Swapnil Thatte, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 13, Issue 4, April 2023, pp. 219-224

RESEARCH ARTICLE OPEN ACCESS

Detection of Distributed Denial of Service (DDoS) Attack Using Machine Learning and Deep Learning algorithms

Swapnil Thatte, Anishkumar Iyer, Siya Doshi, Mihir Bhatkar,
Dr. Dashrath Mane
Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai, Maharashtra.

ABSTRACT
Distributed Denial of Service (DDoS) poses a serious challenge to the security of a network. DDoS attacks cripples the victim's network by saturating its infrastructure with a huge number of requests. With the advent of new advanced internet technologies like 5G and B5G in recent years, the frequency and traffic volume of such outbreaks will be an ever-growing problem. This paper aims to evaluate the efficacy of several deep learning & machine learning classifiers in classifying DDoS attacks using attributes from CICDDoS 2017 dataset. We compare the classification metrics like Accuracy, Precision, Recall & F1-score of the classifiers

Keywords - DDoS, Deep Learning, LSTM, MLP, Neural Networks, Machine Learning, Network Security, Ensemble Learning

---

Date of Submission: 07-04-2023 Date of acceptance: 21-04-2023

I. INTRODUCTION

Distributed Denial of Service (DDoS) is a method of attacking system and internet resources. The magnitude and variety of the attacks has increased unimaginably. Numerous distributed agents quickly deplete important resources at the target and make services unavailable to the intended users. They cause network congestion disrupting the regular internet usage and depriving many legitimate consumers of the services and resources.

GitHub, a well-known online code management site used by millions of engineers, was the victim of one of the most powerful DDoS attacks ever recorded. At 1.3 TB/s, this assault transmitted 126.9 million packets per second. GitHub, on the other hand, was utilising a DDoS monitoring service, which alerted them to the attack within 10 minutes of its beginning. Because of this signal, GitHub had the opportunity to immediately stop the assault after initiating the mitigation step

According to the statistics and predictions made by CISCO, there will be a significant increase in DDoS attacks in the near future. Recent studies suggest that by 2024, there will be twice as many DDoS attacks as there were in 2017. Service providers are under considerable threat as a result of the size and volume of DDoS attacks, with the greatest reported attack volume being 1.7 Tb/s. DDoS attacks have recently resulted in downtime costs that have been considerably high, totalling USD 221,836.80. Attacks on firewalls and IPS devices have increased between 2017 and 2018, from 16 to 31 percent. DDoS attacks on cloud-based services have also increased from 11% to 34% during this time.

Various defence mechanisms have been proposed to deal with DDoS attacks amongst which Statistical Methods and Machine Learning models have shown promising results in detecting DDoS attacks.

It has been found that the Machine Learning models perform marginally better than the Statistical models to predict DDoS attacks. The Mathematical Model is 99.75% accurate, whereas the Machine Learning Model has an accuracy of 100%. To evaluate the efficiency of the DDoS attack prediction, ML classifiers like Naive Bayes and Clustering Algorithms are applied, although Logistic Regression produced superior results to Naive Bayes.

II. BACKGROUND AND RELATED WORK

This section elaborates the recently developed DDoS. Detection Systems that have been

employed using various data sets for their respective case study.

2.1 DDoS Attacks

A Distributed Denial of Service (DDoS) attack is a cyber-threat where an attacker floods a network or website with a high volume of traffic, overloading the targeted system and rendering it inaccessible to normal users.

The attack is called "distributed" because it is launched from several sources, most commonly a botnet or a network of hacked machines. This makes identifying and blocking the source of the attack challenging for the targeted system. DDoS attacks can be classified into various types, including:

1) Volume-Based Attacks - These attacks include flooding the targeted system with a large amount of traffic, often employing amplification or reflection techniques.

2) Protocol-Based Attacks - These attacks exploit vulnerabilities in network protocols such as TCP or UDP, leaving the targeted machine inoperative.

3) Application-layer attacks - These attacks aim to overrun or crash a system by targeting individual apps or services, such as a web server.

2.2 Machine Learning in DDoS Attacks

Machine Learning is a branch of Artificial Intelligence. Recently, Machine Learning has acquired wide recognition in helping to detect DDoS attacks. The two most significant steps in Machine Learning to predict a DDoS attack are Feature Extraction and Model Training. In Model Training, the essential features extracted are used as input to train the model, and to make predictions. Machine Learning can be classified into unsupervised learning and supervised learning approaches. In unsupervised learning, the ML algorithm learns from the input data and finds trends and patterns by itself. Whereas, in supervised learning, the model learns and understands patterns using the data set's features.

