CYBERML – Project

Year : 2025-2026
Lecturer: Pierre Parrend

Project objectives

The goal of the project is to design, deploy and evaluate a data chain for the analysis of cybersecurity data. The data treatment will be performed as batch.

The objective of your analysis is:

(objective 1) ***Classification and Anomaly detection for tracking attacks***

A 20% bonus is applied if you complete

(objective 2) ***Adversarial attacks against classification***

Choose the dataset to analyse among Cybersecurity Datasets for IoT and Industrial IoT:

- CIC IIoT dataset 2025
  - Data and doc : https://www.unb.ca/cic/datasets/iiot-dataset-2025.html
- CIC IoT-DIAD 2024 dataset
  - Data and doc : https://www.unb.ca/cic/datasets/iot-diad-2024.html

Launch

Launch your groups:

- Build a group of 4 people
- Set group number from 1 to 22 in
  - CYBERML_SCIA_2025_2026_project_groups
- Groups will be frozen 4 days after project launch

Deliverables

The deliverables are:

- Analysis notebook, shared on google collab (or similar notebook support)
- Analysis report (20 pages)
- Final oral group presentation (10 min) + demonstration (5 min)

Your report will present the detailed specification and implementation details on:

- The complete deployment of the data handling chain (including classification + anomaly detection)
- Characterization of the dataset under study
- Benchmark of
  - 3 complementary unsupervised algorithms
  - 3 complementary classification algorithms
- The benchmark must include confusion matrix, precision, recall, AUPRC, Balanced accuracy, Matthews Correlation Coefficient




- Conclusions about cybersecurity events in the dataset

The oral presentation is a security analysis review based on the report.

Specifications

The choice of the dataset, the design of the data handling chain as well as the choice of the analysis algorithm is part of the work.