STM JOURNALS

International Journal of
Satellite Remote Sensing

Volume 2, Issue 1, 2024
January–June
DOI (Journal): 10.37591/IJSRS

https://journals.stmjournals.com/ijsrs

Research IJSRS

ML-associated DoS and DDoS Attack Observation in Protection

Vaishnavi Jagtap[1], Sayli Bhosale[1], Pratiksha Malekar[1,] *, Sanika Dhumal[1]

Abstract

DoS and DDoS assaults are significant risks to the availability and integrity of online services and networks. Attack traffic might come from a variety of geographical regions, making it difficult to filter and neutralize the attack. DDoS attacks are far more sophisticated and powerful than DoS attacks. They use a network of compromised devices, known as a botnet, to launch a coordinated attack on a target. Monitoring and evaluating the network for odd trends, such as rapid increases in traffic volume or changes in traffic distribution. This paper presents an approach for the detection of DoS and DDoS attacks using a combination of mathematical and entropy-based methods. The proposed approach lever ages the inherent characteristics of these attacks to develop robust detection mechanisms that enhance network security. Machine learning algorithms, particularly those based on supervised and unsupervised learning, are becoming increasingly prevalent in the detection of DoS and DDoS attacks. This paper provides insights into the application of machine learning for attack classification and the development of predictive models to anticipate new attack vectors.

Keywords: Denial of service (DoS) attack detection, distributed denial of service (DDoS) attack detection, mathematical methods, machine learning, entropy methods

INTRODUCTION

For computer networks and systems to remain available and secure, this article on identifying denial of service (DoS) and distributed denial of service (DDoS) assaults is essential. Mathematical and entropy-based methods are among the various techniques used for detecting and mitigating such attacks. Attack traffic can originate from various geographic locations, making it challenging to filter and mitigate the attack. DDoS attacks are more sophisticated and formidable than DoS attacks. They entail a group of hacked devices, commonly known as a "botnet," that are coordinated to attack a target in a coordinated manner. Monitoring and analyzing network for unusual patterns, such as sudden spikes in traffic volume changes in traffic distribution, are required. This can involve statistical measures like mean, median, standard deviation rate analysis. Entropy-based methods utilize the concept of information entropy to detect abnormal patterns in network traffic. Entropy measures the randomness or uncertainty of data. In network traffic analysis, high entropy can indicate the presence of unusual or malicious activities [1–4].

***Author for Correspondence**
Pratiksha Malekar
E-mail: malekarpratiksha311@gmail.com

[1]Student, Department of Computer Engineering, Rajgad Dnyanpeeth's Shree Chhatrapati Shivajiraje College of Engineering, Dhangwadi, Bhor, Pune, Maharashtra, India

Received Date: May 23, 2024
Accepted Date: June 10, 2024
Published Date: June 21, 2024

Citation: Vaishnavi Jagtap, Sayli Bhosale, Pratiksha Malekar, Sanika Dhumal. ML-associated DoS and DDoS Attack Observation in Protection. International Journal of Satellite Remote Sensing. 2024; 2(1): 17–24p.

MOTIVATION

The motivation behind using mathematical and entropy methods for DoS and DDoS attack detection is rooted in the need to enhance the efficiency, accuracy, and reliability of detecting these types of cyber-attacks.

DoS and DDoS attacks are malicious attempts to interrupt the normal operation of a target system or

© STM Journals 2024. All Rights Reserved

network by flooding it with a large volume of traffic, making it inaccessible to legitimate users.

LITERATURE SURVEY

The authors of one study proposed a technique based on document popularity to detect application layer DoS assaults on well-known websites. The authors propose employing an access matrix to identify any existing spatial–temporal pattern of a usual increase in the number of website visitors. To abstract the matrix, the researchers examined both the major and independent components. A model is created to detect DDoS assaults based on the entropy of the document's popularity.

