COMP4337/9337 Securing Fixed and Wireless Network
Lab 6: Using Machine Learning for Intrusion Detection

Due: Tuesday April 12th, 2022 2359 hrs.

Objectives

This lab is designed to help the students learn:

- Basic machine learning concepts that can be applied to classify malicious traffic samples from benign (non-malicious) traffic.

Lab Overview

Machine learning is a branch of artificial intelligence (AI) focused on building prediction models that learn from data and improve their performance over time without being programmed to do so. Security experts are focusing on the development and application of machine learning techniques for network security in the areas of malware detection, traffic classification, anomaly detection, etc. due to its ability to keep pace with the evolution of threats in networking.

This lab will guide you through an example of using a machine learning (ML) algorithm to train a model for intrusion detection. You have been provided with a Jupyter Notebook with markdown explanations and code, as well as a custom dataset (Further details are provided in the Notebook).

In order to do this lab, you need to have some fundamental knowledge on machine learning. Please refer to a free course, “Applied Machine Learning: Foundations” (https://www.linkedin.com/learning/applied-machine-learning-foundations/) on LinkedIn Learning. LinkedIn Learning is free for all UNSW students and Staff (https://www.myit.unsw.edu.au/services/staff/educational-technology/linkedin-learning). Please follow the instructions at the above link to obtain access. If you already have some knowledge on machine learning, please feel free to skip this step.

Requirements:

To attempt this lab, you require an environment setup that includes Python3, Jupyter Notebook or JupyterLab (https://jupyter.org/install), and some python-based ML packages (Scikit Learn, Pandas, NumPy etc.) installed on your system. You can either install everything separately or install Anaconda (https://www.anaconda.com/products/individual) toolkit.

You are provided with the following two files:

- A Jupyter notebook that contains a working example of using a supervised ML algorithm, K Nearest Neighbor (k-NN).
- A dataset file (.csv) that contains benign and malicious samples of an IoT device.

The lab consists of two parts. In Part A, you are required to use the supplied Jupyter notebook and run the example with k-NN to train a ML model. Next, you will evaluate the performance of this trained model for classification purposes. In Part B, you are required to produce a separate Jupyter notebook by using another supervised ML algorithm, Decision Tree (DT) on the same dataset and compare the results for k-NN and DT. More details are available in the provided notebook.

Assessment and Marking

The Lab assessment details are included as markdown text in the provided Jupyter notebook. The marks will be made available within 2 weeks of the submission date. The details of the marks are as follows:

- Total marks for Lab 7 are 100, these would be scaled down to 4 marks.
    - Part A: Lab performance **(25)**
    - Lab Assignment submission on Moodle **(75)**

Note: Lab performance involves running the first part of Lab following the supplied ML example on the Jupyter notebook. The tutor may ask questions/ provide feedback on your work.