Option A: Detection of Manipulated Pricing in Smart Energy CPS Scheduling

The coursework involves writing programs for the task described below, and you can choose any major programming language such as C/C++, Python, etc. The purpose of this coursework is to understand the linear programming based energy scheduling for smart home cyber-physical system, understand the interdependence between the pricing information and the energy load scheduling, develop detection techniques for pricing attacks, and get familiar with some cyber-physical system security programming skills.

- Consider a small community consisting of only 5 users, each of whom has 10 smart home appliances. The details are included in COMP3217CW2-Input.xlsx
- You are given a set of 10,000 predictive guideline price curves (TrainingData.txt), where half of them are labelled as Normal and the other half are labelled as Abnormal. There are 25 numbers in each predictive guideline price curve, where the first 24 numbers refer to the unit for each hour and the last number is binary indicating whether it is a normal pricing curve (if it is 0) or an abnormal pricing curve (if it is 1). These 10,000 predictive guideline price curves are basically the training data.
- You are also given 100 predictive guideline price curves without labels, which are the testing data (TestingData.txt), where there are 24 numbers in each predictive guideline curve.
- You need to design and implement a technique to model those training data and compute the labels for all testing data (i.e., 0 or 1 for each predictive guideline pricing curve). In the report, you need to clearly indicate the computed label for each predictive guideline pricing curve in the testing data. You also need to output a file (TestingResults.txt) with the same format as the training data and submit it together with your source code.
- For each of those testing data labelled with Abnormal (i.e., labelled with 1) as determined by your code, you will need to compute the linear programming based energy scheduling solution according to that abnormal predictive guideline price curve, and plot the scheduling results showing the hourly energy usage of this community (i.e., 24 bars where each bar shows the total energy usage from all of 5 users during the corresponding hour).
- In the report, you should give a clear description of the problem, the linear programming based scheduling algorithm, and your technique to compute the labels. You need to show the computed labels for all of the 100 testing data. You should also analyse and discuss your results (including, but not limited to, the training error). The accuracy on 100 testing data (testing error) will be an important factor in determining your performance on this coursework, followed by the classification accuracy on 10,000 training data (training error).
- You should make good use of a software repository (e.g., git or github) and your code needs to be maintainable and designed with a make-based build structure.
- Your code should contain sufficient comments.
- Your code should be laid out neatly with consistent indentation.