Nishanth and Mujeeb [5] use Bayesian inference to model and detect flooding-based denial-DoS in wireless ad hoc networks. Flooding-based DoS attacks are a major problem in all types of wireless ad hoc networks, including wireless sensor networks, Mobile Ad hoc Network (MANET), Vehicular Ad Hoc Networks (VANET), and flying area networks. The SYN traffic is modeled using Bayesian inference, and an efficient algorithm is developed for detecting persistent flooding-based DoS attacks that are applicable to different types of wireless ad hoc networks by modifying the mean of the beta distribution [5].

Elsaeidy et al. [6] developed a hybrid deep learning approach for detecting replay and DDoS attacks in smart cities. The methodology's performance was evaluated using synthetically generated replay and DDoS attack data. Attack data was derived from real-life typical behavior observed in the smart city of Queanbeyan, Australia. The methodology's performance was compared with literature-based machine and deep learning models [6].

Li et al. [7], in their article, "Denial of Service (DoS) Attack Detection: A Performance Comparison of Supervised Machine Learning Algorithms," employed a dataset to categorize denial of service attacks using the naive Bayes technique, artificial neural networks, and logistic regression, and compared the results. The experimental results reveal that the neural network method outperformed logistic regression and the naive Bayes algorithm on a dataset with a slightly skewed distribution [7].

EXISTING SYSTEM

In the context of network traffic, abnormal levels of entropy can indicate potential attacks. These systems establish baseline models of normal network behavior and compare real-time traffic against these models. These systems use mathematical models to classify normal and malicious traffic based on features extracted from network packets. DoS or DDoS attacks. Mathematical methods are used to dynamically adjust thresholds based on changing network conditions and traffic patterns. By monitoring communication paths and analyzing traffic flows between network nodes, these systems can identify anomalies indicative of attack traffic. Mathematical methods are applied to analyze network topology and detect abnormal communication patterns [8–12]. Block diagram of existing model is shown in Figure 1.

PROBLEM STATEMENT

DoS and DDoS attacks are malicious activities that try to interrupt the availability of online services by flooding them with a large amount of traffic. Detecting and mitigating these assaults in real time is critical to ensuring the integrity and availability of online services.

METHODOLOGY

Method Overview

1. Gather network traffic data from various sources within the network, such as routers, switches, or intrusion detection systems (IDS).
2. Trigger alarms or alerts when the observed values exceed the defined thresholds, indicating potential DoS or DDoS attack activities.