2.3 Related Work

In [1], It is recommended to use a mathematical model for distributed denial-of-service attacks detection. To distinguish between assaults and regular events, ML classifiers like Naive Bayes and Logistic Regression are used. In this study, the WEKA data mining tool is used and its findings are assessed. This study's conclusion states that machine learning models outperform mathematical models by a little margin.

In [5], a NS2 network simulator is used to construct a clean DDoS data set. The data set used in this paper, has been used to construct several attacks, targeted at the Network layer. Random Forest, Naive Bayes and MLP are applied on the obtained data set to categorise the DDoS as SIDDOS, HTTP-Flood, UDP-Flood and Smurf. MLP had the highest accuracy

Paper [8] utilises a Machine Learning tool, the WEKA, to recognize and describe DDoS attacks. The data set from [5], which has 27 features and 5 distinct classes, is being used. The attacks from the data set were classified using J48, MLP, Random Forest, and Naive Bayes. It was discovered through analysis of the output that the J48 classifier outperformed the other classifiers.

In [3], a TCP-flood was performed using a specially built system to attack the targeted server. Wireshark was used to capture and analyse the internal traffic both before and during the attack. An attack was discovered 3 seconds after the attacker simulated the attack.

III. METHODOLOGY USED

3.1 Data Analysis

We have used the CICDDoS2017 dataset for the purpose of training our models, so as to compare their accuracy and efficiency.

The CICDDoS2017 dataset is a set of network traffic statistics that were produced in a lab setting to model DDoS attacks. The Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick created the dataset, which has over 14 million instances of network traffic, including both legitimate and malicious activity.

CICDDOS2017 is an open-source data set which was used to train various ML and DL models. The dataset has 78 features and two labels namely DDoS and Benign. The data set has a total of 225,745 data points. For researchers and analysts researching DDoS attacks, the network traffic is a useful resource because each instance of traffic is identified with the type of attack that it denotes.

The CICDDoS2017 dataset also has a few extra parameters, like packet size and timing data, that can be used to create more complex models for

identifying and preventing DDoS attacks. This dataset has been extensively used in academic research and has helped to create novel methods for defending against DDoS attacks.

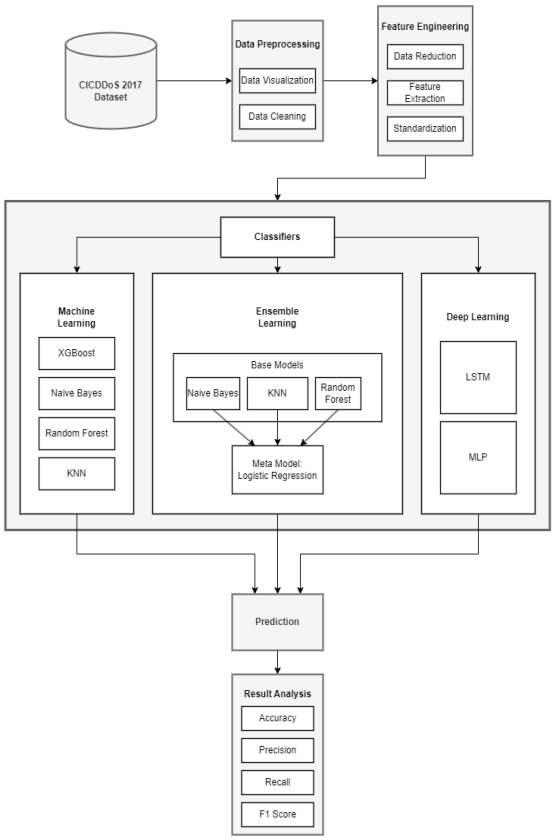

Fig. 1. Methodology

3.2 Machine Learning Approach

3.2.1 Gaussian Naive Bayes (GNB)

The Naive Bayes classification method is a probabilistic classifier, which means it operates on the conditional probability principle. It is based on probability models with high independence assumptions. This type of assumption reduces the model's accuracy since characteristics are occasionally dependent. As a result, they are seen as naïve.

$$P(y|x_1, ..., x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)...P(x_n)}$$

Fig. 2. Naive bayes Formula

In our proposed methodology, we have implemented the Gaussian Naive Bayes algorithm on the selected dataset

3.2.2 K-Nearest Neighbours (KNN)

The k-nearest neighbours algorithm, is a non-parametric, supervised learning classifier that uses Euclidean distance as a similarity measure to group the data points, thereby making a prediction. K value represents the number of nearest neighbours, it can be obtained by appropriate parameter tuning techniques.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

Fig. 3. Euclidean distance formula

In our proposed methodology, we have implemented the KNN classifier with K=10 on the selected dataset

.

