Poster Abstract: "Sensing" the IoT Network: Ethical Capture of Domestic IoT Network Traffic

Diana Andreea Popescu
Vadim Safronov
Poonam Yadav
first.last@cl.cam.ac.uk
University of Cambridge

Roman Kolcun
Anna-Maria Mandalari
Hamed Haddadi
first.last@imperial.ac.uk
Imperial College

Derek McAuley
derek.mcauley@nottingham.ac.uk
University of Nottingham

Richard Mortier
richard.mortier@cl.cam.ac.uk
University of Cambridge

ABSTRACT

As more and more devices are connected to the Internet-of-Things, often made by non-specialist companies or short-lived startups, the likelihood that these devices will be hacked and used for nefarious activity online increases. We seek to support non-expert users in managing the network behaviour of their IoT devices, and assisting them in handling the cases where those devices are hacked. To do so, we wish to enable anomaly detection at the network level, determining when a device starts behaving unusually. This requires capturing data about how devices behave in a diverse range of real deployments, not just lab environments.

To that end, we present *IoTCrowdsourcery*, a toolset for capturing traffic data from real-world IoT deployments. Participants collect packet traces from their IoT devices through our software, and provide them via a crowdsourcing infrastructure. The key challenges to overcome are to make the process straightforward enough for non-expert participants to carry out, and to ensure that legal (notably GDPR) and ethical issues are carefully handled by ensuring that participants understand what they are doing, and are provided with various means to exercise agency in participating, and ultimately to withdraw their participation if they wish. We envisage the captured traces being analysed to develop behavioural models of IoT devices which will be used for anomaly detection, improving the security of our smart homes and more generally of the Internet.

ACM Reference Format:
Diana Andreea Popescu, Vadim Safronov, Poonam Yadav, Roman Kolcun, Anna-Maria Mandalari, Hamed Haddadi, Derek McAuley, and Richard Mortier. 2019. Poster Abstract: "Sensing" the IoT Network: Ethical Capture of Domestic IoT Network Traffic . In *Proceedings of ACM Conference.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Conference, ,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 MOTIVATION

In recent years, considerable research has targeted understanding IoT network vulnerability by collecting network traffic traces to analyse for features such as network usage, privacy breach, and service dependency [2, 6]. In contrast to much of the previous work, we are concerned with developing behavioural models of devices against which we can perform anomaly detection in order to determine when devices should have their network activity actively managed or curtailed. This entails capturing traces of devices in situ, and recording their behaviour in the wide range of deployment contexts they will experience. Given the potentially personal data that might inadvertently be revealed by IoT devices, we must be careful that those submitting packet traces are fully aware of their actions and have multiple opportunities to change their mind, including stopping ongoing collection and withdrawing previously submitted traces.

To meet these challenges, we are developing *IoTCrowdsourcery*, a set of tools to enable us to capture data from real-world deployments of IoT devices (§2). IoTCrowdsourcery enables us to engage non-expert participants to perform longitudinal data collection and upload while ensuring those participants remain in full control of what data is collected and uploaded.

2 SYSTEM WORKFLOW

The end-to-end workflow of our system is presented in Figure 1. Devices connect to a home router for local and Internet connectivity. To ensure coverage of both local network and Internet carried interactions such as between a device and the controller, we thus need to intercept traffic at the home router. The IoT device whose traffic a user wants to monitor is moved to the wireless network advertised by the IoTCrowdsourcery Raspberry Pi 3B+ (subsequently denoted by RPi) along with the device controller, while all other devices remain unaffected and unmonitored. In contrast, IoT Inspector's [4] use of a laptop or similar to intercept traffic via ARP spoofing requires more and ongoing interaction to ensure that longitudinal data collection is still taking place. As we target network trace collection of devices in real domestic deployments, we must take particular account of different technical backgrounds which participants will have [5]. *Non-expert users* may be experienced in deploying and managing IoT devices in their homes. To address

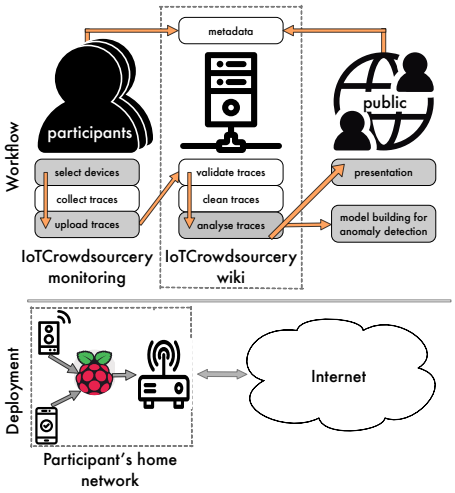

Figure 1: IoTCrowdsourcery workflow and deployment. Participants contribute packet traces from selected devices, alongside metadata describing the particular devices in question. Traces are uploaded to the wiki, validated, cleaned by removing private user application data, and analysed. The results of the analyses are made available, whether for human consumption via the relevant wiki page, or for the construction of models for anomaly detection. The shaded boxes indicate points where participants have explicit control over what is done with their data: they select precisely which devices are captured, they choose whether or not to upload traces, permission is sought to carry out each new analysis added to the wiki on their traces, and they can subsequently decide to remove their traces from the system after seeing their presentation.

