Detecting Consumer IoT Devices Through the Lens of an ISP

Said Jawad Saidi
Max Planck Institute for Informatics

Anna Maria Mandalari
Imperial College London

Hamed Haddadi
Imperial College London

Daniel J. Dubois
Northeastern University

David Choffnes
Northeastern University

Georgios Smaragdakis
TU Berlin and Max Planck Institute for Informatics

Anja Feldmann
Max Planck Institute for Informatics/Saarland University

ABSTRACT

Internet of Things (IoT) devices are becoming increasingly popular and offer a wide range of services and functionality to their users. However, there are significant privacy and security risks associated with these devices. IoT devices can infringe users' privacy by ex-filtrating their private information to third parties, often without their knowledge.

In this work we investigate the possibility to identify IoT devices and their location in an Internet Service Provider's network. By analyzing data from a large Internet Service Provider (ISP), we show that it is possible to recognize specific IoT devices, their vendors, and sometimes even their specific model, and to infer their location in the network. This is possible even with sparsely sampled flow data that are often the only datasets readily available at an ISP. We evaluate our proposed methodology [1] to infer IoT devices at subscriber lines of a large ISP. Given ground truth information on IoT devices location and models, we were able to detect more than 77% of the studied IoT devices from sampled flow data in the wild.

ANRW '21, July 24–30, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8618-0/21/07…$15.00
https://doi.org/10.1145/3472305.3472885

CCS CONCEPTS

• **Security and privacy** → *Network security*; • **Networks** → **Network measurement**; *Network properties.*

KEYWORDS

Internet of Things, IoT detection, IoT security and privacy, Internet Measurement

1 OUR APPROACH

The number of IoT devices is expected to grow exponentially in the next years [2]. IoT devices typically rely on cloud infrastructures to offer their services. While doing so, they may expose information, including their destinations [3]. Internet Service Providers (ISPs) are developing strategies for dealing with the large-scale coordinated attacks from these devices. Identifying IoT devices in the network is useful to block attacks, isolate vulnerable devices, and inform their users [4]. On the other hand, due to the presence of several middleboxes, and traffic sampling at ISPs, it is challenging to identify and isolate the misbehaving devices among millions of connected IoT devices at subscribers premise [5].

Proposed solutions, either rely on DNS data [6] that raise privacy concerns, or on in-situ scans by anti-virus software that are not scalable [7]. In this paper we describe a methodology for detecting the presence of IoT devices at subscriber lines at scale, using sparsely sampled flow captures (i.e., NetFlow [8]). We set up two testbeds consisting of 56 different IoT products from 40 manufacturers across six different categories to tackle this challenge. We first identify backend infrastructures for many IoT devices, using DNS queries, web certificates, and banners. We then use the traffic signatures to identify broadband subscriber lines using IoT devices. We apply our methodology to a large residential ISP in Europe.

Results show that IoT devices typically use internet backend infrastructures to offer their services and it is possible

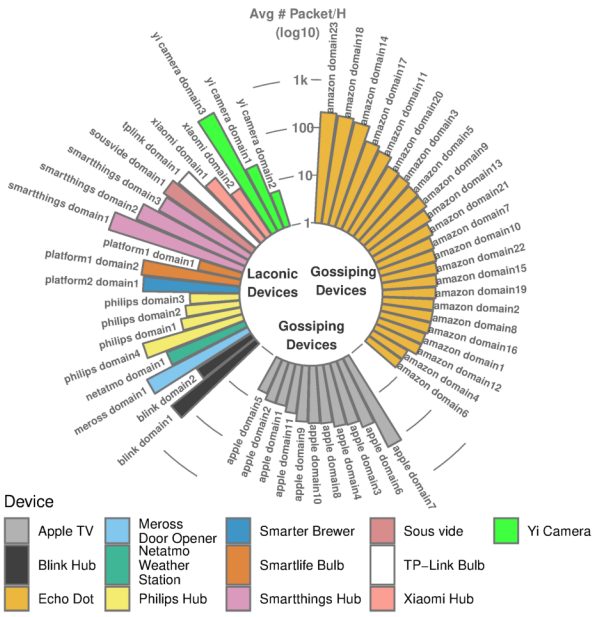

Figure 1: Circular bar plot of average # of packets/hour per domain (log y-scale). The domains belong to 13 IoT devices and separated into three groups: one for laconic and two for gossiping devices (Echo Dot and Apple TV).).

to identify IoT devices even if the device is not actively used. Using our methodology, we recognized that 20% of 15 million subscriber lines used at least one of the IoT products we analyzed. We were able to detect the presence of devices from 77% of our target devices within hours, sometimes minutes.