3.2.3 Random Forest (RF)

Random Forest is an ensemble learning technique that creates many decision trees, each of which is trained using a different subset of data and set of characteristics. Random Forest makes guarantees that the trees are different and uncorrelated by choosing random subsets of data and attributes for each tree. After each tree has been constructed, they combine to provide the final prediction. This method enhances generalisation, reduces the risk of overfitting, and is resistant to noisy and missing data. For optimum performance, it could necessitate a precise selection of hyperparameters and be computationally expensive.

In our proposed method, we have implemented the Random Forest classifier with max-depth=3 on the selected dataset

3.2.4 XGBoost (XGB)

XGBoost is an open-source library for gradient-boosting decision trees, which is a popular ensemble learning technique that combines multiple weak models to create a strong model. XGBoost uses iterative gradient boosting to add decision trees to a model, such that each new tree focuses on correcting the errors made by the previous trees.

3.2.5 Ensemble Learning Approach

Ensemble learning is a ML approach that combines different models to improve a system's predictive performance. The core idea is to train various models on the same dataset but with different algorithm configurations, such as different hyperparameters or learning algorithms.

Stacked generalisation or Blending is a form of ensemble strategy that comprises Level-0 or multiple Base models whose predictions are then supplied to a Level-1 or Meta-model, which picks the optimal approach to combine those predictions.

We have made use of this Blending technique in which the predictions are made by the base models (Naive Bayes, Random Forest, K Nearest Neighbours) and Logistic regression as Meta model for combining those predictions.

3.3 Deep Learning Approach

3.3.1 Multi-Layer Perceptron (MLP)

MLP, a feedforward NN consisting of several layers of connected neurons. The input data is processed by one or more hidden layers of neurons; each layer computes a weighted sum of the inputs and applies an activation function that results in non-linearity. The output layer, which may include one or more neurons, creates the network's final output.

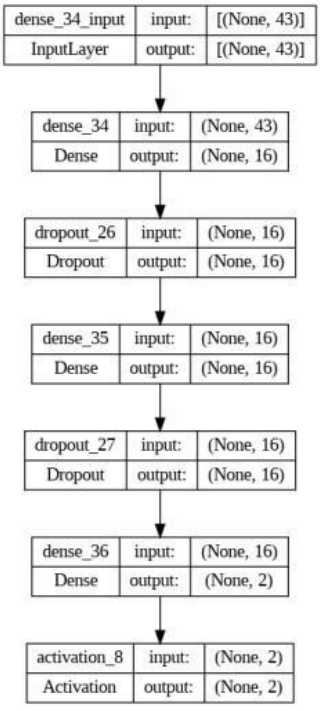

Fig. 4. MLP Architecture

3.3.2 Long Short-Term Memory (LSTM)

LSTM, a sequential network, is a variant of Recurrent NN (RNN) architecture that can process and make predictions on sequential data. To properly handle long-term dependencies in sequential data, LSTM networks feature a special memory cell that

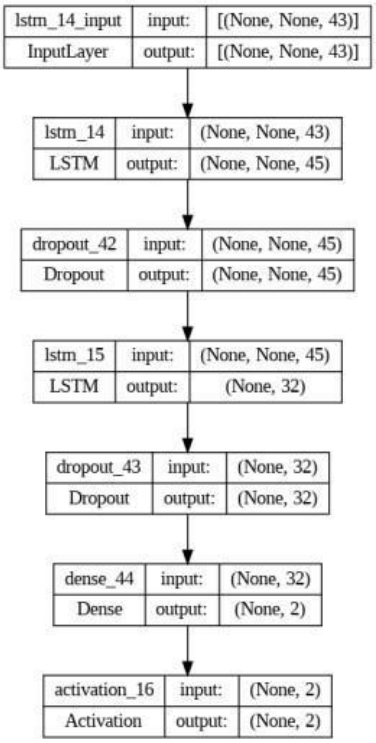

Fig. 5. LSTM Architecture

can selectively recall or forget information depending on the input it gets.

