DDoS Protection System for Cloud: Architecture and Tool

Panel No. 06
Supervisor Name
Dr. Balaji Srikaanth P, AP/NWC
Dr. S. Nagendra Prabhu, AP/CINTEL

Batch No. NW000156
Shaurya Singh Srinet – RA2111032010006
Shounak Chandra – RA2111032010026
Charvi Jain – RA2111047010113

Functional Document for User Story 3: Anomaly Detection Implementation

1. Introduction

This user story involves developing an anomaly detection script using the Isolation Forest algorithm. The script will analyse simulated traffic data and detect deviations that indicate potential DDoS attacks. The goal is to build a reliable detection mechanism that identifies anomalies with high precision.

2. Product Goal

To implement an anomaly detection system capable of analysing network traffic data and identifying suspicious patterns indicative of DDoS attacks. The output will be visualized to validate the model's effectiveness.

3. Demography (Users Location)

- **Target Users:** Data scientists, security analysts, AI developers.
- **User Characteristics:** Users experienced in machine learning and anomaly detection frameworks.
- **Location:** Organizations globally, particularly those with cloud-based systems.

4. Business Processes

- **Model Development:**
  - Import and preprocess the labelled dataset.
  - Train the Isolation Forest model using normalized data.
- **Anomaly Detection:**
  - Use the model to predict anomalies in traffic logs.
  - Output results with anomaly scores for further analysis.

- **Visualization**:
  - Generate plots (e.g., anomalies_plot.png) to highlight detected anomalies.
  - Validate detection results through visual inspection.

5. Features

- **Model Training:**
  - Train the Isolation Forest model on the labelled dataset.
  - Optimize hyperparameters to enhance detection accuracy.
- **Anomaly Scoring:**
  - Assign anomaly scores to traffic entries.
  - Classify traffic as normal or anomalous based on threshold values.
- **Visualization:**
  - Generate visualizations to present detected anomalies.
  - Use plots for result validation and debugging.

6. Authorization Matrix

ROLE	**Access Level**
Data Scientist	Access to raw datasets and preprocessing tools.
Developer	Full access to model training and testing scripts.
Security Analyst	Access to detection results and visualizations.

7. Assumptions

- The labelled dataset is accurate and pre-processed.
- The Isolation Forest algorithm is appropriate for the use case.
- Visualization tools are installed and configured correctly.

8. Target Audience

- **Audience:** Security teams, data scientists, cloud solution architects.
- **Effort Estimation:** Approximately 3-4 days for model implementation and testing.

9. Acceptance Criteria

- The Isolation Forest model is successfully trained and tested.
- Traffic entries are scored and classified accurately.

- Visualization plots are generated and reviewed for correctness.

10. Checklist

- Dataset pre-processed and normalized.
- Isolation Forest model trained and optimized.
- Anomaly scores assigned to traffic entries.
- Visualization plots generated and validated.