Classification and Prediction Techniques for DDoS Attack Using Machine Learning

[1]Swetha G, [2]Selva Lakshmi B, [3]Priya Dharshini S

Student
Cyber Forensics and Information Security
Dr. M.G.R. Educational and Research Institute, Chennai, India.

Abstract - **Distributed network attacks are commonly referred to as Distributed Denial of Service (DDoS) attacks. These assaults exploit specific constraints that pertain to every arrangement asset, such as the framework of the authorized organization's website. This project proposes a machine learning approach for DDoS attack type classification and prediction. The classification algorithms GBC and MLP are employed in this project's work. StandardScaler is used to pre-process the datasets. StandardScaler removes the mean and scales the data to the unit variance. For the purpose of identifying the performance of the model, this proposed project produced a confusion matrix. The GBC classifier algorithm is utilized in the first classification for both Precision (PR) and Recall (RE). In the second classification, the MLP classifier technique is used to classify both Precision (PR) and Recall (RE). This project is implemented using python software.**

INTRODUCTION

DDoS assaults on a daily basis, thus stopping them necessitates further study and exploration of the issue. Algorithms for machine learning are frequently used to security issues as well as several other uses. The primary objective of utilizing a classification algorithm in a DDoS detection system is to recognize and categorize DDoS attack-related requests among regular traffic. In actuality, a variety of factors influence accuracy and model training duration. For example, the performance is influenced by the classification algorithm selection. Both the model's accuracy and training time are directly impacted by the amount of the dataset. In addition to choosing the best features, many researchers think about shrinking the size of the dataset and, as a result, shortening the time needed for model training. Another element that affects performance and the delay in model training is the setting of the parameters in machine learning algorithms. Each application will experience these elements' effects differently. Accuracy is one of the most, if not the most, essential requirements for all applications. However, for some applications, low training time may be required, even be critical, as identifying DDoS requests with the normal traffic may under time constraints to be useful. Obviously, a trade-off exists between higher accuracy and shorter training times

OBJECTIVES

1. **High Prediction Accuracy:** Evaluate the model performance using metrics such as Precision, recall, F1 score and accuracy to access their ability to correctly classify instances of DDoS attacks.
2. **Hyperparameter Tuning:** Optimize hyperparameters of GBC and MLP algorithms to improve their predictive performance.
3. **Scalability and Robustness:** Access the scalability of the models to handle varying levels of network traffic. Enhance the robustness of the models to adapt to evolving DDoS attack strategies and patterns.

EXISTING SYSTEM

The technologies used in the existing system are KNN Classifier algorithm and DNN Classifier algorithm. This proposed work uses a machine learning method to classify and predict the different kinds of DDoS attacks. The initial stage of data preprocessing involves an analysis of the input dataset. The datasets are pre-processed using StandardScaler. The dataset's useless data is handled using data pre-processing techniques. Following data pre-processing, the stage of data analysis includes researching the observed data. For the data, the following machine learning classifier technique is allowable. The KNN classifier algorithm is used in the first classification for both Precision (PR) and Recall (RE). Finally, the data are evaluated to achieve a prediction output.

PROBLEM STATEMENT

For both labeled and unlabeled datasets, it is crucial to advance from unsupervised learning to supervised learning. Furthermore, the identification of DDoS assaults will be impacted by non-supervised learning techniques, particularly when non-labeled datasets are taken into consideration.

PROPOSED SYSTEM

- **DESIGN METHODOLOGY**

In this project, machine learning method is used to classify and predict the different kinds of DDoS attacks is proposed. The initial stage of data pre-processing involves an analysis of the input dataset. The dataset's useless data is handled using pre-processing techniques. Following data pre-processing, the stage of data analysis includes researching the observed data. For the data, the following machine learning classifier technique is allowable. The machine learning classifier technique uses the Gradient Boosting Classifier algorithm (GBC). The GBC is used in the first classification for both Precision (PR) and Recall (RE). In the second classification, the MLP classifier technique is used to classify both Precision (PR) and Recall (RE). Finally, the data are evaluated to achieve a prediction output.

