IDS Deception using Generative Adversarial Networks

Borys Bryndak, Allan Cheng

Motivation

- It is important to detect various attacks on the networks so that preventative measures can be taken
- Intrusion detection systems may not be able to adapt to obfuscated network traffic
- We are developing a system that is able to test the resilience of IDS by masking the traffic of network attacks

Problem

- Using dataset CSE-CIC-IDS2018, train a generative adversarial network that is capable of taking a real network attack and modifying some of its non-essential features in such a way as to confuse an IDS into believing it to be normal traffic

- To this end, it is necessary to identify all functional features of each attack type (those that allow us to reconstruct a valid packet sequence that constitutes an actual attack)

- Utilize Conditional Wasserstein GANs to create obfuscated network traffic for each attack type and replace the functional features to create a valid attack

Challenges

- In the dataset, by using a linear neural network, it was apparent that infiltration attacks are hard to determine without using time data

- It is difficult to identify functional features in the records containing 80+ features for each of 13 attack types

(Preliminary) Solution

- Preprocessed the data by replacing categorical features like the protocol and the destination port with dummies, splitting attacks and benign traffic into two datasets for training GAN and normalizing the features

- Created a network (the generator) that takes in an attack type and some random noise and outputs an attack record with the functional features replaced by those of a valid attack

- Also trained a set of intrusion detection systems using different methods (SVM, MLP, Nearest Neighbours, Naive Bayes, Decision Trees)

(Preliminary) Results

- By comparing the detection rate of the intrusion detection systems on the attack records from the dataset and those produced by the generator, we can see that the generator is successful in confusing the IDS

- However, these results are not quite valid yet, since we haven’t identified all the functional features of the attacks, and so the output of the generator may not be used to reconstruct a real attack