Network Overload & DoS Attack Detection using Logistic Regression

This repository implements a supervised machine learning pipeline for detecting network overload and Denial of Service (DoS) attacks from wireless network traffic data. The project focuses on identifying anomalous traffic patterns that deviate from normal baseline behaviour, with an emphasis on high detection sensitivity and controlled false positive rates.

The model is trained on a large, preprocessed dataset (439,171 samples, 15 features) containing both benign traffic and multiple attack types, including ARP spoofing, ARP storms, SYN floods, and PING floods. Feature selection prioritises traffic intensity, temporal pressure, and transport-layer state (e.g. frame.time_delta, data.len, TCP flags), enabling effective characterization of abnormal flow behaviour while excluding non-informative identifiers.

Logistic regression is selected as the primary classifier due to its suitability for linearly separable problems, probabilistic output, interpretability, and scalability on large datasets. Model evaluation prioritises recall and precision–recall AUC to reflect the asymmetric cost of missed attacks versus false alarms. Results demonstrate strong and balanced performance, achieving high recall (>0.93), precision (>0.89), F1-score (>0.91), and precision–recall AUC (>0.98), indicating reliable anomaly detection under class imbalance conditions.

The repository also includes analysis of feature correlations, discussion of threshold tuning and regularization for false positive control, and a comparative assessment against k-nearest neighbors, highlighting trade-offs in interpretability, computational cost, and deployment suitability.

This project is intended as a practical, interpretable baseline for network anomaly detection rather than a fully exhaustive intrusion detection system, and is designed to be extensible with additional features, regularization strategies, or alternative classifiers.