Network Traffic Behavioral analysis for detection of DDOS Attacks

Alma Lopez , Asha Mohan , Sukumaran Nair - Southern Methodist University

**Goal :** To Classify whether is was a benign or Attack( Types of Attack: SlowLoris, SlowHttpTest, Hulk, HeartBleed, GoldenEye, Botnet, DDOs).

**Data-Source/Data-Set:** CICIDS2017

**Data-Preprocessing:** There are 85 features, appropriate features were extracted using Recursive Feature Elimination (RFE) techniques. Also feature importance was checked as classification algorithms were run.

**Feature-Selection:** Correlation between the features were analyzed to eliminate features that contribute the same information about the data. As part of identification of important features to detect anomaly, random forest feature importance attribute was examined. For continuous variables, a correlation matrix with a Pearson’s correlation coefficient will define the linear relationship between two attributes.

**Resolving Imbalance DataSet:** Stratified Shuffle Split is a technique provided by scikit-learn that returns a stratified randomized split.

**ML-Modelling:** Multinomial Logistic Regression(MLR), K – Nearest Neighbor Classifier(KNN), Random Forest(RF), Naïve Bayes Classifier, Multi-Layer Perceptron, Dense Neural Network(DNN).

**Validation:** Precision, Recall, F1-Score, Confusion Matrix, false negative predictions rate of an attack, Receiver Operating Characteristics(ROC).

**Result:** Random forest model has high overall classification accuracy and low percentage of false negative predictions with an acceptable training time. Random Forest model outperforms all other algorithms for anomaly detection and can be trained in order of seconds.

**Analysis Performed:** They use flow based traffic characteristics to analyze the difference in pattern between normal vs anomaly packet.They evaluate several supervised classification algorithms using metrics like maximum detection accuracy, lowest false negatives prediction, time taken to train and run. They prove that decision tree based Random Forest is the most promising algorithm whereas Dense Neural network performs equally well on certain DDoS types but require more samples to improve the accuracy of low sampled attacks. The top flow based features that are helpful in identifying anomalies are total length of backward packets, average packet size, total length of forward packets, backward packets per second, flow duration and mean flow inter-arrival time.