Project proposal:malware detection using network traffic

woyu li

November 1, 2025

1 Project summary

Network traffic contains both benign and malicious flows. Accurately identifying malware traffic is critical for cybersecurity. In this class project, I choose to explore and reproduce the malware classification result provided by nPrient pcapML benchmark.

2 data

The dataset for this project is available under netML Malware Detection website. The dataset contains 1.1GB traffic data. In addition, we can do 2 class classification, and 19 class classification on this dataset.

3 Machine learning

I will try tree based model and deep neural networks and compare their result.

4 evaluation

I will use balanced accuracy suggested by the website as my main metric of evaluation. However, only relying on accuracy will lead to bias. For example, if malware is very rare in the dataset, classifying every traffic as benign will lead to good accuracy, but the model will be useless in that case. Therefore, I will also use recall, precision, F1 score as auxilary metrics.

5 Learning objective

My learning objective is to improve my skill of processing raw data by doing feature engineering. And learn how to build a ML pipeline on real world data to consolidate my understanding of machine learning. In addition, I want to improve my paper reading ability by exploring existing methods.