Journal of IoT in Social, Mobile, Analytics, and Cloud (ISSN: 2582-1369)
www.irojournals.com/iroismac/

DDoS Detection using Machine Learning Techniques

R. Amrish[1], K. Bavapriyan[2], V. Gopinaath[3], A. Jawahar[4], C. Vinoth Kumar[5]

Department of Electronics and Communication Engineering,Sri Sivasubramaniya Nadar College of Engineering, Rajiv Gandhi Salai (OMR), Kalavakkam, Tamil Nadu, India

E-mail: [1]amrish18015@ece.ssn.edu.in, [2]bavapriyan18029@ece.ssn.edu.in, [3]gopinaath18049@ece.ssn.edu.in, [4]jawahara@ssn.edu.in, [5]vinothkumarc@ssn.edu.in

Abstract

A Distributed Denial of Service (DDoS) attack is a type of cyber-attack that attempts to interrupt regular traffic on a targeted server by overloading the target. The system under DDoS attack remains occupied with the requests from the bots rather than providing service to legitimate users. These kinds of attacks are complicated to detect and increase day by day. In this paper, machine learning algorithm is employed to classify normal and DDoS attack traffic. DDoS attacks are detected using four machine learning classification techniques. The machine learning algorithms are tested and trained using the CICDDoS2019 dataset, gathered by the Canadian Institute of Cyber Security. When compared against KNN, Decision Tree, and Random Forest, the Artificial Neural Network (ANN) generates the best results.

Keywords: Internet of Things (IoT), DDoS (Distributed Denial of Service), KNN, ANN, Random Forest, Decision Tree, Machine Learning

1. Introduction

The Internet of Things (IoT) is the integration of electronic devices embedded with sensors, software, processing ability, networking capability, and other technologies to operate, actuate, connect, and communicate autonomously with other devices and systems over the internet in order to enhance and enable services in a variety of areas. In applications including smart cities, industry automation, health care, and agriculture, the IoT is envisioned to promote broad sensing and effective resource utilization. Privacy and security are important concerns for a variety of IoT applications due to this rapid expansion, the enormous scale of the network, and the critical nature of the data generated by these devices.

DOI: https://doi.org/10.36548/jismac.2022.1.003
Received: 21.03.2022, received in revised form: 13.04.2022, accepted: 30.04.2022, published: 13.05.2022

The limitations of IoT devices in terms of energy, processing capacity, and memory, exacerbate security and design issues. New effective protocols and security methods that can suit the specifications and requirements of existing as well as new devices are required in order for secure data transmission between a myriad of devices. Distributed Denial of Service (DDoS) attacks is a serious security risk in IoT systems. DDoS makes up for a significant portion of the cyber-attacks that take place. More than 10 million DDoS attacks were observed in 2020 alone, reported NetScout [5]. In the first half of 2021, cybercriminals launched roughly 5.4 million DDoS attacks, up 11% from the same period in 2020. DDoS attacks are anticipated to rise from 7.9 million in 2018 to 15.4 million in 2023, according to forecasts. These figures, as well as the attack traffic and times, demonstrate the seriousness of DDoS attacks hence the need for improved security.

A Distributed Denial of Service (DDoS) assault is an attempt to flood a network or service with internet data in order to disrupt regular traffic. DDoS attacks leverage multiple compromised internet-connected computers and other devices as the source of attack traffic, making them more effective. Computers and other related resources, such as IoT devices, are instances of exploited machinery. In the event of a DDoS attack, the server's bandwidth and connectivity are seriously affected, causing major interruptions in all network services. The primary objective of DDoS attacks is to damage the network and resource availability for authentic users. The network is overloaded beyond its bandwidth capabilities in a malicious flood attack resulting in downtime and the disruption of services. Targets range from banks and health care providers to low-profile public networks.

In a DDoS assault, it's impossible to tell the difference between attack traffic and legitimate traffic because it's so similar. They act very much alike to conventional network packets, but in larger amounts and with a higher concentration on the victim. It's easier to identify and mitigate a malicious attack from a small set of nodes. The number of nodes in a typical DDoS attack is usually pretty high, and the overall behaviour of these nodes reduces the chances of valid requests being fulfilled significantly.

Machine Learning, a subset of Artificial Intelligence, uses algorithms to uncover patterns in data, which are then utilised to create a data driven model. The versatility of machine learning makes it a great option for scenarios where the data is ever changing and the task's difficulty is continually shifting. Machine learning techniques that can detect the maliciousness of the packets are used to combat various sorts of DDoS attacks. DDoS

detection has been demonstrated using machine learning approaches such as KNN, Decision trees and Random forests. In terms of accuracy, deep learning, which involves machine learning and multiple abstraction layers, provides more accurate results from the trained neural network. Deep learning has also improved device capabilities, making it appropriate for IoT devices.

