Machine Learning-Based Network Intrusion Detection

Ayoub Al-Mubaydeen

Bashar Rasheed

Mohammad Al-Ashqar

Project Overview

This project presents a comprehensive three-stage machine learning pipeline for network intrusion detection and attack classification, designed to enhance accuracy, efficiency, and interpretability in cybersecurity applications.

Motivation for a Multi-Stage Pipeline

Network traffic data is inherently complex and noisy, often dominated by benign activity with a wide variety of protocols and attack types. A single monolithic model struggles to effectively process this heterogeneity and volume.

- By decomposing intrusion detection into three specialized stages, this project addresses challenges of:
- Data volume and noise reduction through early traffic type filtering.
- Class imbalance by isolating suspicious flows before detailed attack classification.
- Computational efficiency by narrowing the scope at each stage.
- Enhanced interpretability by providing modular insights at multiple levels of analysis.

Pipeline Stages

Traffic Type Classification
The initial model systematically categorizes network traffic by protocol (e.g., DNS, HTTP, FTP). This multi-class classification facilitates targeted analysis and reduces irrelevant data noise in subsequent stages, improving downstream model focus and performance.

Suspicious Traffic Detection

Leveraging binary classification, the second model discriminates between benign and potentially malicious traffic. This gatekeeping step optimizes resource allocation by focusing attention and computational effort on suspicious flows, while minimizing false alarms from benign traffic.

Why This Project Matters

- Real-world Applicability: Reflects practical IDS workflows and network traffic complexity.
- Layered, Modular Architecture: Enhances accuracy and interpretability at each detection phase.
- Clear Evaluation: Standard metrics allow objective performance assessment and continuous improvement.
- Scalability: Easily extendable to larger networks, new traffic types, and emerging attack vectors.

Dataset Explanation – CIC-IDS-2017

- **Source:** Canadian Institute for Cybersecurity (UNB)
- **Period:** July 3–7, 2017
- **Types:** Benign, DoS, DDoS, Brute Force, Heartbleed, Web Attacks, Infiltration, Botnet
- **Format:** PCAP → CSV via CICFlowMeter
- **Features:** 79 columns (78 numeric, 1 label), over 2.8M flows

Attack Types & Distribution

- **Attack Categories:** DoS, DDoS, Brute Force, Heartbleed, Web Attacks, Infiltration, Botnet
- **Class Imbalance:** Most records labeled as 'Benign'
- **Link:** https://www.unb.ca/cic/datasets/ids-2017.html

XGBoost?

1 What is XGBoots

Extreme Gradient Boosting (XGBoost) is a fast, accurate machine learning algorithm for structured/tabular data.

Builds an ensemble of decision trees sequentially, each correcting errors from previous trees.

Uses gradient boosting: fits new trees to gradients (and Hessians) of the loss function for precise optimization.

Incorporates regularization (L1 & L2) to control model complexity and prevent overfitting.

Optimized for speed and scalability: supports parallel processing, missing data handling, and efficient memory use

2 How XGBoost Works & Why Use It?

Training process per iteration:

Compute first and second derivatives (gradients & Hessians) of loss to guide tree splits.

Build trees by selecting splits that maximize gain while factoring in regularization penalties.

Update predictions by adding outputs from the new tree.

Final output: Sum of all trees' predictions; applies activation function depending on task.

3 Advantages:

High accuracy in classification and regression tasks.

Fast, scalable training on large datasets.

Robust to overfitting thanks to regularization and subsampling.

Flexible with support for custom loss functions.

Data Preprocessing

Missing & Infinite Values & duplicated records:

- [ ] Identify missing (NaN) and infinite values from raw data.
- [ ] Replace missing and infinite values impute with statistical methods (mean/median).
- [ ] Replace missing and infinite values impute with statistical methods (mean/median).

Introduction to Multi-Stage Pipeline

Cybersecurity faces a growing challenge with increasingly sophisticated network attacks.

Intrusion Detection Systems (IDS) must be accurate, efficient, and adaptable.

Single-model IDS often struggle with noisy data and imbalanced classes.