© STM Journals 2024. All Rights Reserved

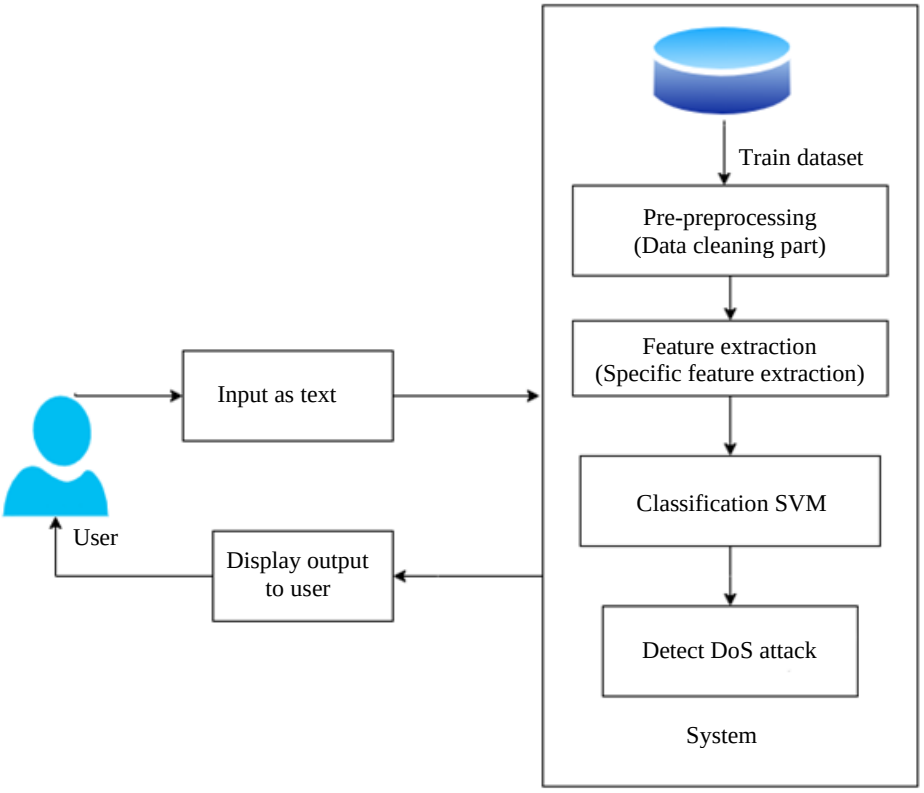

Figure 1. Block diagram of existing model.

3. Put in place automatic response systems to reduce reaction times and successfully lessen the impact of the attack.
4. Evaluate the effectiveness of the detection methodology by comparing detected anomalies with known attack instances and assessing the accuracy of the detection mechanism.
5. Upon detecting a potential attack, initiate response mechanisms such as traffic filtering, rate limiting, or blacklisting to mitigate the impact of the attack.

DATA PREPROCESSING

The dataset collected from the packet tracer contains various fields like duration, Total forward packet, total forward packet, total backward packet, total length of forward packets, total length of backward packets, forward packet length max, and many more fields. Effectively, duration, total forward packets, total backward packets, down up ratio, act data packet forward, min seg size forward, label, average packet size this are required in the proposed work. Hence under data preprocessing unwanted fields are removed.

SUPPORT VECTOR MACHINE

The support vector machine, or SVM, is a popular supervised learning technique for classification and regression issues. But in Machin, it is mostly used for categorization problems. The SVM algorithm seeks to produce the optimal line or decision boundary that can partition n-dimensional space into classes to facilitate future classification of new data points [13].

A hyperplane is the border of the best option. SVM is used to choose the extreme points and vectors that contribute to the formation of the hyperplane. These extreme situations are known as support vectors, and the method is known as a support vector machine. Examine Figure 2's diagram, which uses a decision boundary or hyperplane to classify two distinct groups. They can be utilized to circumvent the challenges of utilizing linear functions in high-dimensional feature space, and the optimization problem is turned into dual convex quadratic programming [14].

© STM Journals 2024. All Rights Reserved

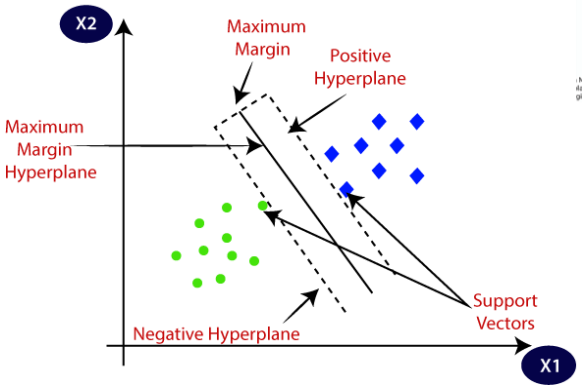

Figure 2. Random forest.

Working

One popular machine learning method that falls under the supervised learning category is called Random Forest (see Figure 2). It can be applied to jobs involving both regression and classification. The foundation of this approach is ensemble learning, which combines several classifiers to address challenging issues and enhance model performance. In order to improve accuracy, Random Forest aggregates many decision trees constructed using various data subsets and then averages their predictions. Unlike relying on a single decision tree, the random forest takes into account the predictions from all trees and generates the final output based on the majority vote, as illustrated in Figure 3. As the number of trees in the forest increases, the accuracy improves, and the risk of overfitting decreases.. The graphic in Figure 2 illustrates how the random forest algorithm works.

DECISION TREE

Although they are usually utilized for classification problems, decision trees are supervised learning techniques that can be applied to both classification and regression applications. This model is structured like a tree, where internal nodes represent features of the dataset, branches correspond to decision rules, and the terminal nodes (leaves) represent the final outcomes [15, 16]. As seen in Figure 4, a decision tree has two different kinds of nodes: decision nodes and leaf nodes.

