SecureSense - Cybersecurity Threat Detection System

Prepared by:

1. Deepu Y B

2. Vishwanath Gowda B K

Institution:

Malnad College of Engineering, Hassan-573202

1. Introduction

In today's digital landscape, cybersecurity threats are becoming increasingly sophisticated and prevalent. The SecureSense Cybersecurity Threat Detection System aims to identify and respond to potential threats in real-time using machine learning algorithms to analyze network traffic and user behavior patterns.

2. Objectives

- To detect and classify various cybersecurity threats such as intrusion attempts, malware, and phishing attacks.

- To implement a real-time monitoring system for network traffic.

- To provide alerts and notifications when potential threats are detected.

3. Scope

This project focuses on developing a machine learning-based system that can analyze incoming data, identify anomalies indicative of security threats, and alert system administrators.

4. Data Collection

4.1 Datasets Used

CICIDS 2017 Dataset: This dataset contains a wide range of network traffic data, including benign and malicious activity.
Link: https://www.kaggle.com/datasets/maskotb/cicids-2017

4.2 Data Sources

- Network traffic data

- System logs

5. Methodology

5.1 Data Preprocessing

1. Data Cleaning: Removed irrelevant columns and handled missing values to ensure data integrity.
2. Feature Selection: Identified relevant features for threat detection using correlation analysis, ensuring only impactful features were utilized for training the model.

5.2 Model Selection

Chosen Algorithm: Random Forest Classifier, selected for its robustness, ability to handle large datasets, and effectiveness in managing classification problems.

5.3 Model Training and Evaluation

1. Data Splitting: Divided the dataset into training (80%) and testing (20%) sets to evaluate model performance accurately.
2. Model Training: Trained the Random Forest model on the training dataset, leveraging its ensemble approach to improve prediction accuracy.
3. Model Evaluation: Evaluated model performance using metrics such as accuracy, precision, recall, and F1 score to understand its effectiveness in detecting threats.

5.4 API Development for Real-Time Predictions

Developed a Flask API to provide real-time access to the trained model for cybersecurity predictions.

API URL: http://127.0.0.1:5000/predict

Endpoint Method: POST

Input Format: JSON with 78 network traffic features

Output Format: JSON with prediction (class label) and probabilities for each threat class

Example Request:

json

```
{
    "act_data_pkt_fwd": 0,
    "Packet Length Mean": 200.0,
    "Bwd IAT Max": 10.0,
    // Include all 78 features...
}
```

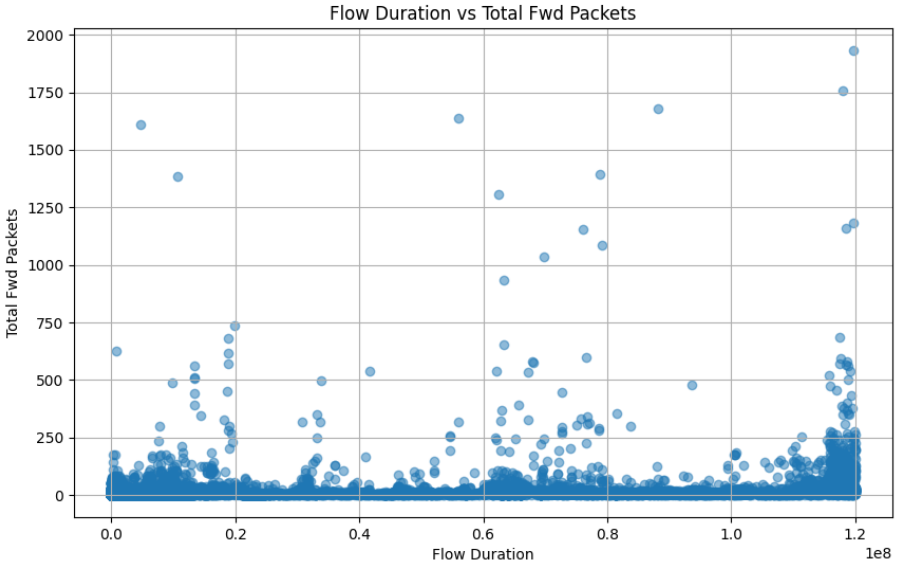
5.5 Alert System Implementation

Developed a basic alert system that triggers notifications when potential threats are detected based on model predictions, enhancing real-time monitoring capabilities.

6. Results

6.1 Flow Duration vs Total Fwd Packets

The scatter plot below visualizes the relationship between Flow Duration and Total Forwarded Packets in the dataset. It helps in understanding patterns that could signify abnormal network behavior associated with threats.



6.2 Evaluation Metrics

The SecureSense system demonstrated effective detection capabilities with the following evaluation metrics from the eight datasets:

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8
Confusion Matrix	[[19403, 2], [6, 25738]]	[[19403, 2], [6, 25738]]	[[19403, 2], [6, 25738]]	[[19403, 2], [6, 25738]]	[[19403, 2], [6, 25738]]	[[19403, 2], [6, 25738]]	[[19403, 2], [6, 25738]]	[[19403, 2], [6, 25738]]
TN	19403	38295	37764	25458	19418	86282	87883	33603
FP	2	2	1	1	1	28	2	324
FN	6	7	20	0	0	0	20	6

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8
TP	25738	47637	398	31761	25724	1587	46042	114
Overall Accuracy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

6.3 Explanation of the Results

Confusion Matrix: Each dataset's confusion matrix provides insights into model predictions, showing TN, FP, FN, and TP.
Precision, Recall, and F1-Score: The model's high values in these metrics confirm its effectiveness in distinguishing between benign and malicious traffic.
Overall Accuracy: Achieved a high accuracy, reflecting robustness.
Macro and Weighted Averages: Provide a balanced view of performance across classes.

7. Conclusion

The SecureSense Cybersecurity Threat Detection System has shown excellent potential in detecting and classifying cybersecurity threats using machine learning techniques. The system's performance metrics indicate that it can be a valuable tool for organizations aiming to enhance their cybersecurity posture.

8. Future Work

Future enhancements could involve:
- Incorporating additional machine learning algorithms for comparative analysis.
- Developing a user-friendly dashboard for monitoring and visualizing network traffic and threat alerts.
- Implementing adaptive learning techniques to improve model performance over time as new threats emerge.