2. Related Work

Saini et al., [7] proposed a machine learning approach that detects DDoS attacks and classifies the type of DDoS attack. The machine learning model is validated using the dataset proposed by Mouhammd et al [1], which contains various types of modern DDoS attacks like HTTP-flood, SIDDoS, smurf and UDP-Flood. Also, the dataset consists of 27 features and 5 different classes. For detecting the DDoS attack a machine learning tool WEKA was used. Four machine learning algorithms like Random Forest, MLP, Naive Bayes, and J48 were implemented. Among the four algorithms, J48 gave the best results compared to other classifiers.

Doshi et al., [4] presented a machine learning approach to distinguish the normal and DDoS attacks from consumer IoT devices. For detecting the DDoS attack, stateless (Inter-packet Interval, Packet size) and stateful features (Bandwidth) were considered. The real-time dataset was obtained using middlebox (Raspberry Pi v3 WiFi access port) and other devices like Wemo smart switch, YI camera, and many other devices and the authors mainly focused on UDP Flood, TCP SYN Flood and HTTP GET Flood classes of DDoS attacks. Five different ML algorithms: KNN, LSVM, Decision tree, random forest and neural networks were tested and validated using the dataset. All five algorithms had an accuracy of above 99% and it also observed that the stateless features outperformed the stateful features. However, this paper has no evaluation of well-known datasets for detecting various types of DDoS attacks.

The use of blockchain to mitigate the DDoS attacks and solutions are discussed by Singh et al. [9]. It compares and examines the existing blockchain-based techniques for defending against DDoS attacks. In particular, it reviews four papers which use blockchain, smart contracts and software defined networks (SDN) for DDoS mitigation. For combating DDoS attacks, all four studies leverage blockchain as the primary component. This paper highlights the various issues present in these techniques such as the use of public blockchains

which are a big security concern. It also highlights the lack of security structures to protect the blockchain from direct attacks. It emphasizes that the blockchain technology is still in its beginnings as far as development and implementation are concerned and therefore, the lack of security measures to protect the blockchain.

Wani et al., [6] discussed DDoS attacks, their impact and the losses incurred by these attacks. They provide a comprehensive and systematic analysis of blockchain based DDoS mitigation techniques. The mitigation techniques presented in the paper are divided into various types such as Software Defined Networking (SDN), Artificial Intelligence, Blockchain and Collaborative platforms. The techniques discussed use machine learning, deep learning, fog computing and SDN along with blockchain to mitigate DDoS attacks. The paper lists out the various drawbacks of these techniques and suggests future directions to improve these techniques. It also emphasizes the challenges on mitigation techniques as most of them are only applicable to specific architectures and the need for scalability as these methods should be effective in real world scenarios.

Manikumar et al., [2] discussed the use of machine learning models and Blockchain to mitigate the DDoS attack. Machine learning techniques are adopted for classifying and detecting the malicious packet and temporarily blocks the blacklisted IP using blockchain. The target server distributes the requests to helper nodes which classify them as malicious or benign requests. For this classification, machine learning techniques are adopted and various classification algorithms are used. If the helper nodes classify a request as a malicious request, it blocks the request and stores the blacklisted IP address on the Ethereum Blockchain along with the timestamp. This is stored in the Blockchain for a threshold time and gets unblocked after that time. This paper has also discussed the machine learning algorithms that can be used for classifying the request as benign or malicious. The dataset used for training is the CICDDoS-2019 [4] which has various types of DDoS attacks. Extra Tree Classifier is used for feature selection. The algorithms selected for the purpose of classification are Decision Tree, KNN and Random Forest Algorithm and have evaluated the accuracy performance of the three algorithms. From the accuracy evaluation, the better classifying algorithm was found to be Random Forest Classification algorithm with an accuracy of 95%. This paper also illustrates blocking of the malicious IP with the Ethereum Blockchain and Smart Contracts. Sharma et al., [8] proposed a methodology for enabling generic security for the IoT network through the use of interest-based and physical-aware

coalitions, deep learning and blockchain. The presented solution addresses the problems of security, energy efficiency, scalability and decentralization. Devices are clustered into coalitions such that devices can communicate only with other devices of the same coalition and each group has a coalition head which is responsible for the coalition formation and maintaining security in that group. This minimizes the number of connections required. A multi-layered neural network is used to classify the nodes. Nodes classified as benign exist on the blockchain.

