US Ignite MODBUS Network Flows

Data Preparation

A similar process was executed on the US Ignite PCAP file as the 5G PCAPs. A shell script calls a tshark command which extracts the previously selected fields from each packet and saves them to a CSV file. The CSV file is loaded into the extractor where partitioning and feature extraction occurs as before. However, since the US Ignite data captures MODBUS network traffic rather than 5G, the data rates are significantly lower. To account for this, the extractor was modified to convert bytes to kilobits instead of megabits. Other than this minor adjustment, the extractor remains the same.

The US Ignite data was run through the extractor using both timeout and fixed interval sub-partitioning methods. Setting the timeout interval greater than 5 seconds resulted in unbroken flows and setting it lower resulted in many subflows with a short duration (< 2 sec). I decided to try the fixed interval method and compare my results. I initially tested with a 10 second interval but switched to 5 upon noticing many subflows with 3 to 5 second durations. This method results in fewer subflows with short durations, and I feel it better captures the average network traffic.

Training and Testing

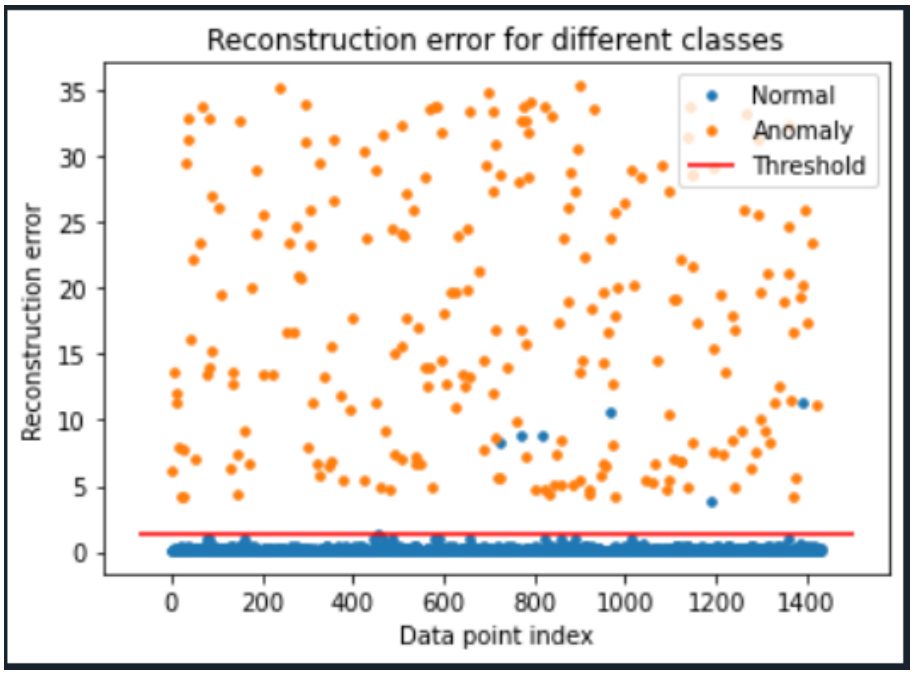
The US Ignite features were trained on the same sixteen model designs as the 5G features. To ensure consistency, no shuffling occurs during training (the data is already shuffled by the extractor). Different combinations of Principle Component Analysis (PCA) principles, activation functions, and optimizers were applied to the autoencoder, and the trained model from each test was saved. Saved models can be loaded on demand. The detection threshold is set at 3 standard deviations, which includes approximately 99.7% of nominal data points.

DDoS Simulation

The loader program allows for the generation of synthetic flows, which is useful for simulating DDoS attacks against monitored devices. To simulate a malicious traffic flow, network statistics outside the nominal range must be generated. To accomplish this, a function was created which takes as input a packet size and number of packets per second and returns a feature vector of network statistics. These two parameters can be adjusted to control the data rate, and a data frame with any number of these attacks can be saved for testing purposes.

Each anomalous flow has equal-sized packets which makes statistics easy to calculate. Our tool can generate several flows at once, each with different data rates if desired. This allows us to visualize a gradient of synthetic flows with increasing data rates. Currently the gradient feature works by fixing the packets per sec and increasing packet sizes. It can be modified to work vice versa, but the idea of increasing the data rates would remain the same.

Example using inward flows with model 9




```
accuracy:   0.9958071278825996
recall:     1.0
precision:  0.9755102040816327
f1-score:   0.9876033057851239
```