AI-Driven Dynamic Fuzz Testing for IoT Security

Panel No. 06	Batch No. NW000156
Supervisor Name	Shaurya Singh Srinet – RA2111032010006
Dr. Balaji Srikaanth P, AP/NWC	Shounak Chandra – RA2111032010026
Dr. S. Nagendra Prabhu, AP/CINTEL	Charvi Jain – RA2111047010113

Functional Document for User Story 2: Generate Dataset for GNN Model

1. **Introduction**

   This user story focuses on generating the dataset needed for training the GNN model. The dataset will include traffic data from IoT simulations, comprising both benign and DDoS attack traffic. This data is crucial for developing a robust AI-driven security framework capable of detecting and mitigating DDoS attacks in real-time.

2. **Product Goal**

   The goal is to generate a labeled dataset from the NS-3 simulation environment, which will be used to train the GNN model for identifying and mitigating DDoS attacks in IoT networks.

3. **Demography (Users Location):**

   a. Target Users: AI developers, security analysts, researchers.

   b. User Characteristics: Individuals with experience in AI model training and network security.

   c. Location: Academic and research institutions globally.

4. **Business Processes:**

i. **Data Collection:**

   a. Run network simulations to generate diverse traffic patterns.

   b. Capture network logs in XML format.

ii. **Data Processing:**

   a. Convert XML logs to CSV format.

   b. Label data entries as either benign or DDoS traffic.

iii. **Data Balancing:**

   a. Ensure the dataset has a balanced representation of benign and DDoS traffic.

5. **Features:**

- Traffic Simulation: Execute simulations in NS-3 to generate IoT traffic data.
- Data Conversion: Convert and process XML logs to CSV for model training.
- Data Labeling: Label data entries based on the traffic source as benign or DDoS.

6. **Authorization Matrix:**

Role	Access Level
Developer	Access to raw simulation data and conversion tools.
Researcher	Access to the labeled dataset for analysis.
Data Scientist	Full access to processed and labeled datasets.

7. **Assumptions:**

- NS-3 simulations are run successfully without errors.
- The XML to CSV conversion script is functional and accurate.
- Labeling criteria are clearly defined and adhered to during data processing.

8. **Target Audience:**

- Audience: Data Scientists, AI Developers, Researchers.
- Effort Estimation: Approximately 1 week for complete dataset generation and processing.

9. **Acceptance Criteria:**
    - Simulation data is captured and logged correctly.
    - XML logs are converted to CSV format without errors.
    - Data is accurately labeled as benign or DDoS traffic.
    - The dataset is balanced and ready for model training.

10. **Checklist:**
    - Traffic data generated via NS-3 simulations.
    - XML logs converted to CSV format.
    - Data entries labeled correctly.
    - Dataset reviewed for balance and accuracy.