Machine Learning Engineer Nanodegree

Capstone Proposal

Cody Durden

August 27th, 2018

Proposal

Domain Background

Intrusion Detection Systems (IDS) are a type of software that monitors for malicious activity or deviations from set policy. Typically, IDS's are hosted on a company's network and monitor traffic. This serves as a good basis for protecting against unwanted intruders in a company's network. But as all things evolve so does the way these nets of protection should protect. There are generally two different ways an IDS protects the network; signature and anomaly detection. Signature detection uses known signatures in network traffic to identify malicious traffic. Anomaly detection searches for deviations in normal traffic to try and identify potential malicious traffic. As is usually the case both used in tangent can serve as more beneficial.

Ever since I got into computers I have been drawn to computer security. Maybe it was the movies or just something latent but being able to break into a computer has always been fascinating to me. So, on the flip side I would like to use my skills to prevent the real bad guys from causing havoc in a system.

Problem Statement

The problem with IDS's are that they don't pick up all threats. The reason for this is that things evolve; signature detection that previously caught 99% of threats is now 80% (just an example). The bad guys aren't going to sit back idle and throw their hands up; no, they will develop more innovative ways to get into a network.

Datasets and Inputs

The dataset being used is the 'Intrusion Detection Evaluation Dataset (CICIDS 2017). The dataset includes 5 days of activity. The first day (Monday) includes only normal activity. The following days (Tuesday-Friday) include normal plus attack traffic. The goal of this dataset was to include realistic background (normal traffic) traffic as well as the most up to date common attack traffic.

Generating realistic background traffic was top priority in this dataset and this was obtained by using a B-Profile system (Sharafaldin, et al. 2016) which captures the abstract interactions of humans and created naturalistic benign background traffic. Also, this dataset was created to address some of the shortcomings of past IDS datasets. 11 criteria were named as being relevant to be considered a benchmark dataset.

Solution Statement

The problem is that the current state of IDS's are very good at detecting previously determined attacks but have less of an ability to determine new and forming attacks. The solution to this is by using state of the art learning algorithms we can better classify and identify new attacks. The ultimate determination of improvement will be the accuracy of detection. If the newly trained algorithms can better detect new and existing attacks then there is immediate benefit from the training.

Benchmark Model

The benchmark for testing would be the current detection methods. Currently the two main detections methods are signature-based and anomaly-based. Comparing newly developed methods versus these two methods will allow for an excellent benchmark test.

Typically, the use of binary classification is suitable for reliability testing of a particular IDS. I can compare the newly trained model against this same metric. If it shows that the ratio of true positives far exceeds the current standard, then we have shown to improve the current state of the IDS.

Evaluation Metrics

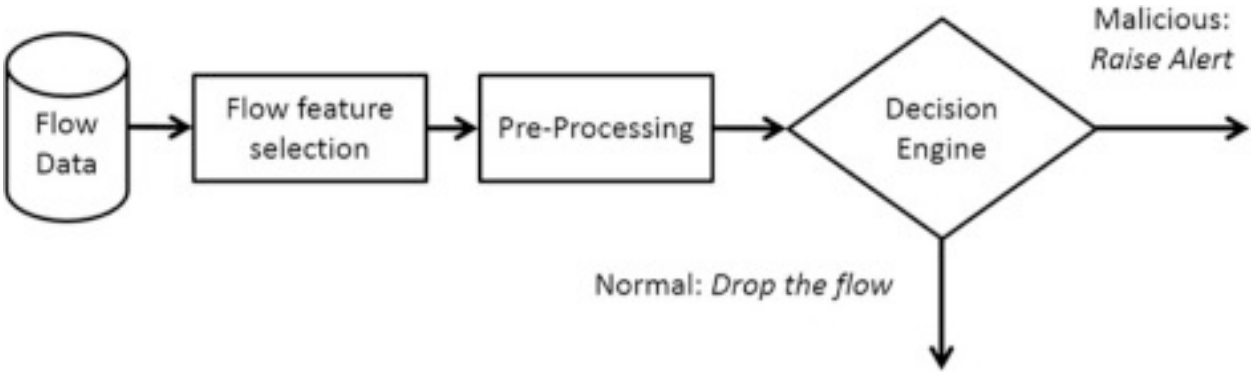
Evaluation of the IDS can be determined using binary classification. By using this metric we can accurately state if improvement was obtained. For example, if the current

IDS determined 10 true positives and 5 false positives in our training dataset and the newly trained IDS determined 12 true positive and 3 false positives I would say the new trained IDS has exceeded and improved on the current state.

Project Design

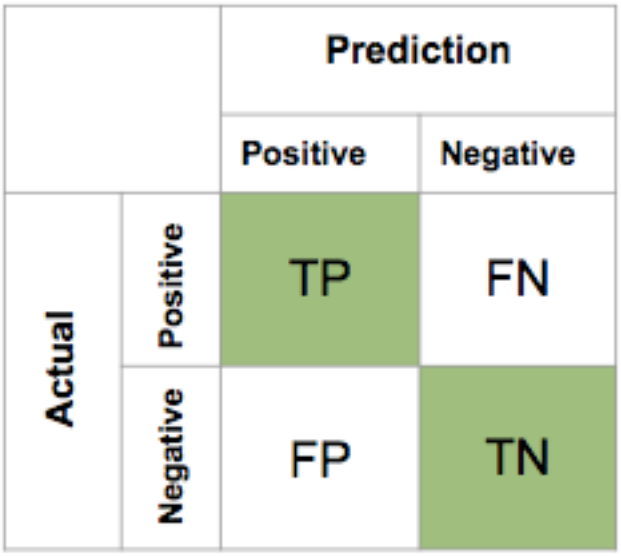
The design of this project is relatively simple in theory. I will examine an existing dataset prepared by the University of New Brunswick. Using this existing training and testing dataset I will be able to fast pace into training. The goal is to train the data using a variety of methods including a deep learning neural network, Naïve Bayes and Logistic Regression machine learning algorithms. By using these methods the expected outcome is that we can achieve a better prediction of attacks than the current anomaly and signature detection. For comparison to current methods I hope to place this dataset into an IDS with current rulesets.

Below is a graphic of a simplified IDS.



1) The information flows into the IDS
2) Some decision is made
3) Traffic is passed thru or dropped

The quality of the learning will be quantified using binary classification.



For a quality IDS most of the results should be a True Positive, True Negative, or False Positive. The worst thing that can happen is to have a lot of False Negatives. This would increase the likelihood and intruder could infiltrate the network. The metrics being used will be the F-score and Accuracy F-score.

a. $F_\beta = (1+\beta_2) \cdot precision \cdot recall (\beta_2 \cdot precision) + recall$

2) Accuracy
   a. $accuracy = TP/(TP + FP)$

The above metrics will show how well the predictions label the data. By using the different training algorithms and determining their ability for good predictions we should have a good idea of how much these algorithms can enhance the current IDS prediction abilities.