IoT-Based Networks Using Machine Learning Models: A Survey and Research Directions." Electronics 12.14 (2023): 3103.

[2]. Bhayo, Jalal, et al. "Towards a machine learning-based framework for DDOS attack detection in software-defined IoT (SD-IoT) networks." Engineering Applications of Artificial Intelligence 123 (2023): 106432.

[3]. Alhalabi, Wadee, et al. "Machine learning-based distributed denial of services (DDoS) attack detection in intelligent information systems." International Journal on Semantic Web and Information Systems (IJSWIS) 19.1 (2023): 1-17.

[4]. Tuan, Tong Anh, et al. "Performance evaluation of Botnet DDoS attack detection using machine learning." Evolutionary Intelligence 13 (2020): 283-294.

[5]. Potluri, Sirisha, et al. "Detection and prevention mechanisms for ddos attack in cloud computing environment." 2020 11th international conference on computing, communication and networking technologies (ICCCNT). IEEE, 2020.

[6]. Pande, Sagar, et al. "DDOS detection using machine learning technique." Recent Studies on Computational Intelligence: Doctoral Symposium on Computational Intelligence (DoSCI 2020). Springer Singapore, 2021.

[7]. Rajapraveen, K. N., and Rohitha Pasumarty. "A Machine Learning Approach for DDoS Prevention System in Cloud Computing Environment." 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS). IEEE, 2021.

[8]. Bagyalakshmi, C., and E. S. Samundeeswari. "DDoS attack classification on cloud environment using machine learning techniques with different feature selection methods." International Journal of Advanced Trends in Computer Science and Engineering 9.5 (2020).

[9]. Arunkumar, M., and K. Ashok Kumar. "Malicious attack detection approach in cloud computing using machine learning techniques." Soft Computing 26.23 (2022): 13097-13107.

[10]. Manjunath, C. R., et al. "Cloud Based DDOS Attack Detection Using Machine Learning Architectures: Understanding the Potential for Scientific Applications." International Journal of Intelligent Systems and Applications in Engineering 10.2s (2022): 268-271.

[11]. Sambangi, Swathi, and Lakshmeeswari Gondi. "A machine learning approach for ddos (distributed denial of service) attack detection using multiple linear regression." Proceedings. Vol. 63. No. 1. MDPI, 2020.

[12]. Mishra, Anupama, et al. "Classification based machine learning for detection of ddos attack in cloud computing." 2021 ieee international conference on consumer electronics (icce). IEEE, 2021.

[13]. Peneti, Subhashini, and E. Hemalatha. "DDOS attack identification using machine learning techniques." 2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2021.

[14]. Makkawi, Ahmed Mohammed, and Adil Yousif. "Machine Learning for Cloud DDoS Attack Detection: A Systematic Review." 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). IEEE, 2021.

[15]. Wani, Abdul Raoof, Q. P. Rana, and Nitin Pandey. "Machine learning solutions for analysis and detection of DDoS attacks in cloud computing environment." Int. J. Eng. Adv. Technol 9.3 (2020): 2205-2209.

[16]. Singh Samom, Premson, and Amar Taggu. "Distributed denial of service (DDoS) attacks detection: A machine learning approach." Applied Soft Computing and Communication Networks: Proceedings of ACN 2020. Springer Singapore, 2021.

[17]. Ouhssini, Mohamed, and Karim Afdel. "Machine Learning Methods for DDoS Attacks Detection in the Cloud Environment." International Conference on Advanced Intelligent Systems for Sustainable Development. Cham: Springer International Publishing, 2020.

[18]. Santos, Reneilson, et al. "Machine learning algorithms to detect DDoS attacks in SDN." Concurrency and Computation: Practice and Experience 32.16 (2020): e5402.

[19]. Revathi, M., V. V. Ramalingam, and B. Amutha. "A machine learning based detection and mitigation of the DDOS attack by using SDN controller framework." Wireless Personal Communications (2021): 1-25.

[20]. Luong, Tan-Khang, Trung-Dung Tran, and Giang-Thanh Le. "Ddos attack detection and defense in sdn based on machine learning." 2020 7th NAFOSTED Conference on Information and Computer Science (NICS). IEEE, 2020.