COMP3217 COURSEWORK 2

MUHAMAD ABEED SAHARUDIN (STUDENT ID: 33371806)

May 9, 2024

1 Problem Statement

The problem that were given is to detect physical cyber-attacks on a power system framework configuration using machine learning (ML). There are many different scenarios that can happen on the framework that can indicate whether a physical cyber attack is targeted towards the framework [3], however for this coursework only 3 are described and analysed, which are natural faults, data and remote command injection attacks, and normal events.

This coursework is divided into 2 parts. The first part is concerned with only detecting between normal events and data injection attack events. The second part is concerned with the same events, with the addition of remote command injection attack events.

This report explains the data composition, ML pipeline design, and training evaluation for both parts of the courswork.

2 Data Composition

The training data given to train the ML algorithm are 2 `csv` files, each containing 6,000 system traces, with each of those containing 128 features. The first training file is for the first part of this coursework, where the last column is a marker that serves as a label indicating whether an event to the framework is a normal event (labelled as 0) or a data injection attack event (labelled as 1). The second file is for the second part of this coursework, which contains the same data marker, with the addition of the third event, a remote command injection attack (labelled as 2).

3 ML Pipeline Design

It is required that the design of the ML workflow be done in Google Notebook. The link is provided in the References section [1, 2].

3.1 Pre-Processing

Before training, data pre-processing was done to better prepare the ML models for maximum performance on training and better generalisation on unseen data. One of the pre-processing methods used is standardising the dataset so its data distribution has mean 0 and unit variance.

3.2 ML Models

For this coursework, the `RandomForest` classifier from `scikit-learn` was used to

4 Training Evaluation

5 Conclusion

6 References

[1] *Google Colab*. URL: https://colab.research.google.com/drive/1D-d-fE6HDtyurjdwe30B6qVHKBR9
scrollTo=yyC0y3AKpFEh.

[2] *Google Colab*. URL: https://colab.research.google.com/drive/1LSB2kUiGJ3uFcUVLtb6Fy5C2c3EF
scrollTo=yyC0y3AKpFEh.

[3] Mississippi State University and Oak Ridge National Laboratory. *Power system attack datasets*. Apr. 15, 2014. URL: http://www.ece.uah.edu/~thm0009/icsdatasets/PowerSystem_Dataset_README.pdf.