3.3.3 Loss Function Used: Categorical Cross Entropy

Used for Multi-class classification problems where the output is a single integer representing the predicted class from a collection of classes that are mutually exclusive, with the objective of minimising the cross-entropy between the predicted and true distributions.

IV. EXPERIMENTS AND RESULTS

Firstly, the performance of each individual ML algorithm was evaluated on the dataset. As mentioned in the methodology, the following algorithms were used: KNN, RFC, GNB, XGB. Among these algorithms, the LSTM and Ensemble model achieved the highest accuracy rates of 99.87% and 99.85%, respectively.

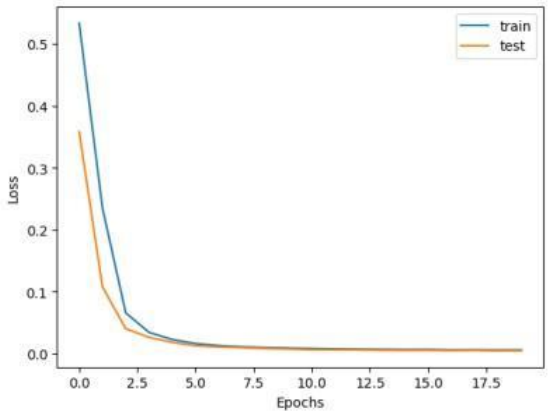

Fig. 6. LSTM Loss vs Epochs

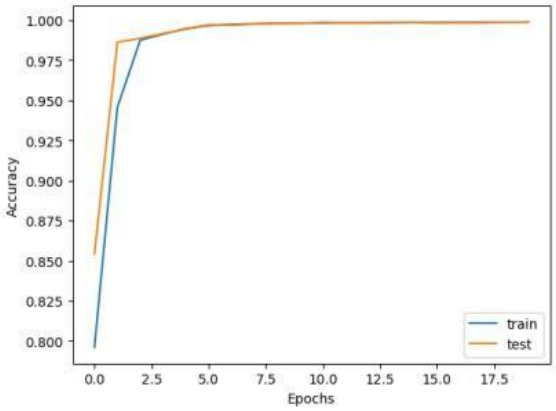

Fig. 7. LSTM Accuracy vs Epochs

Multilayer Perceptron achieved accuracy of 99.83% on the CICDDoS 2017 dataset.

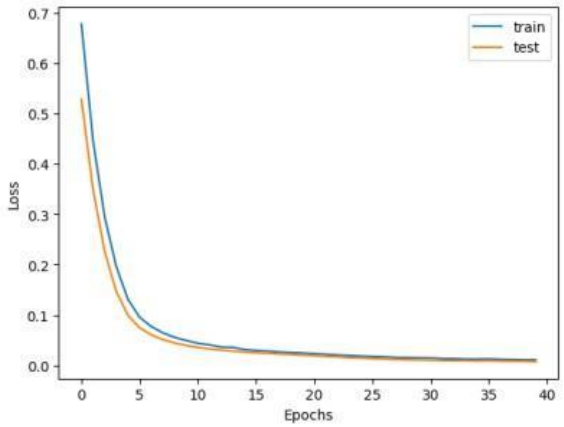

Fig. 8. MLP loss vs. Epoch

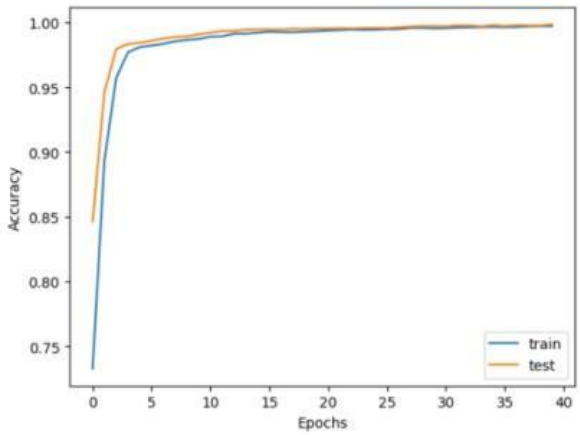

Fig. 9. Accuracy vs Epochs

The below graph gives graphical representation performance evaluation metrics of various algorithms.