This is a list of all coalition device connections that have been approved. The transaction must be mined, validated, and recorded in the blockchain before a node may build a genuine link between another device and itself. The paper suggests a new approach for mining, the PoR (Proof of Reliance) algorithm. This algorithm works on the basis of maliciousness of the nodes. Benign nodes are given an easier problem to solve while malicious nodes are given harder problems to solve for mining. This algorithm is more energy efficient than the PoW (Proof of Work) algorithm. Checking is a process performed by coalition heads to assure that malicious nodes are by no means able to mine its transactions. The receiving node is permitted to establish a connection link with the sender only on the condition that the transaction is available on the blockchain. The overall accuracy obtained was 95%.

3. Methodology

3.1 Dataset Description and Pre-processing

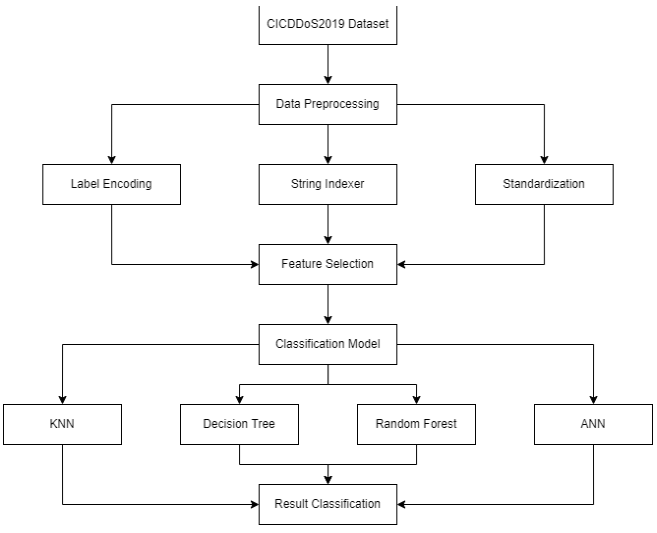

Figure 1. Flowchart for training the classification models

The classification models were trained and tested using the CICDDoS2019 dataset. It comprises of several DDoS attacks such as NTP, DNS, LDAP etc. The dataset contains one class attribute and 88 features that are used to determine if a packet is authentic or malicious. The sample size for training was reduced due to the excessive memory requirements. The steps adopted for training the classification models are shown in Fig 1. The data was pre-processed by removing infinite, empty or missing values and the outliers from the data and then the data was shuffled. The categorical labels were encoded into integer format, where 1 denotes benign and 0 denotes malicious, so that binary classification can be done. The string values were indexed and the data was then standardized to remove any misclassification.

3.2 Feature Selection and Classification

The ExtraTreesClassifier is used for feature selection from the pre-processed data. The ExtraTreesClassifier randomises some decisions and data subsets to reduce overfitting and over-learning. The ExtraTreesClassifier is used to choose the best 15 features from a batch of 88 features. The best 15 features chosen by ExtraTreesClassifier to reduce computational time are depicted in Fig 2. The data containing only the top 15 features is then extracted for training the classification models. Multiple classification models were trained to select the best model that provides the highest accuracy and performance results. The classification models taken into consideration were 1) K-Nearest Neighbours 2) Decision Tree 3) Random Forest and 4) Artificial Neural Network. The artificial neural network is trained by keras library and the classification models based on machine learning is trained by scikit library. The Decision Tree model and the Random Forest model were trained with the default values. The k-value was set to 3 for the KNN classifier model.

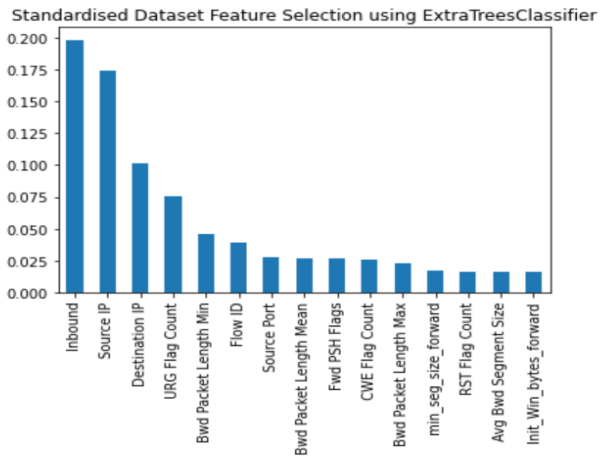

Figure 2. Best 15 Features selected by ExtraTreesClassifier

The neural network used consists of five layers, three dense layers and two dropout layers. The sigmoid activation function and the ReLu activation function are employed in the output layer and other layers, respectively. The optimizer is based on Adam and the loss function is constructed using binary cross entropy to train the model. The batch size was set at 1000, with a total of 40 epochs.

