Denial of Service (DoS) Attack Detection Using Machine Learning

Taha Shah (21K-4930), Aimal Amir (21K-3339), Jaleel Abbas (21K-3383)

***Abstract*—Denial of Service (DoS) attacks are malicious attempts to disrupt network services or systems, rendering them inaccessible to legitimate users. This research applies machine learning techniques, including Gradient Boosting, Support Vector Machine, Naive Bayes, and K-Nearest Neighbors, to detect and mitigate DoS attacks. Gradient Boosting and Support Vector Machines demonstrated superior accuracy and efficiency compared to other models, making them suitable for real-world implementation.**

***Index Terms*—Denial of Service, Cybersecurity, Machine Learning, Gradient Boosting, Naive Bayes, Decision Tree, Random Forest**

I. Introduction

The increasing frequency and sophistication of Denial of Service (DoS) attacks pose significant challenges to network security. These attacks disrupt normal operations by overwhelming systems with illegitimate requests, rendering them inaccessible to legitimate users. Traditional detection methods often fail to adapt to the evolving nature of these threats. Machine Learning (ML) offers a promising solution by automating detection processes and improving accuracy. This paper explores ML-based approaches to detect and mitigate DoS attacks effectively.

II. Motivation

High-profile incidents, such as the 2.3 Tbps attack on AWS in 2020, underscore the critical need for robust detection mechanisms. Attackers exploit TCP/UDP protocols and IP spoofing, making it challenging to distinguish between legitimate and malicious traffic. With the rapid increase in network activity and vulnerabilities, it has become essential to develop automated solutions to identify and neutralize DoS threats effectively.

III. Methodology

The methodology involves leveraging Python-based frameworks for data preprocessing, model training, and evaluation. The process is divided into the following components:

A. Dataset

The dataset, sourced from Kaggle, comprises 346,869 instances with 78 features. The data was filtered to focus on two classes: ”Benign” (normal traffic) and ”DoS Slow Loris” (attack traffic), resulting in 214,475 instances. Data splits of 80:20, 75:25, and 70:30 were used to train and evaluate the models under different conditions.

B. Data Preprocessing

To prepare the data for machine learning, missing values were imputed using column-wise mean values. Categorical data were converted to numerical formats using label encoding, ensuring compatibility with ML algorithms. Redundant features were removed to enhance model efficiency and performance.

C. Machine Learning Models

The following machine learning models were employed for detection:

- **Gradient Boosting:** Combines multiple weak learners to form a strong predictive model. Each iteration reduces error, making it highly accurate for classification tasks.
- **Naive Bayes:** A probabilistic classifier based on Bayes' theorem, effective for high-dimensional data and text classification problems.
- **Decision Tree:** A tree-structured model where decisions are made at nodes based on feature values, providing interpretability and simplicity.
- **Random Forest:** An ensemble of decision trees that aggregates their predictions, reducing overfitting and improving accuracy.

IV. Figures

Figures below illustrate the working mechanisms of key ML models:

A. Gradient Boosting

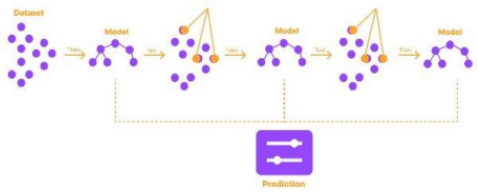

Fig. 1. Visualization of Gradient Boosting Process

B. Decision Tree

C. Naive Bayes

D. Random Forest

V. Results

The proposed system demonstrated significant improvements in detecting DoS attacks. Gradient Boosting achieved

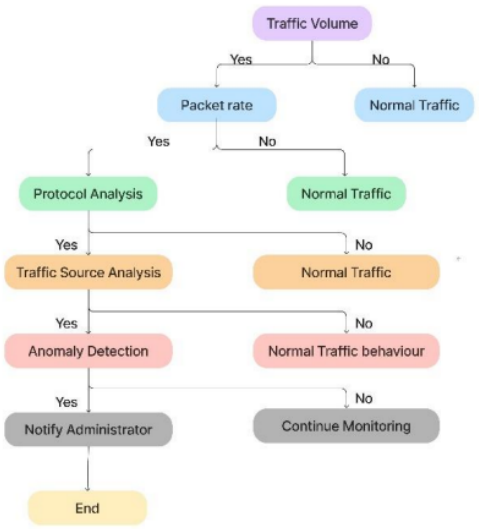

Fig. 2. Structure of a Decision Tree

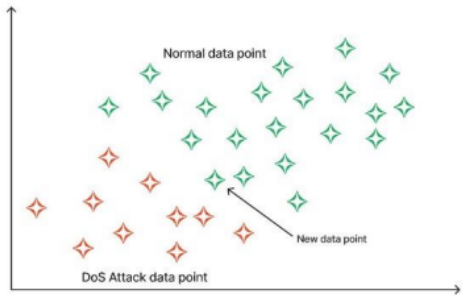

Fig. 3. Naive Bayes Classifier Workflow

the highest accuracy, outperforming other models in datasets with slightly imbalanced distributions. The Random Forest model provided robust results, reducing overfitting tendencies, while Naive Bayes and Decision Trees offered acceptable performance with faster computation times.

VI. Conclusion

This study emphasizes the efficacy of machine learning in addressing cybersecurity challenges posed by DoS attacks. The Gradient Boosting and Support Vector Machine models proved highly effective, demonstrating superior accuracy and efficiency. Future work will focus on incorporating advanced algorithms and exploring the integration of blockchain technology to further enhance system security. Extending the system to detect other types of cyberattacks, such as DDoS and Man-in-the-Middle attacks, will also be explored.

References

[1] Y. Mirsky et al., "Vesper: Using echo analysis to detect man-in-the-middle attacks in LANs," IEEE TIFS, vol. 14, no. 6, pp. 1638–1653, 2019.

[2] P. S. Saini et al., "Detection of DDoS attacks using machine learning algorithms," INDIACom, pp. 16–21, 2020.

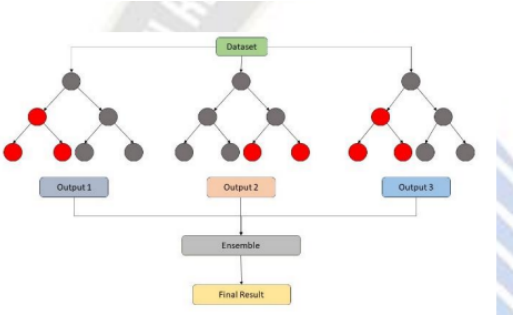

Fig. 4. Random Forest Model Workflow

[3] J. Cheng et al., "DDoS Attack Detection Using IP Address Feature Interaction," Proc. of 2009 International Conference on Intelligent Networking and Collaborative Systems, pp. 113-118.