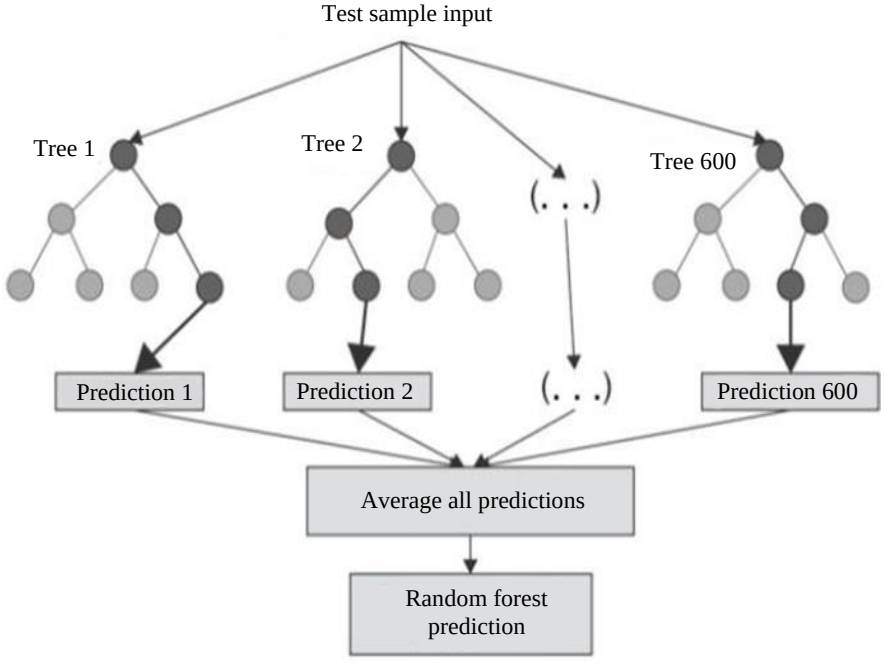

Figure 3. Average predictions based on prediction of various trees.

© STM Journals 2024. All Rights Reserved

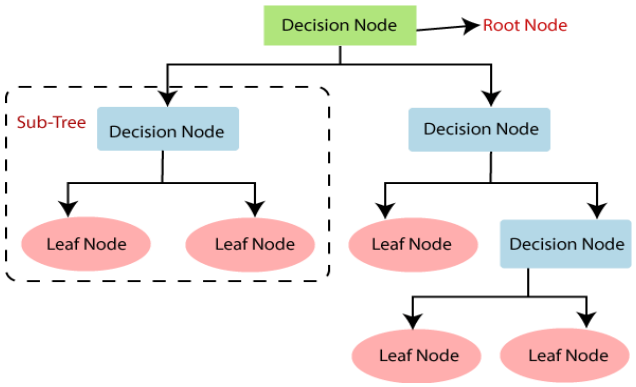

Figure 4. Division of decision node.

While leaf nodes indicate the results of such decisions and lack further branches, decision nodes are in charge of making decisions and have several branches. A decision tree works by posing a question and, depending on the answer (yes/no), splits into subtrees. The usual structure of a decision tree is depicted in the diagram below.

CLASSIFICATION REPORT

Precision

Precision is a performance metric used to assess a machine learning model, specifically measuring the accuracy of its positive predictions. It is determined by dividing the number of true positives (TP) by the total number of positive predictions, which includes both FP and TP.Precision determines the proportion of correctly identified positive instances out of all those predicted as positive [17]. As a result, the precision calculation formula is:

Precision = TP / (TP+ FP)

Recall

The ability of a model to distinguish positive examples (true positives) from all of the real positive samples in the dataset is measured by a metric called recall. Recall is computed by dividing the number of true positives by the total number of positive cases, which includes both false negatives and true positives. It is often referred to as the true positive rate (TPR). It displays the model's ability to include every pertinent instance in a dataset. The ratio of true positives to the total of true positives and false negatives (FNs) is the mathematical definition of recall.