4. Model Analysis and Results

The metrics that were used to compare the machine learning algorithms are given below.

Accuracy: It is the percentage of correctly classified observations to the total number of observations.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Precision: It is defined as the percentage of successfully predicted positive observations compared to all anticipated positive observations.

$$\text{Precision} = TP / (FP + TP)$$

Recall: It is the percentage of successfully predicted positive observations to the total number of observations in the class.

$$\text{Recall} = TP / (FN + TP)$$

F1-score: It is defined as the number of observations that have been erroneously categorized.

$$\text{F1-score} = 2 * TP / (2 * TP + FN + FP)$$

Table 1. Comparison of Various Parameters of the Different Classification Models

Model	TP	TN	FP	FN
KNN	132770	862	116	29
Decision Tree	132886	304	0	587
Random Forest	132886	764	0	127
ANN	132886	826	0	65

where TP, FP, FN and TN denote True Positive, False Positive, False Negative and True Negative respectively which are shown in Table I. The simulation results of the different classification models observed are shown in Table II.

Table 2. Comparison of Various Metrics of the Different Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
KNN	99.89	99.98	99.91	99.94
Decision Tree	99.50	99.56	100	99.78
Random Forest	99.90	99.90	100	99.95
ANN	99.95	99.95	100	99.97

The best performing model was the Artificial Neural Network model with an accuracy score of 99.95%. The Decision Tree model performed the worst with the false negative rate being comparatively higher than the other three models.

One of the noticeable properties of the models were that the false positives were practically non-existent while false negatives were more common although very less in number. This may result in some genuine traffic being blocked but it ensures that no malicious traffic will be misclassified as genuine.

5. Conclusion

In this paper, machine learning algorithm is used for detecting DDoS attacks. The CICDDoS2019 dataset is used which contains 88 features, out of which the best 15 features are extracted. This study is evaluated using four distinct algorithms, including ANN, KNN, Random Forest, and Decision Tree to determine which classification model performs best in detecting malicious IP addresses. It is noted that the ANN model outperforms the other classifier models with an accuracy of 99.95%.

Further, this work can be expanded to develop and deploy a blockchain system to store and blacklist the IP Address of the node, if the traffic is classified as malicious by the machine learning algorithm. The use of the blockchain provides an extra layer of security measure so that the data cannot be tampered with.

References

[1] Alkasassbeh, M.; Al-Naymat, G.; Hassanat, A.B.; Almseidin, M. (2016) “Detecting Distributed Denial of Service Attacks Using Data Mining Techniques.” Int. J. Adv. Comput. Sci. Appl.

[2] D. V. V. S. Manikumar and B. U. Maheswari (2020), ‘‘Blockchain based DDoS mitigation using machine learning techniques’’ in Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA), pp. 794–800.

[3] DdoS Evaluation Dataset (CICDDoS2019). https://www.unb.ca/cic/datasets/ddos-2019.html

[4] Doshi, R., Apthorpe, N., & Feamster, N. (2018). “Machine Learning DDoS Detection for Consumer Internet of Things Devices”. IEEE Security and Privacy Workshops (SPW).

[5] Netscout Systems (2021)."Netscout Threat Intelligence Report". https://www.netscout.com/threatreport

[6] S. Wani, M. Imthiyas, H. Almohamedh, K. M. Alhamed, S. Almotairi and Y. Gulzar (2021), ‘‘Distributed denial of service (DDoS) mitigation using blockchain—A comprehensive insight’’ Symmetry, vol. 13, no. 2, p. 227.

[7] Saini, P. S., Behal, S., & Bhatia, S (2020). “Detection of DDoS Attacks using Machine Learning Algorithms”. 7th International Conference on Computing for Sustainable Global Development (INDIA.Com).pp;16-21,.

[8] Sharma, M.; Pant, S.; Kumar Sharma, D.; Datta Gupta, K.; Vashishth, V.; Chhabra, A (2020). “Enabling security for the Industrial Internet of Things using deep learning, blockchain, and coalitions.” In Transactions on Emerging Telecommunications Technologies; Wiley: Hoboken, NJ, USA; Volume 32, p. e4137.

[9] Singh, R., Tanwar, S., Sharma, T.P. (2020), “Utilization of Blockchain for mitigating the distributed denial of service attacks”. Secur. Priv. 3(3), 1–13.