Network Intrusion Detection System

TEAM MEMBERS:

Ziad Alaa Osman	Data Scientist
Adham Mohamed Ibrahim	Data Scientist
Mohamed Atta Abu Elhamd	Data Scientist
Sara Khaled Abd El-Samie	Data Scientist
Mahmoud Sayed Mahmoud	Data Scientist

TEAM LEADER: Ziad Alaa Osman

DESCRIPTION

With the rapid increase in cyberattacks, traditional rule-based intrusion detection systems struggle to adapt to evolving threats. There is a need for an intelligent system that can automatically detect anomalies in real-time network traffic and differentiate between normal and malicious activities. The formulated problem is: How can we design a scalable, adaptive, and accurate AI-powered system to detect suspicious activities in network traffic, so This project aims to develop a machine learning–driven solution for network threat detection. The system leverages advanced AI methods such as anomaly detection, Natural Language Processing (for log analysis), and classification algorithms to recognize malicious behaviors. It analyzes multiple network features including packet flow rates, inter-arrival times, TCP flags, and communication patterns to detect threats

OBJECTIVE

- Provide real-time monitoring of network traffic.
- Detect cyber threats early, before escalation.
- Automate alerts to reduce the workload on security teams.
- Ensure scalability to handle large volumes of traffic.
- Enhance cybersecurity resilience through proactive and adaptive defense mechanisms.

TOOLS AND TECHNOLOGIES:

- Python
- NumPy
- Pandas
- Scikit learn
- Matplotlib
- Seaborn

Milestones:

- **Milestone 1: Data Collection**
  In this phase, relevant data is gathered from various sources such as public datasets. The main goal is to collect high-quality, comprehensive, and representative data that aligns with the project objectives.

- **Milestone 2: Data Cleaning**
  This stage focuses on improving data quality by handling missing values, removing duplicates, correcting inconsistencies, and addressing outliers. Clean data ensures accuracy and reliability for analysis and model training

- **Milestone 3: Eda and visualization**
  EDA involves examining data patterns, distributions, and relationships using statistical summaries and visualizations. It helps uncover insights, detect anomalies, and guide further preprocessing or modeling steps.

- **Milestone 4: Data preprocessing for machine learning**
  In this step, the data is prepared for machine learning models. Tasks include encoding categorical variables, scaling numerical features, splitting data into training and testing sets, and handling class imbalances if present.

- **Milestone 5: Applying machine learning Algorithms**
  This milestone involves selecting and implementing suitable machine learning algorithms to solve the problem. Models are trained, validated, and evaluated based on performance metrics to identify the best-performing approach.

- **Milestone 6: Deployment using MLOPS**
  The final stage focuses on deploying the trained model into a production environment using MLOps practices. This includes model versioning, continuous integration/continuous deployment (CI/CD), monitoring, and ensuring scalability and reliability in real-world use.

KPIs (Key Performance Indicators):

1. Data Quality

Percentage of missing values handled:100%

Data accuracy after preprocessing: ≈99%

Dataset diversity (representation of different categories):high diversity across 7 classes including 6 attack types

2. Model Performance

Model accuracy (Accuracy/F1-Score):99.88%,99.8%

Model prediction speed (Latency): ≈20–50 ms per request (estimated)

Error rate (False Positive/False Negative Rate): <0.2% FP, <0.5% FN

3. Deployment & Scalability

API uptime ≈99.9% (expected)

Response time per request: 20–50 MS

4. Business Impact & Practical Use

Reduction in manual effort: ≈85–90% reduction

Expected cost savings: ≈70% savings

User satisfaction: ≥95% satisfaction (expected for real-world use)