Crowdsourcery: Defence Against Dark Artefacts in Smart Homes

Vadim Safronov

University of Cambridge
vs451@cam.ac.uk

With the advance of ubiquitous computing and communication technologies over the last few decades, IoT devices have become more independent and autonomous.This tendency is especially noticeable in smart home environment where almost every IoT device reports data to the cloud and can automatically trigger a variety of smart home functions.

Despite the benefits which a smart home claims to bring to the end user, there are a number of serious security gaps in its network infrastructure. The lack of isolation between groups of IoT devices brings ones home network to severe security threats. If an adversary hack the weakest link in the network of devices, they can gain total control over the smart home functions turning the owner into a victim in their own house.

The DADA (Defence Against Dark Artefacts) project, initiated by Cambridge, Imperial and Nottingham universities, addresses these challenges by:

- designing and implementing mechanisms for device traffic monitoring with a precise look at packet traces and device profiles;
- application of learning technologies to detect devices' abnormal behavior;
- introducing techniques to deal with traffic anomalies and restoring home network operability;
- putting the homeowner in the center of management by informing them on possible security threats and offering a choice of mitigation options.

The above DADA objectives raise a number of important technical questions. First, to apply learning techniques, it is necessary to collect sufficient behavioral statistics about an IoT device in order to detect traffic anomalies. Second, a decent number of participants (i.e. smart home owners) is required to collect enough device traces in order to form training data sets.

Inspired by these technical challenges, we are currently investigating how crowdsourcing (which is successfully applied in wiki communities) can help the project in collection of device statistics needed for the application of learning algorithms. Being a part of DADA solution, a crowdsourcing platform can solve the problem with participants and data sets. By collaborative work and information exchange between smart home owners and industrial partners the platform could gain enough traffic traces in a reasonable time to form training data sets for each considered device type. Another key objective of this platform is to ensure the absence of private information leakage while training models on the collected data sets.

As the DADA project is currently in its start stage, we would appreciate initial feedback on the proposed concept. We will present the role of crowdsourcing in DADA, show relevant use-cases, demo the current project state and discuss possible future steps.