Efficacy of using SVM on Iot Security Threat Identification under Reduced Feature Set Architecture Compared to J48 and Random Forest Algorithms using UCI Benchmark Data Set

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

More than 5 billion IoT devices are on the Internet due to smart technology development in 2020. Most of them are connected to the Internet and they will create an easy gateway to hackers. So the network security has become an immense challenge. Maintaining minimal resource consumption of IoT devices is also difficult. Here we try to build up light weight anomaly detection mechanism for smart devices. Supervised machine learning algorithms were used for building a classifier for threat detection and their performances were compared.

Danmini Doorbell (DDb) data in UCI repository is used in this experiment. 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. In order to test the hypothesis, we used Gafgyt and Mirai traffic data and benign traffic used to train the model for SVM.

Feature importance is determined using the Gini index which was obtained using Random Forest algorithm. We picked top n=20 from the total N features to train the one class SVM, J48 and Random forest which is used to detect potential attacks in test data. Comparing the total N feature classification against reduced n feature classification equal accuracy were obtained in each algorithm. According to the benchmark testing the computational cost were reduced in each algorithm. However, when comparing other algorithms with SVM, the SVM had the lowest accuracy and performance (Table 1)

Table 1. Algorithm comparison results

Algorithm	Mean time (s)
Random Forest	1.640045
KSVM	184.698795
J48	1.046536

SVM can be trained using benign traffic but others need malicious traffic to train the system. Although SVM shows less accuracy and high resource consumption it can be used to identify unknown threats. So the SVM is the only algorithm among these three algorithms that can be employed to identify zero day attack. These results can be used the reduced feature set in IoT anomaly detection. It should be further investigated with different data sets and different algorithms. This becomes an area for potential research in future.

Keywords: Anomaly detection, IoT, Machine learning, Network security, UCI

Acknowledgements: Authors 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