this, we make the observation that one interaction that many non-expert users can certainly carry out is to connect their IoT devices to their network; else they would be unable to use such devices. We apply this observation by providing a custom RPi Linux image that configures the RPi to provide the IoTCrowdsourcery Wi-Fi network routed to the default home router over the wired interface while capturing all traffic observed on both its wired and wireless interfaces. Participants connect to this network the devices they are willing to have monitored by our system.

3 SYSTEM ARCHITECTURE

The two main components in our system are: (1) LinuxKit Image running IoTCrowdsourcery deployed on a RPi, and (2) IoTCrowdsourcery Wiki. We have deployed the Wiki platform running on Apache2 with MySQL backend on a modestly configured Ubuntu 18.04 VM allocated a 1.8 GHz CPU, 2GB RAM, 16GB disk space and shared 2x10 Gbps link speed.

Our deployment consists of a purpose-built Linux image that has routing capabilities and traffic capture mechanisms, and it is deployed on a RPi. The Linux image is built using LinuxKit, a toolkit for building customised, immutable, and minimal Linux distributions [1]. The advantage of using LinuxKit is that the image incorporates the minimal set of desired functions, improving efficiency and security. Our custom image targets non-expert users. A user can simply burn the image on a microSD card which, when plugged into the RPi, causes the RPi to become an IoTCrowdsourcery measurement node. We build our IoTCrowdsourcery LinuxKit image for the ARMv8 architecture. The LinuxKit configuration file specifies the set of packages that are included in the image, which comprises the package that offers Ethernet connection and the Wi-Fi network for the devices that are monitored, along with routing functionality to forward the traffic appropriately, and the collection script for capturing network traffic. The current total image size comprising the kernel and the initial RAM disk is 112MB, and the configuration has nine packages.

The RPi is equipped with a Wi-Fi card supporting 2.4 GHz and 5 GHz 802.11 b/g/n/ac; our current setup uses 2.4GHz 802.11g. We generated traffic using *iperf* from a laptop connected to the wireless network of the RPi. Simultaneously, we ran the network collection script on the RPi to check that it is capable of capturing the incoming traffic and writing it to storage. In our preliminary experiment, we measured network bandwidth up to 21.9 Mbps that can be handled by the collection script using *tcpdump* on LinuxKit. For comparison, we performed a similar experiment on the Raspbian [3] operating system, where the peak network bandwidth was just below 20 Mbps.

4 CONCLUSION AND FUTURE WORK

We presented a toolset for capturing traffic data from real-world IoT deployments. Currently, the features of each trace are extracted on our wiki servers. We envision that, in the future, the analysis could be done directly on the monitoring box. The participant would be able to see the results before any upload takes place, giving them greater control over what is shared, and also how the shared data is treated. Extracted features will be added to the appropriate device dataset, and learning models will be trained using these features. The extracted features will be compared against aggregated device behavioural statistics built on previous analysis of similar devices within similar smart-home settings.

Acknowledgements. This work is supported by the grant EPSRC EP/R03351X/1. The authors would like to thank the LinuxKit community for answering questions about LinuxKit.

REFERENCES

[1] Developers. [n. d.]. LinuxKit. https://github.com/linuxkit/linuxkit. ([n. d.]). [Online; accessed August 2019].
[2] H. Haddad Pajouh *et al.* 2019. A Two-layer Dimension Reduction and Two-tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks. *IEEE Transactions on Emerging Topics in Computing* (2019), 1–1. https://doi.org/10.1109/TETC.2016.2633228
[3] Raspberry Pi Foundation. [n. d.]. Raspbian. https://www.raspberrypi.org/downloads/raspbian/. ([n. d.]). [Online; accessed August 2019].
[4] Danny Y. Huang, Gunes Acar, Noa Apthorpe, Frank Li, Arvind Narayanan, and Nick Feamster. [n. d.]. IoT Inspector. https://iot-inspector.princeton.edu/. ([n. d.]). [Online; accessed May 2019].
[5] Mortier, Richard *et al.* 2012. Homework: Putting Interaction into the Infrastructure. In *Proceedings of the 25th ACM UIST'12 (UIST '12)*. ACM, New York, NY, USA, 197–206. https://doi.org/10.1145/2380116.2380143
[6] Poonam *et al.* Yadav. 2019. Network Service Dependencies in Commodity Internet-of-things Devices. In *Proceedings of the IoTDI '19 (IoTDI '19)*. ACM, New York, NY, USA, 202–212. https://doi.org/10.1145/3302505.3310082