CMPEN 462

Mini-Project #2a

Cyber Security and ML for Wireless Comms:: Intrusion Detection

Due: April 10, 2025 (11:59PM EST)

(worth 20 pts total)

Cyber security is a broad term that refers to the use of algorithms for malware, intrusion, and phishing detection as well as the area of cryptography and privacy techniques. This project will focus on intrusion detection learning software to detect network attacks from various users (unauthorized or insiders who may not be authorized but have ready access). This project will require you to implement a binary classifier to distinguish normal connections from attacks followed by an implementation of a deep-learning based approach where you will get to train the network and then make a comparison to the classical methods from the first task. In-class lectures, homework, and mini-labs will be providing the necessary theory elements to equip you to perform this MP.

You will be gaining hands-on experience (maybe for the first time) of implementing machine learning algorithms, gaining insight and understanding of how cybersecurity is related to the protection of our wireless networks to malicious actors, and the challenges associated with distinguishing 'good' and 'bad actors.

Project:

You will build software to detect network intrusions. 'Intrusion detection' protects a network from unauthorized users, including insiders. The intrusion detector mini-project is oriented towards building a predictive model (i.e. a 'classifier') capable of distinguishing between ``bad'' connections, called intrusions or attacks, and ``good'' normal/benign connections.

1. You will build a binary classifier to distinguish normal connections from threat connections (attacks).
   a. Design and implement (the most successful) classical ML classifiers: Logistic Regression, Support Vector Mahine, and Random Forests
   b. Design and implement state-of-the-art Deep Learning-based approaches and train NN
2. Compare the results of the two classifiers (classical vs. state-of-the-art)

It is important to note that the test data is not from the same probability distribution as the training data, and it includes specific attack types not in the training data. This makes the task more realistic. Some intrusion experts believe that most novel attacks are variants of known attacks, and the "signature" of known attacks can be sufficient to catch novel variants. The KDD90 datasets contain a total of 24 training attack types, with an additional 14 types in the test data only.