Algorithm	Colour
MLP	
LSTM	
XGBoost	
ENSEMBLE MODEL	
KNN	
RANDOM FOREST	

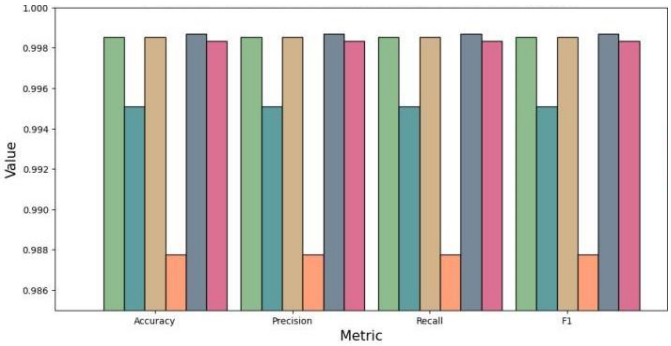

Fig. 10. Metrics of all classifiers

The table below shows the evaluation metrics of various algorithms in tabular format with each value rounded off to 3 decimal places.

Model	Accuracy	Precision	Recall	F1
LSTM	99.870	99.859	99.870	99.867
ENS	99.852	99.849	99.852	99.849
RFC	99.852	99.840	99.852	99.849
MLP	99.834	99.835	99.834	99.831
XGB	98.773	98.627	98.773	98.756
KNN	99.507	99.525	99.507	99.498
GNB	80.889	87.320	80.889	78.680

V. CONCLUSION

In conclusion, this study used machine learning, ensemble learning, and DL techniques on the CICDoS 2017 dataset to improve the detection of DDoS attacks. The findings of this study showed that ensemble learning techniques were more effective than individual machine learning models in identifying DDoS attacks. In addition, compared to other algorithms examined in this study, the DL model was found to be most accurate at identifying DDoS attacks.

The LSTM neural network displayed the highest accuracy of 99.870% followed by an ensemble learning algorithm with accuracy of 99.852%.

REFERENCES

[1]. Kumari, K., Mrunalini, M. Detecting Denial of Service attacks using machine learning algorithms. J Big Data 9, 56 (2022).

[2]. Shriram Rajesh, Marvin Clement, Sooraj S. B., Al Shifan S. H, Jyothi Johnson - Real Time DDoS Attack Detection Based on Machine Learning Algorithms. The International Conference on Emerging Trends in Engineering, Govt. Engineering College Kozhikode, 24th September 2021

[3]. A. Ramzy Shaaban, E. Abdelwaness and M. Hussein, "TCP and HTTP Flood DDOS Attack Analysis and Detection for space ground Network," 2019 IEEE International Conference on Vehicular Elec-tronics and Safety (ICVES), Cairo, Egypt, 2019, pp. 1-6, doi:10.1109/ICVES.2019.8906302.

[4]. Suresh M., Anitha R. (2011). Evaluating machine learning algorithms for detecting DDoS attacks. Communications in Computer and Information Science, 196 CCIS, 441–452.

[5]. Alkasassbeh, M., Al Naymat, G., Hassanat, A. B. & Almseidin, M. (2016). Detecting Distributed Denial of Service Attacks Using Data Min-ing Techniques. International Journal of Advanced Computer Science and Applications, 7(1):436–445.

[6]. Alzahrani R. J., Alzahrani A. (2021). Security analysis of ddos attacks using machine learning algorithms in network traffic. Electronics (Switzerland), 10(23).

[7]. Sambangi S., Gondi L. (2020). A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression. 51.

[8]. P. S. Saini, S. Behal and S. Bhatia, "Detection of DDoS Attacks using Machine Learning Algorithms," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 16-21, doi: 10.23919/INDIACom 49435.2020.9083716.

[9]. Jiangtao Pei, Yunli Chen, Wei Ji - A DDoS Attack Detection Method Based on Machine Learning. Jiangtao Pei et al 2019 J. Phys.: Conf. Ser.1237 032040

[10]. Sambangi S., Gondi L. (2020). A Machine Learning Approach for DDoS(Distributed Denial of Service) Attack Detection Using Multiple Linear Regression. 51.

[11]. Mona Alduailij, Qazi Waqas Khan, Muhammad Tahir, Muhammad Sardaraz, Mai Alduailij, Fazila Malik (2022). Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method. Symmetry, 14(6), 1095.

[12]. Hriram Rajesh, Marvin Clement, Sooraj S. B., Al Shifan S. H, Jyothi Johnson - Real Time DDoS Attack Detection Based on Machine Learning Algorithms. The International Conference on Emerging Trends in Engineering, Govt. Engineering College Kozhikode, 24th September 2021