- **ARCHITECTURE OF PROPOSED SYSTEM**

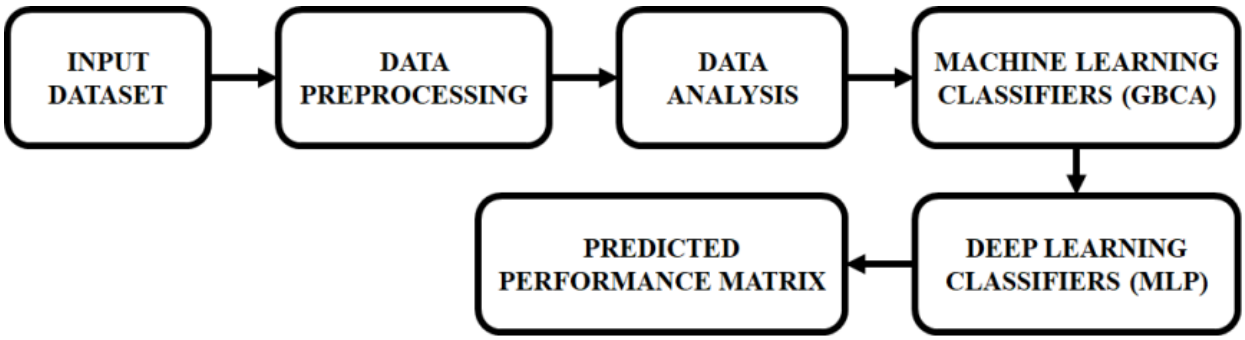

Figure:1 Represents the architecture of proposed system

MODULES

- **DATA PREPROCESSING**

Any kind of processing done on raw data to get it ready for another data processing step is referred to as data preprocessing, which is a part of data preparation. Preprocessing the data is crucial to enhancing its quality. During preprocessing, the computer receives the gathered datasets as input and is trained appropriately.

- **STANDARDSCALER**

In Deep Learning, StandardScaler is used to resize the distribution and normalize values.

- **DATA ANALYSIS**

The process of methodically using statistical approaches to summarize, assess, and describe data is known as data analysis. Data analysis is the process of taking usable information out of data and using that knowledge to guide a decision

- **GRADIENT BOOSTING CLASSIFIER**

The powerful boosting method known as the Gradient Boosting Classifier turns several weak learners into one powerful one. With the use of gradient decent, this approach is trained to minimize the loss function of the prior model, such as mean squared error or cross-entropy. The residual errors from the previous model are used as labels in the predictor's training. After that, the new model's prediction is included in the ensemble, and the procedure is continued until a stopping requirement is satisfied.

- **MULTI LAYER PERCEPTRON**

It is a feed forward artificial neural network model that maps input sets to output proper sets. A multilayer perceptron is made up of multiple layers, each of which is connected to the next. Each node is a processing element or a neuron that has a non-linear activation function except the input nodes. It employs back propagation, a supervised learning technique, to train the network.

- **PREDICTED PERFORMANCE MATRIX**

The module predicts the classified datasets and evaluates the values of accuracy, recall, time, F1 score, precision and confusion matrix. Finally compared the performance of the two classification methods.

APPLICATIONS

- Cloud and Digital Transformation.

- RAN Monitoring and Troubleshooting.
- Cable/ MSO and Fixed Networks.
- Internet Of Things (IoT).

CONCLUSION

This project proposes a machine learning method to classify and predict the different kinds of DDoS attacks. Gradient Boosting Classifier algorithm (GBC) classification is a typical machine learning algorithm. The Deep Learning technique includes the classification algorithm Multi-Layer Perceptron (MLP). Deep Learning techniques are used to aid in the detection and categorization of Distributed Denial of Service (DDoS) assaults. Multi-Layer Perceptron (MLP) is an artificial neural network feed-forward model that maps input sets to output sets. This suggested project created a confusion matrix to determine the performance of the model. Gradient Boosting Classifier (GBC) algorithm classifications achieve a high accuracy performance of 99.69%.

REFERENCES:

1. K. S. Sahoo et al., "An Evolutionary SVM Model for DDOS Attack Detection in Software Defined Networks," in IEEE Access, vol. 8, pp. 132502-132513, 2020.
2. Lau P, Wang L, Liu Z, Wei W and Ten C W , A coalitional cyber-insurance design considering power system reliability and cyber vulnerability, IEEE Transactions on Power Systems, vol. 36, no. 6, pp.5512-5524, 2021.
3. W. Zhijun, X. Qing, W. Jingjie, Y. Meng and L. Liang, "Low-Rate DDoS Attack Detection Based on Factorization Machine in Software Defined Network," in IEEE Access, vol. 8, pp. 17404-17418, 2020.
4. Wisanwanichthan T and Thammawichai, M, A double-layered hybrid approach for network intrusion detection system using combined naive bayes and SVM, IEEE Access, vol. 9, pp.138432-138450, 2021.
5. J. A. Pérez-Díaz, I. A. Valdovinos, K. -K. R. Choo and D. Zhu, "A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning," in IEEE Access, vol. 8, pp. 155859-155872, 2020.
6. Ferrag, Mohamed Amine, Leandros Maglaras, Ahmed Ahmim, Makhlouf Derdour, and Helge Janicke, Rdtids: Rules and decision tree-based intrusion detection system for internet-of-things networks, no.3, 2020.
7. N. Agrawal and S. Tapaswi, "Defense Mechanisms Against DDoS Attacks in a Cloud Computing Environment: State-of-the-Art and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3769-3795, Fourthquarter 2019.
8. K. Hong, Y. Kim, H. Choi and J. Park, "SDN-Assisted Slow HTTP DDoS Attack Defense Method," in IEEE Communications Letters, vol. 22, no. 4, pp. 688-691, 2018.