IOT Device Identification Using Machine Learning Techniques

Idan Mosseri, Dean Eckert

# Abstract

In this work we apply machine learning algorithms on IOT network traffic data for accurate identification of devices. In order to train and evaluate the classifier, data was collected from different IOT devices. Using supervised learning, we trained a classifier being able to distinguish between a specific use and the rest. The overall IoT classification accuracy of our model is 99.281%.

# Introduction

The “Internet of Things” (IOT) relates to networks of physical devices and items embedded with electronics, sensors, actuators, software and connectivity which enables the communication between these devices and their exchange of data.

In recent years, more organizations allow IOT devices to be connected to their networks which might impose a security threat to these networks. Hence, organizations must be able to identify which devices are connected to their networks and whether these devices are considered legitimate and do not impose a risk.

Leveraging network traffic in order to identify devices in general has been gaining in popularity in previous works. Specifically, there is an increasing interest in the domain of IOT device identification due to the importance of identifying such devices in an organizational environment (in terms of security).

In this work, we address the challenge of identifying an IOT device by analyzing its high-level network traffic data using machine learning techniques. We would like to develop a method for identifying such device, even if its IP address has been spoofed (which can be done easily) and to allow us identify an abnormal behaviour which may indicate which device is in-use.

Since we can’t rely on the IP address as an identifier of the device (since this value can be spoofed), we would like to analyze the traffic’s high-level data (which means the metadata and traffic statistics, rather than analyzing the content).

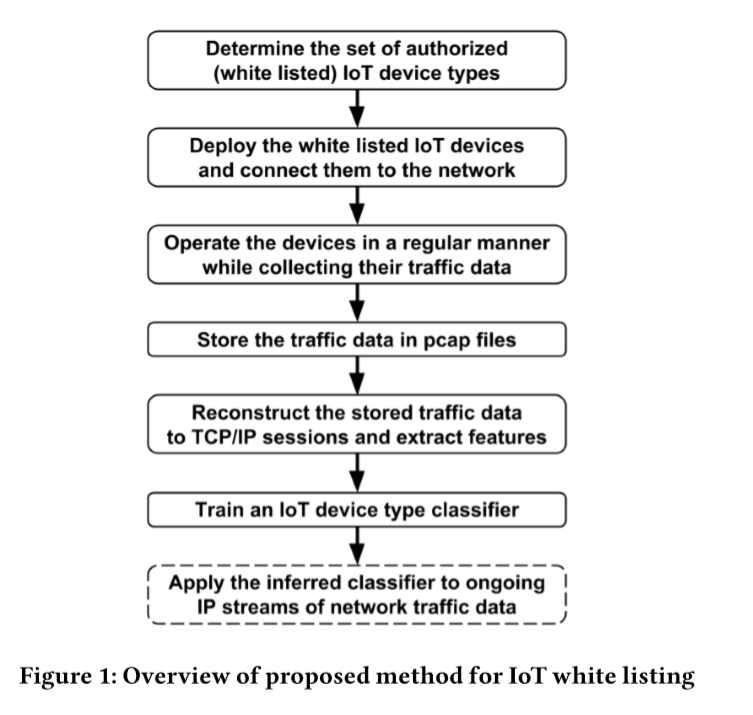
**Research questions:** This research proposes a novel approach in order to identify IOT devices using their network traffic data and analysis made using machine learning techniques. Specifically, we focus on the following question: Is it possible to identify an IOT device using machine learning techniques by analyzing its high-level network traffic data?

**Summary of contributions:**

# Related Works

Previous work has already been conducted about the topic of identifying an IOT device in an organizational environment. The most relevant work in this area is described in a paper about a system called ProfilIOT [1]. The paper describes an experimental environment in which network traffic data was collected from different nine devices (where seven of them were IOT devices). Following the data collection stage, the data was transformed into the form of sessions (TCP connection from its SYN packet to the FIN packet) and a meta-classifier was trained and evaluated. The meta-classifier used session-based classifier which predicts the probability that a given session was originated from this specific device. Following a definition of a threshold, the classifier was also being able to predict whether it was actually originated from the device or not. Eventually, the meta-classifier used these device-specific classifiers in order to predict the device in which a sequence of sessions was originated from. This work used some specific machine learning models like Random Forest, GBM and XGBoost.

In another paper [2], a similar mechanism was used for the detection of unauthorized IOT devices in the network using machine learning techniques. Traffic was collected from 17 distinct IOT devices, representing 9 types of IOT devices. Based on a classifications of a 20-sessions sequences and the use of the majority rule the classifier managed to understand whether the device was part of the whitelisted devices. The machine learning model used in this paper was the Random Forest model. The general method proposed in the paper as follows:



As can be seen in the diagram above, the whitelisted devices should be defined in advance, learn the “normal” behavior of these devices and training a relevant classifier. Following this training stage, the classifier will be able to determine whether the next ongoing IP streams are of a legitimate device or not.

It is also worth mentioning that there are previous works related to the usage of network traffic data analysis in general and not only in relation to the IOT domain. An example for such work is a paper describing a method of unknown malware detection using network traffic classification [3]. In this paper, a method using a supervised machine learning techniques is described and the way it can be used for detection of malicious communication like the interaction with a command and control servers. The solution was based on different layers of the network stack and different protocols. The paper is useful for understanding the different approaches of network data analysis like packet-level vs flow-level analysis, port-based attributes vs payload based attributes vs statistical based attributes etc.

# Proposed Method

# Evaluation

# Results

# Discussion

# Appendix

1. ProfilIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis - Yair Meidan , Michael Bohadana , Asaf Shabtai , Juan David Guarnizo , Mart´ın Ochoa , Nils Ole Tippenhauer , and Yuval Elovici (2017)

2. Detection of Unauthorized IoT Devices Using Machine Learning Techniques - Yair Meidan, Michael Bohadana, Asaf Shabtai, Martin Ochoa, Nils Ole Tippenhauer, Juan David Guarnizo, Yuval Elovici (2018)

3. Unknown Malware Detection Using Network Traffic Classification - Dmitri Bekerman, Bracha Shapira, Lior Rockach, Ariel Bar (2015)