Internship proposal 2016/2017

Topic: Detecting anomalies in home IoT devices

Duration: 4 to 6 months

Hosting team: Muse, Inria Paris (https://team.inria.fr/muse/)

Contact: renata.teixeira@inria.fr

Mentor:
Vassilis Christophides, Senior Researcher, Inria
Renata Teixeira, Directeur de Recherche, Inria

Keywords: IoT, device discovery, anomaly detection, Internet measurements, network traffic

Description:

Home networks consist of a plethora of vulnerable devices, and securing them often becomes impossible for a normal user. Currently, smart device owners are required to configure each device separately without understanding what data the device will communicate over the network, the security mechanism it will use, and how multiple devices will interact with each other. Furthermore, due to the heterogeneous nature of IoT devices, users need to manage each device separately and ensure separate credentials. This results in sacrificing security for ease of accessibility. Our long-term goal is to design a software-defined IoT platform that can manage a network of off-the-shelf IoT devices with a limited level of administrator support. The proposed system involves a controller that (1) collects measurements, such as discovering the devices and monitoring the traffic, (2) makes decisions, based on high-level security policies and device network profiles extracted from the measured data, and (3) configures the network to enforce security policies and adapt to the prevailing conditions.

The goal of this internship is to develop new mechanisms for improving the security of the Internet of Things at home. In particular, we plan to tackle research problems in the discovery and fingerprinting of network-connected devices and the detection of anomalous "inbound" and "outbound" traffic events.

Device discovery and fingerprinting. Detecting anomalies and analyzing the device footprint first requires being able to determine what devices are connected to the network, how they are connected (e.g., via what physical medium), and where they are connected (e.g., directly to the Internet, indirectly via a hub inside the home). Unfortunately, it is currently difficult to identify the devices that are connected to the network, due to the sheer diversity of the devices, and the lack of standard discovery mechanisms. In lieu of a standard discovery mechanism, we propose to develop classification algorithms to identify connected devices based on various network-level features and traffic patterns. A combination of features, ranging from a device's hardware/MAC address to certain traffic patterns (e.g., the set of destinations that the device regularly contacts) may ultimately serve as useful features for identification. The research will involve both (1) identifying the features that are useful for representing devices and (2) developing classification algorithms that can accurately identify these devices using lightweight, real-time algorithms.

Anomaly detection. We intend to build on our capabilities for accurately discovering and fingerprinting devices to characterize traffic patterns for these devices and automatically detect anomalies. A main focus of our work on anomaly detection will involve developing the appropriate taxonomy of devices and usage patterns to allow us to develop models for use in anomaly detection. A second aspect of detection will involve characterizing the behavior of various types of devices in light of different usage patterns. For example, devices may generate different traffic profiles based on whether the user is present (i.e., whether the user is home or away) and what they are doing at any given time (e.g., a motion-activated camera may generate more data when a user is actively triggering it). We will develop models for "normal" traffic from various devices by learning individual user patterns and detecting anomalies (e.g., a device is generating more traffic than it otherwise would at a particular time of day) and other discrepancies across devices (e.g., general traffic patterns suggest that a user is not present, but one device is generating an abnormal amount of traffic indicating user interaction). Features that we may use may include the frequency of data transmission from a device and the set of destinations that a device contacts during a particular time interval. In the future, we will apply machine learning techniques we developed for device fingerprinting to anomaly detection.