CYBERML – Project

Year : 2024-2025
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) ***Anomaly detection for tracking attacks***

A 25% bonus is applied if you complete

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

Choose the dataset to analyse among Cybersecurity Datasets for core networks:

- Hardware In The Loop
  - Data and doc : WDT2022
- Secure Water Treatment
  - SWaT.A3_dataset_Jul 19_labelled
  - Doc : go2016SWAT.pdf

Launch

Launch your groups:

- Build a group of 4 people
- Set group number from 1 to 17 in: MLSecu_SCIA_2024-2025_project_groups.xlsx
- Groups will be frozen
  - on 10/12/2024 evening for KB
  - on 11/12/2024 evening for LYS

Intermediate presentation

Present your first results:

- 1 dataset
- 1 learning algorithm
- Characterisation of 1 type of attack

Deliverable:

- Notebook + export (HTML or PDF) + prés PPT

Deliverables

The deliverables are:

- Analysis notebook, shared on google collab
- 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 analysis algorithms, including 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 data handling chain will comply with following requirements:

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.