LIGHTWEIGHT MACHINE LEARNING MODEL FOR CLOUD DDoS DETECTION USING EARLY TRAFFIC FEATURES

Dr. S. Porkodi
Assistant Professor, Department of Computer Applications
Score Vellore Institute of Technology, Vellore, Tamilnadu, India
porkodi.s@vit.ac.in

DOI: doi.org/10.34293/shanlax.9789361631474.ch018

Abstract

In the recent years, Distributed Denial of Service (DDoS) attack still remains threat in the cloud environment due to rapid increase of big data. Features of the minimal network flow is analysed for early detection and reduction of the response time. This paper aims to develop a lightweight multi class DDoS detection model using Light Gradient Boosting Machine (LightGBM) classification. Basic transport layer features and flow level features such as port number, protocol type, package count, handshake time, etc are only focused and other features which will introduce the data leakage when developing the model are deliberately avoided. The developed classifier model acquires overall accuracy of 89.07% when classifying across attack, benign and suspicious traffic classes. Even with a limited dataset the model effectively detects the DDoS attack making the system effective for the real time cloud security applications.

Keywords: DDos attack detection, Cloud Security, Network Traffic Analysis

1. Introduction

Due to the high availability and scalability traits of the cloud environment, it is targeted with DDoS attack [3]. The attackers aims to exhaust the network and computational resources in turn leading to disruption of services ultimately resulting in financial losses. The traditional Intrusion Detection System (IDS) won't be much suitable for real time deployment [4] and DDoS detection since it relies on the extensive statistical feature analysis and deep packet inspection which will be more expensive on computation aspects. In recent days, machine learning [2] and artificial intelligence models [1] have demonstrated better results in DDoS detection [5]. Yet, many research works employ the usage of inter arrival statistics, packet flags, percentage of the flow level which leads to data leakage and overfitting which in turn resulting in unrealistically high accuracy. This work mainly focuses on features of network flow that are in the early stage of the connection. A simple light weigh multi class classification model is built using LGBM to differentiate attack, benign and suspicious traffic of the network. The proposed system ensures robustness and suitability for the cloud environment in the real time analysis.

2. Dataset Collection

BCCC-cPacket-Cloud-DDoS-2024 which is a standard cyber security dataset, published by York University, Canada is collected for the experimentation. The dataset consists of

traffic captured on 4 days of the network consisting of approximately 85000 network flows in each day with 324 features. The labelled dataset consists of network flow records collected in the cloud during attack and benign scenarios where for each flow, timings, packet level features and protocol related features are listed. The labelled dataset is categorized into attack, benign and suspicious flow. To ensure realistic evaluation and eliminate the possibility of creating a biased model, attributes such as timestamp, flow ID, IP address were removed before the training phase and only flow level information is focused which are available at the beginning of the communication.

3. Feature Selection and Preprocessing

Rather than relaying on hundreds of engineered features that are preferred in the existing works, the proposed work deliberately restricts the features to a small set of early stage attributes of the network flow. Thus, the selected features are,

- Flow duration
- Protocol type
- Total packet count
- Handshake duration
- Source and destination port numbers
- Forward and backward packet counts

All these listed features can be acquired during the initial stage of the network communication and need not require full flow completion. The categorical attributes like, protocol type are label encoded whereas missing data and invalid entries are handled by assigning zero imputation. All there ultimately results in reducing the computational complexity, minimizing the risk of overfitting and develop much reliable DDoS attack detection in the real time systems.

4. Proposed Model

LightGBM which is based on the decision tree boosting is one of the efficient ensemble learning model which is very much suitable for handling big data. Since the model has the capacity to train faster while utilizing much low memory usage. In the proposed work, LightGBM acts as a multi classifier to classify attack, benign and suspicious traffic of the network flow. The dataset is split into 80-20 for training and testing respectively whereas, stratified sampling is applied to the dataset for preserving the distribution of the classes. Hyper parameters such as, subsampling, column sampling, max depth are tuned to improve the detection accuracy of the model whereas the overfitting is also prevented resulting in a generalized model. Key hyperparameters includes, Max depth = 6, learning rate = 0.05, subsample = 0.08, columnsample_bytree = 0.8

5. Experimental Results

The evaluation of the proposed work is done by calculating the precision, recall, F1=-score and accuracy of the model which are presented in Table 1.

Table 1. Eevaluation metric of the proposed model

Class	Precision	Recall	F1-Score	Support
Attack	0.8206	0.9094	0.8628	45,694
Benign	0.9307	0.9458	0.9382	82,640
Suspicious	0.9261	*0.4336*	0.5906	11,821
Accuracy			**0.8907**	140,155
Macro Avg	0.8925	0.7629	0.7972	140,155
Weighted Avg	0.8944	0.8907	0.8843	140,155

The proposed LGBM model yields an overall accuracy of 89.07% and benign traffic is classified better with high recall value. The misclassification occurs only between suspicious and attack classes due to the overlapping traffic characteristics and is expected. The confusion matrix in Table 2 indicates that the model does not have any data leakage or overfitting issues. These results ensure that the proposed work is suitable for detecting DDoS attack with the early stage features of the network flow.

Table 2. Confusion matrix of the proposed multi class LGBM model

	Predicted: Attack	Predicted: Benign	Predicted: Suspicious	Total Actual
Actual: Attack	**41,556**	3,823	315	45,694
Actual: Benign	4,384	**78,162**	94	82,640
Actual: Suspicious	4,698	1,998	**5,125**	11,821

6. Conclusion

LIghtGBM multi classifier is proposed in this work to detect the DDoS attack by only relaying on the early stage features of the network flow. The features are deliberately restricted to transport layer attributes and the proposed LGBM model emerges as a strong classifier in differentiating attack, benign and suspicions network flow. The proposed model is safe from data leakage and avoids overfitting issues ensuring that it is much suitable for real time DDoS attack detection in cloud environment.

Reference

1. Iseal, S. (2025). AI for Detecting and Mitigating Distributed Denial of Service (DDoS) Attacks in Cloud Networks.
2. Alazmi, A. N. D., & Alharbi, Y. O. (2025, April). Classification-Based Machine Learning For Detection of DDoS Attack In Cloud Computing. In *2025 4th International Conference on Computing and Information Technology (ICCIT)* (pp. 210-214). IEEE.
3. Pimpalkar, A. S., Akshansh, S., Gour, A., Junankar, M., & Ghule, A. (2025). Detection and Mitigation Techniques for Defending DDoS Attacks in Cloud Environment. *Resilient Community Microgrids*, 303-314.
4. Terawi, N., Ashqar, H. I., Darwish, O., Alsobeh, A., Zahariev, P., & Tashtoush, Y. (2025). Enhanced detection of intrusion detection system in cloud networks using time-

aware and deep learning techniques. *Computers*, *14*(7), 282.

5. Fu, X., Lou, S., Zheng, J., Chi, C., Yang, J., Wang, D., ... & Zhu, X. (2025). Deep learning techniques for DDoS attack detection: Concepts, analyses, challenges, and future directions. *Expert Systems with Applications*, *291*, 128469.