Recall= TP / TP + FN

F1 Score

A machine learning model's performance is assessed using the F1 score, which strikes a balance between recall and precision.. It's commonly used for binary and multiclass classification tasks. The F1 score runs from 0 to 1, with 1 representing a perfect outcome and 0 representing the lowest possible result. A high F1 score implies a well-balanced performance, whereas a low F1 score frequently indicates a compromise between recall and precision. The performance of two classifiers is mostly assessed using the F1 score. It might not, however, accurately depict how well the model performs when dealing with unbalanced data. This is because the traditional F1 score assigns equal weights to precision and recall. The following formula is used to calculate the F1 score:

F1 = 2 * (Precision + Recall) / (Precision * Recall)

Accuracy

© STM Journals 2024. All Rights Reserved

Accuracy is a performance metric that represents the proportion of correct predictions the model generates. This statistic gives a broad overview of how well the model performs in each class. It is described as:

Accuracy = Number of correct predictions / Total Number of predictions.

PROPOSED SYSTEM

A proposed system for detecting DoS and DDoS attacks using mathematical and entropy methods would involve the integration of various algorithms and techniques designed to analyze network traffic patterns and identify anomalous behavior indicative of potential attacks. Extract features from the preprocessed data that are relevant for detecting DoS and DDoS attacks. Utilize mathematical models such as statistical analysis, machine learning algorithms, or time-series analysis to detect anomalies in the network traffic. Trigger alarms or alerts when the observed values exceed the defined thresholds, indicating potential DoS or DDoS attack activities. Upon detecting a potential attack, initiate response mechanisms such as traffic filtering, rate limiting, or blacklisting to mitigate the impact of the attack. Put in place automatic response systems to reduce reaction times and successfully lessen the impact of the attack.

RESULT ANALYSIS

DoS Attack

DoS attack detection using thresholding and entropy-based method in the form of graph is shown in Figure 5.

DDoS Attack

DDoS attack detection using thresholding and entropy-based method is shown in Figure 6.

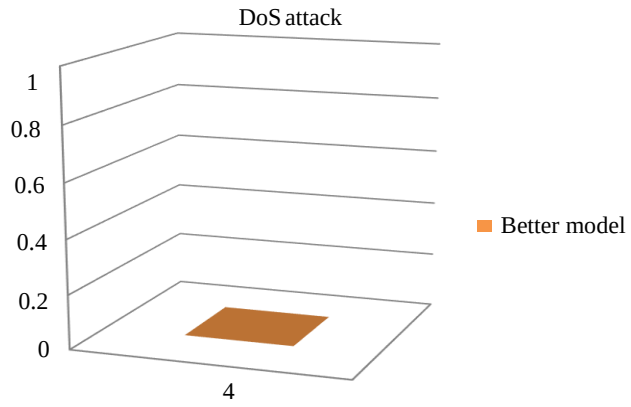

Figure 5. Denial of service (DoS) attack detection using thresholding and entropy-based method.

