Machine Learning-Based Cross-Dataset Validation Framework for IoT Network Intrusion Detection

Simran Kaur Hora
Department of Information Systems, Vilnius Gediminas Technical University (VILNIUS TECH), Lithuania

Nikolaj Goranin
Department of Information Systems, Vilnius Gediminas Technical University (VILNIUS TECH), Lithuania

Antanas Čenys
Department of Information Systems, Vilnius Gediminas Technical University (VILNIUS TECH), Lithuania

Abstract:
The development of intrusion detection systems for IoT environments heavily relies on the availability of labeled network datasets. Nowadays, many datasets are being generated using emulated IoT devices due to the high cost and complexity associated with deploying large-scale actual IoT devices. The practical utility of datasets generated employing emulated IoT devices is commonly validated using training and testing splits from the same dataset. Such within-dataset results primarily demonstrate how well a model adapts to that dataset generated using emulated IoT devices. However, it does not adequately reflect how models trained on datasets generated employing emulated devices will perform when applied to real-world IoT environments. To address this research gap, this paper proposes a cross-dataset validation framework to check whether the learned patterns from emulated IoT device datasets remain effective across actual IoT network datasets. The framework is instantiated with machine-learning models and applied to the EmuIoT-VT dataset and the performance of the models is reported using standard evaluation metrics.

Keywords:
Internet of Things (IoT), Cybersecurity, Anomaly Detection, IoT Network Dataset, Intrusion Detection.