HCL Hack 2020

Bug Slayers

CR2

Workflow

The question that we chose was DDOS attack detection, why we might add hold on.
The data was given to us in the form of PCAP files with millions of packets. The dpkt library in python was used to parse the data into CSV from which the features were extracted. We chose the bidirectional node pairs (no difference between source and destination and destination and source) as our data points to detect the traffic. The features we selected are a result of reading a lot of research papers and selecting the most appropriate ones out of them. They're indicated in the python file.

Moreover, we had to assume that the labelling is done such that all traffic within benign files are labelled as benign, and all traffic within malicious files are labelled as malicious.

Observations and Results:

Logistic regression on the data gave us an accuracy of 75.8% which implied that the data points are non-linear. We trained the data on the Random Forest Classification model. The confusion matrix is as follows:

```
3909     0
 4      1001
```

Accuracy: 0.999
F2 Score: 0.999