© STM Journals 2024. All Rights Reserved

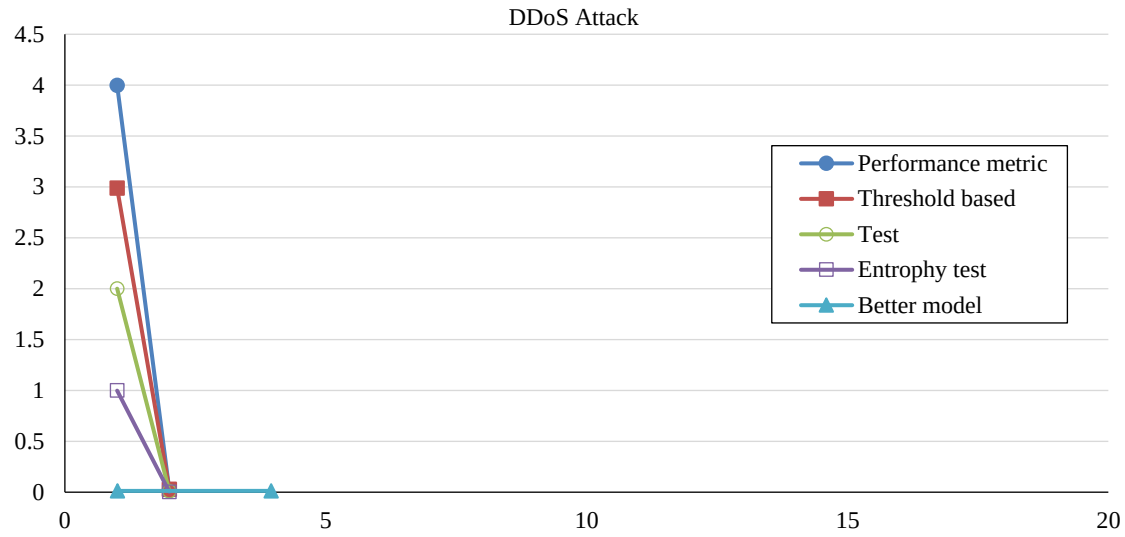

Figure 6. Detection of distribute denial of service (DDoS) attack using thresholding.

APPLICATION AND ADVANTAGE

The advantages are

1. This provides a reliable way to distinguish between normal and malicious activities.
2. These methods can be applied at various scales, from individual network devices to larger network segments.
3. This scalability enables detection and prevention at multiple levels within network architecture.
4. Establishing baseline patterns of typical network behavior is a common step in mathematical and entropy methodologies.
5. By comparing current network traffic against these base- lines, anomalies caused by attacks can be detected.

What Is Machine Learning?

Machine learning is a subfield of artificial intelligence that creates algorithms by learning the underlying patterns of datasets and using them to generate predictions on new comparable types of data without being explicitly coded for each task. By merging data and statistical algorithms, traditional machine learning forecasts an output that can be utilized to provide significant insights. Applications for machine learning are numerous and include automated tasks, fraud detection, recommendation systems, natural language processing, image and speech recognition, portfolio optimization, and more. Autonomous cars, drones, and robots are also powered by machine learning models, which increase their intelligence and adaptability to changing situations.

Applications of Machine Learning

• Automation: Without human assistance, machine learning functions entirely on its own in any field. For instance, in manufacturing plants, robots perform crucial process steps.

- Finance Sector: The banking industry is seeing an increase in the use of machine learning. Banks primarily utilize it for identifying patterns in data and for fraud prevention.
- Government: Governments use machine learning to manage public safety and services. For instance, in China, facial recognition technology is widely used, and AI systems help address issues such as jaywalking. One of the first sectors to use machine learning for picture recognition was the healthcare sector.
- Marketing: Due to the vast amounts of available data, artificial intelligence is frequently used in marketing. Before the big data era, researchers applied complex mathematical techniques like Bayesian analysis to assess customer value. With the growth of data, marketing teams now rely on AI to enhance customer engagement and optimize marketing efforts.. The retail industry uses machine learning to evaluate customer behavior, estimate demand, and manage

© STM Journals 2024. All Rights Reserved

inventories. It also enables companies to personalize the shopping experience for each client by proposing products based on previous purchases and interests.

- Transportation: To improve overall system efficiency, reduce fuel consumption, and optimize routes, the transportation industry uses machine learning. It also has an impact on autonomous vehicles, which use machine learning algorithms to make navigation and safety judgments.

CONCLUSION

In this study, we compare two models: mathematical and entropy based. Considering the performance metrics such as accuracy, precision, recall, and F1-score, we conclude that the entropy-based approach, which combines cumulative entropy trend analysis with thresholding, yields superior results. Border cases like rejection of ping, or a computer request to a blocked IP address giving a ping very similar to a system under attack as seen in the packet tracer is also identified in the second model.

Future Scope

Future work in DoS and DDoS attack detection will likely involve a multidisciplinary approach, combining advances in machine learning, network security, and data analytics to stay ahead of increasingly sophisticated attackers. Collaboration, research funding, and knowledge-sharing within the cyber security community will be key drivers of innovation in this field.

