**Project Report Summary: Adversarial Network Traffic Generation for IDS Evasion**

**Objective**

The project aims to explore the vulnerability of Intrusion Detection Systems (IDS) by generating adversarial network traffic using a Fast Gradient Sign Method (FGSM) attack. The goal is to perturb malicious network traffic in such a way that it evades detection while remaining realistic and functional. The generated adversarial traffic is saved as .pcap files, which can be analyzed using tools like Suricata to evaluate the effectiveness of the attack.

**Brief Description**

**Data Preprocessing :**

The dataset (MachineLearningCVE.csv) contains labeled network traffic data with features such as Source Port, Destination Port, Protocol, Packet Length, and TCP Flags.

Labels are binarized: 0 for benign traffic and 1 for malicious traffic.

Infinite values are replaced with NaN and removed. Features are normalized using StandardScaler to ensure consistent scaling during model training.

**Neural Network Model :**

A simple feedforward neural network (IDSModel) is trained as a surrogate model to classify network traffic as either benign or malicious.

The model architecture consists of three fully connected layers with ReLU activations and a sigmoid output layer for binary classification.

**Adversarial Attack (FGSM) :**

The FGSM attack generates adversarial examples by perturbing the input features of malicious traffic in the direction that maximizes the loss function.

The perturbation magnitude is controlled by the parameter EPSILON. The adversarial examples are designed to evade detection by the surrogate model.

**Packet Generation :**

Scapy is used to create realistic network packets from both original and adversarial feature vectors.

Packets are crafted based on protocol type (TCP, UDP, or raw IP) and include attributes like source/destination ports, packet length, and TCP flags.

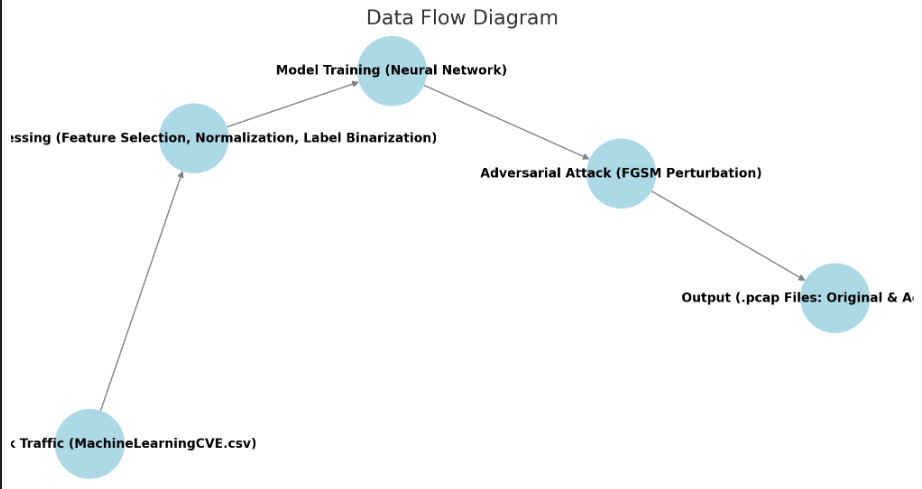
**Evaluation :**

Two .pcap files (original\_traffic.pcap and adversarial\_traffic.pcap) are generated.

These files can be analyzed using Suricata to compare the number of alerts triggered by the original malicious traffic versus the adversarial traffic.

Visuals

Data Flow Diagram :



Input : Raw network traffic data (MachineLearningCVE.csv).

Preprocessing : Feature selection, normalization, and label binarization.

Model Training : Neural network trained on preprocessed data.

Adversarial Attack : FGSM perturbation applied to malicious traffic.

Output : Original and adversarial .pcap files.

Alert Comparison Chart :

A graph showing a red and blue bar chart

Description automatically generated

Alert Comparison

X-axis: Traffic Type (Original vs. Adversarial).

Y-axis: Number of Alerts Triggered by Suricata.

**Outcomes**

Reduced Alert Counts :

The adversarial traffic successfully evades detection by the surrogate model, resulting in fewer alerts when analyzed by Suricata.

This demonstrates the vulnerability of machine learning-based IDS to adversarial attacks.

Realistic Packet Generation :

The use of Scapy ensures that the generated adversarial packets are valid and resemble real-world network traffic, making them suitable for testing IDS robustness.

Insights into IDS Vulnerabilities :

The project highlights the importance of incorporating adversarial robustness into IDS models to mitigate evasion attacks.

**Conclusion**

This project successfully demonstrates how adversarial attacks can be used to craft malicious network traffic that evades detection by an IDS. By leveraging FGSM to perturb features and Scapy to generate realistic packets, the study provides valuable insights into the limitations of current IDS systems. Future work could focus on enhancing model robustness through adversarial training or exploring more sophisticated attack methods.