This project proposes a three-stage pipeline for layered detection:

- Stage 1: Traffic Type Classification — sorts traffic by protocol to reduce irrelevant noise early.
- Stage 2: Suspicious vs. Benign Detection — binary filter prioritizing resources on potentially harmful traffic.
- Stage 3: Attack Type Classification — detailed categorization of attack types for effective response.
- This modular approach improves interpretability, performance, and scalability

Model 1 – Traffic Type Classification

```
Accuracy: 0.9996
Precision: 0.9994
Recall: 0.9996
F1 Score: 0.9995
```

Classification report:

	precision	recall	f1-score	support
DNS (UDP)	1.00	1.00	1.00	175377
FTP (TCP)	1.00	0.98	0.99	1950
HTTP/HTTPS (TCP)	1.00	1.00	1.00	203047
NTP (UDP)	1.00	1.00	1.00	4267
NetBIOS (UDP/TCP)	1.00	1.00	1.00	1250
Other	1.00	1.00	1.00	115729
RDP (TCP)	0.00	0.00	0.00	29
SMB (TCP)	1.00	1.00	1.00	418
SNMP (UDP)	0.00	0.00	0.00	27
SSH (TCP)	1.00	1.00	1.00	2066
accuracy			1.00	504160
macro avg	0.80	0.80	0.80	504160
weighted avg	1.00	1.00	1.00	504160

- [ ] Goal:

Classify incoming network traffic into distinct types such as DNS, HTTP, FTP, etc.

- [ ] Why it matters:

Organizes network traffic by protocol for better analysis.

Filters irrelevant noise early, improving downstream detection accuracy.

- [ ] Modeling Approach:

Multi-class classification problem

- [ ] Algorithms: XGBoots
- [ ] Evaluation Metrics:

Accuracy

Precision & Recall per traffic class

Confusion matrix to visualize classification performance

Model 2 – Suspicious vs Benign Detection

```
Accuracy: 0.9939
Precision: 0.9965
Recall: 0.9673
F1 Score: 0.9817
```

Classification report:

	precision	recall	f1-score	support
Normal Traffic	0.99	1.00	1.00	419012
Attack Traffic	1.00	0.97	0.98	85148
accuracy			0.99	504160
macro avg	0.99	0.98	0.99	504160
weighted avg	0.99	0.99	0.99	504160

- [ ] Goal:

Binary classification to detect whether network traffic is benign or suspicious.

- [ ] Why it matters:

Prioritizes security monitoring on potentially malicious traffic.

Reduces false alarms by filtering benign flows.

- [ ] Modeling Approach:

Binary classification problem

- [ ] Algorithms: XGBoost

Evaluation Metrics:

- [ ] Accuracy

Precision, Recall, F1-Score

Model 3 – Attack Type Classification

```
Accuracy: 0.9954
Precision: 0.9954
Recall: 0.9954
F1 Score: 0.9949
```

```
Classification report:
                          precision    recall  f1-score   support

                  BENIGN       1.00      1.00      1.00    419012
                     Bot       1.00      0.39      0.56       390
                    DDoS       1.00      1.00      1.00     25603
           DoS GoldenEye       0.99      0.97      0.98      2057
                DoS Hulk       0.99      0.98      0.98     34569
        DoS Slowhttptest       0.92      0.99      0.96      1046
           DoS slowloris       0.99      0.99      0.99      1077
             FTP-Patator       1.00      0.99      1.00      1186
              Heartbleed       1.00      0.50      0.67         2
            Infiltration       1.00      0.86      0.92         7
                PortScan       0.99      1.00      0.99     18139
             SSH-Patator       1.00      0.91      0.96       644
  Web Attack � Brute Force       1.00      0.09      0.16       294
Web Attack � Sql Injection       1.00      0.25      0.40         4
          Web Attack � XSS       1.00      0.02      0.05       130

                accuracy                           1.00    504160
               macro avg       0.99      0.73      0.77    504160
            weighted avg       1.00      1.00      0.99    504160
```

- [ ] Goal:

Classify suspicious traffic into specific attack categories (e.g., DoS, DDoS, Brute Force, Heartbleed).

- [ ] Why it matters:

Provides granular identification of attack types for targeted responses.

Enhances threat intelligence and mitigation strategies.

- [ ] Modeling Approach:

Multi-class classification problem

- [ ] Algorithms: XGBoost

Evaluation Metrics:

- [ ] Accuracy

Precision, Recall, F1-Score per attack type

Confusion matrix to understand misclassification trends

Summary & Key Takeaways

1. Designed a detailed, multi-stage intrusion detection pipeline.
2. Used strong classical and boosting ML models tuned for cybersecurity data.
3. Tackled major challenges: data imbalance, noise, interpretability.
4. Set foundation for continuous improvement with future work directions.

Future Work

1. Incorporate anomaly detection techniques to identify zero-day or unknown attacks.
2. Experiment with deep learning models on time-series traffic data.
3. Expand dataset with new threats and simulate emerging attack vectors.
4. Develop real-time streaming analytics for continuous network monitoring.
5. Enhance user interface and alerting mechanisms for operational environments.

Thank you all