REFERENCES

1. Saxena U, Sodhi JS, Singh Y. An analysis of DDoS attacks in smart home networks. In: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, January 29–31, 2020. pp. 272–276.
2. Qin X, Xu T, Wang C. DDoS attack detection using flow entropy and clustering technique. In: 2015 11th International Conference on Computational Intelligence and Security (CIS), Shenzhen, China, December 19–20, 2015. pp. 412–415. doi: 10.1109/CIS.2015.105.
3. Zlomislic V, Fertalj K, Sruk V. Denial of service attacks: An overview. In: 2014 9th Iberian Conference on Information Systems and Technologies (CISTI), Barcelona, Spain, June 18–21, 2014. pp. 1–6. doi: 10.1109/CISTI.2014.6876979.
4. Vanitha KS, Uma SV, Mahidhar SK. Distributed denial of service: attack techniques and mitigation. In: 2017 International Conference on Circuits, Controls, and Communications (CCUBE), Bangalore, India, December 15–16, 2017. pp. 226–231. doi: 10.1109/CCUBE.2017.8394146.
5. Nishanth N, Mujeeb A. Modeling and detection of flooding-based denial-of-service attack in wireless ad hoc network using Bayesian inference. IEEE Syst J. 2020; 15 (1): 17–26.
6. Elsaeidy AA, Jamalipour A, Munasinghe KS. A hybrid deep learning approach for replay and DDoS attack detection in a smart city. IEEE Access. 2021; 9: 154864–154875.
7. Li Z, Zhang H, Shahriar H, Lo D, Qian K, Whitman M, Wu F. Denial of service (DoS) attack detection: performance comparison of supervised machine learning algorithms. In: 2020 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Calgary, Alberta, Canada, August 17–22, 2020. pp. 469–474.
8. Fan C, Kaliyamurthy NM, Chen S, Jiang H, Zhou Y, Campbell C. Detection of DDoS attacks in software defined networking using entropy. Appl Sci. 2022; 12 (1): 370. doi: 10.3390/app12010370.
9. Daneshgadeh S, Kemmerich T, Ahmed T, Baykal N. Online DDoS attack detection using Mahalanobis distance and kernel-based learning algorithm. J Netw Computer Appl. 2020; 168: 102756. doi: 10.1016/j.jnca.2020.102756.
10. Mladenov B, Iliev G. Searching for optimal software defined network controller against DDoS attacks. In: Proceedings of the 2020 International Symposium on Networks, Computers and Communications (ISNCC), Montreal, Quebec, Canada, October 20–22, 2020. pp. 1–4.

© STM Journals 2024. All Rights Reserved

11. Sanjeetha R, Benoor P, Kanavalli A. Mitigation of DDoS attacks in software defined networks at application level. In: Proceedings of the 2019 PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS), Bangalore, India, August 18, 2019. pp. 1–3.
12. Yadav SK, Suguna P, Velusamy RL. Entropy based mitigation of distributed-denial-of-service (DDoS) attack on control plane in software-defined network (SDN). In: Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, July 6–8, 2019. pp. 1–7.
13. Pham TND, Yeo CK, Yanai N, Fujiwara T. Detecting flooding attack and accommodating burst traffic in delay-tolerant networks. IEEE Trans Veh Technol. 2018; 67 (1): 795–808.
14. Kumari K, Mrunalini M. Detecting denial of service attacks using machine learning algorithms. J Big Data. 2022; 9 (1): 1-17. doi: 10.1186/s40537-022-00616-0.
15. Vanitha KS, Uma SV, Mahidhar SK. Distributed denial of service: attack techniques and mitigation. In: 2017 International Conference on Circuits, Controls, and Communications (CCUBE), Bangalore, India, December 15–16, 2017. pp. 226–231. doi: 10.1109/CCUBE.2017.8394146.
16. David J, Thomas C. DDoS attack detection using fast entropy approach on flow-based network traffic. Procedia Computer Sci. 2015; 50: 30–36.
17. Gu Y, Li K, Guo Z, Wang Y. Semi-supervised K-means DDoS detection method using hybrid feature selection algorithm. IEEE Access. 2019; 7: 64351–64365. doi: 10.1109/ACCESS.2019.2917532.

© STM Journals 2024. All Rights Reserved