Cyber-Attack Classification Project Report

---

1. Project Description

In this project, we aim to build a machine learning classification model that can accurately distinguish between different types of cyber-attacks (e.g., DDoS, DoS, Mirai, Recon, MITM) and benign traffic, based on per-flow network traffic features. By analyzing features such as flow durations, packet rates, flag counts, and protocol indicators, the model will learn characteristic patterns of each attack type to enable real-time detection and response.

---

2. Dataset Overview

- **Source:** Network traffic records collected from simulated or real-world environments.
- **Rows:** 938,583 flow records.
- **Columns:** 22 features describing each flow, plus 1 target label.
- **Task:** Predict the `label` (attack type or benign) given the observed features.

---

3. Feature Categories

We group the columns into three categories:

1. **Numerical Features** (continuous or count data)
2. **Binary Features** (flags or protocol indicators)
3. **Categorical Target** (the label to predict)

3.1 Numerical Features

Feature	Description	Usage in Model
flow_time	Total duration of the flow (seconds).	Longer durations may indicate slow scanning or persistent attacks.
header_size	Size (bytes) of packet headers.	Larger headers can suggest application-layer protocols (e.g., HTTP).

Feature	Description	Usage in Model
packet_duration	Time (ms) between first and last packet.	Distinguishes bursty traffic from more uniform transfer.
overall_rate	Average data transfer rate of the flow (bytes/sec).	High rates are typical of volumetric attacks (e.g., DDoS).
src_rate	Transfer rate from the source IP (bytes/sec).	Asymmetry in rates can signal attack flows (e.g., reflection attacks).
dst_rate	Transfer rate to the destination IP (bytes/sec).	Low dst_rate with high src_rate can indicate reflection attacks.
fin_packets	Count of packets marked with the FIN flag.	FIN-heavy flows may reflect clean connection teardowns.
urg_packets	Count of packets with the URG flag.	Rare in normal traffic; can signal reconnaissance or malformed packets.
rst_packets	Count of packets with the RST flag.	Excessive resets often accompany scanning or aborted connections.
max_value	Maximum payload byte value observed in the flow.	Can reflect payload content characteristics.
value_covariance	Covariance of payload byte values over the flow.	Captures payload variability—higher in complex protocols.

3.2 Binary Features

Feature	Description	Usage in Model
fin_flags	Any packet had a FIN flag? (0 = no, 1 = yes)	Helps detect proper connection teardowns vs. abrupt drops.
syn_flags	Any packet had a SYN flag?	SYN-heavy flows can denote connection attempts or port scans.
rst_flags	Any packet had an RST flag?	Spike in resets often correlates with aborted scans or attacks.
psh_flags	Any packet had a PSH flag?	Used in streaming or real-time data; abnormal in certain attacks.
ack_flags	Any packet had an ACK flag?	ACK patterns help profile completed handshakes vs. half-open connections.
protocol_http	HTTP traffic indicator (1 if HTTP seen).	Distinguishes web traffic from other protocols.
protocol_https	HTTPS traffic indicator.	Encrypted web sessions flagged separately.

Feature	Description	Usage in Model
protocol_tcp	TCP protocol indicator.	Core transport-layer protocol; basis for flag analysis.
protocol_udp	UDP protocol indicator.	Connectionless traffic (e.g., DNS, streaming) identified.
protocol_icmp	ICMP protocol indicator.	Ping and network diagnostics traffic flagged.

3.3 Categorical Target

Feature	Description
label	Type of traffic in the flow:• DDoS• DoS• Mirai• Recon• MITM• BenignTraffic

The classifier will learn to predict this target category.

---

4. Conclusion

This structured overview of the dataset—defining each column, its role, and grouping by feature type—will guide data preprocessing, exploratory analysis, and model development phases. By carefully handling numerical scaling, class imbalance, and leveraging binary protocol indicators, one can build a robust classifier to detect and categorize cyber-attacks effectively.