Effective feature reduction for classifying IoT security threats using one class – SVM

Samanwickrama W.D.[a*], Munasinghe L.[b], Perera L.D.R.D.[a]

[a]Faculty of Applied Sciences, Wayamba University of Sri Lanka
[b]Software Engineering Teaching Unit, Faculty of Science, University of Kelaniya

*Corresponding author (email: dumindu@wyb.ac.lk)

Abstract

IoT usage has shown rapid increase in the smart world. It has estimated that the number of IoT devices will be over 5 billion by 2020. With the rapid increase in IoT usage, ensuring network security of IoT has become an immense challenging due to miniaturization of the IoT devices.

Launching bot attacks to steal information or create a network outage on tiny connected devices is not tough task for hackers. So here we try to build up light weight anomaly detection mechanism for tiny devices. We used a supervised machine learning algorithm for building a classifier for threat detection. First, we analyzed the feature importance for threat detection. Feature importance is determined using the Gini index which was obtained using Random Forest algorithm trained on training data. Second, we picked top n features interns of feature importance to train one class SVM which is used to detect potential attacks in test data. In this case we used n=20 which means 20 most important features to train one class SVM.

In this experiment, the Danmini Doorbell (DDb) data in UCI repository retrieved from *https://archive.ics.uci.edu* used to model the classifiers. The DDb dataset contains three types of web traffic data — benign traffic, Mirai traffic and Gafgyt traffic. Each record contains N=115 features which were generated by the publishers of the dataset using raw attributes of network traffic. As both Gafgyt and Mirai traffic produced by the attack activity, the two data-sources were combined to construct the overall set of malicious data for this experiment. For testing our hypothesis, we used Gafgyt and Mirai traffic data and benign traffic used for train the model.

Finally compared the results of threat classification (threat detection) accuracy of one class SVM trained with top n features and one class SVM trained with whole set of features (N). Resource utilization also tested using benchmark testing.

Comparing the total N feature classification against reduced n feature classification equal accuracy were obtained. 99% accuracy were obtained in n

feature model in figure 1. According to the benchmark testing the computational cost were reduced.

So the reduced feature set can be used in IOT anomaly detection. Hybrid approach should be employed for feature reduction. To removing a feature from classification should use domain expertise and statistical analysis together. So this remains for future enhancement.

Table 1: Confusion Matrix of n feature model

Prediction	False	True
False	316650	2025
True	0	7884

Keywords: Anomaly detection; IoT; Machine learning; Network security

Acknowledgement: We are grateful to Dr. Harsha Kalutarage at the School of Computing Science and Digital Media, Robert Gorden University, UK for his valuable advice and comments on this research work.