2 GENERATING IOT SIGNATURES

In order to identify IoT devices, we first use controlled experiments, where we tunnel the traffic of two IoT testbeds to an ISP. This generates ground truth IoT traffic within the ISP. We identify IoT devices signatures using DNS queries, IP addresses and port numbers. We then apply the signatures to the flows captured from the ISP.

The circular bar plot in Figure 1 depicts the average number of packets/hour per domain for 13 devices when they are in their idle state. We observe that most devices are supported by their own set of domains and for many IoT devices, this is a small set containing less than 10 domains. We classify those as *Laconic devices*. Other devices gossip and communicate with more than ten domains, we name those *Gossiping devices*. However, not all of these domains can be used to generate signatures. We filter out **generic domains**, i.e., domains that are not primarily used by IoT devices and can be contacted by a wide range of users and humans, e.g., wikipedia.com or domains that belong to Content Delivery Networks (CDNs) or shared infrastructure. Considering that an IP address belonging to a shared infrastructure

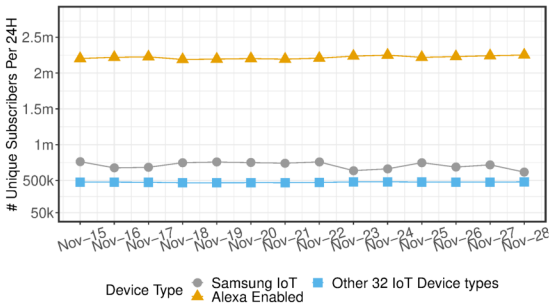

Figure 2: ISP subscriber lines with IoT activity daily trends (Alexa Enabled, Samsung IoT, and others).

may serve many domains, we use passive DNS dataset [9], similar to the methodology in [10], to identify the domains and, consequently, IP addresses that belong to the shared infrastructure.

3 DEVICES IN-THE-WILD

We apply our methodology on NetFlow data from a large European ISP. The ISP has over 15 million subscribers and does not deploy carrier-grade NAT. We focus on the two weeks of data from November 15-28, 2019. Using our ground truth dataset we first check how long it takes for our methodology to detect the presence of the IoT devices we test. On average, by requiring the evidence of at least 40% of domains, we are able to detect 72/93/96% of IoT devices that are detectable at manufacturer or product level within 1/24/72 hours in the active mode. Even in idle mode their the percentage is 40/73/76% with 1/24/72 hours.

Figure 2 shows the number of ISP subscriber lines for which we inferred IoT-related traffic. Even though multiple IoT devices can reside at an ISP subscriber, we count each subscriber only once. The figure shows the daily IoT-related activities of subscriber lines. Alexa Enabled device is any device that responds to Alexa Voice Service commands. For roughly 20% of subscriber lines, we detected activity from at least one of the devices in our testbed. Our results show activity related to Alexa Enabled devices for roughly 14% of the subscriber lines.

4 FUTURE WORK

To foster further research in this area, we make all the signatures available at https://moniotrlab.ccis.neu.edu/imc20/.

After detecting the presence of IoT devices our next step is to identify the non-essential traffic generated by them, at subscriber lines at scale. We will define general rules for blocking unnecessary destinations in smart home environments [11], in a manner similar to ad blockers in browsers. The goal of our methodology is to monitor non-essential traffic trends over time and detect any common non-required destinations among different IoT devices even in large residential ISPs.

ACKNOWLEDGEMENTS

The research in this paper was partially supported by the European Research Council (ERC) Starting Grant ResolutioNet (ERC-StG-679158), the EPSRC (Databox EP/N028260/1, DADA EP/R03351X/1, HDI EP/R045178/1, and Impact Acceleration Account (IAA)), the NSF (BehavIoT CNS-1909020, ProperData SaTC-1955227), and Consumer Reports (Digital Lab Fellowship for Daniel J. Dubois).

REFERENCES

[1] Said Jawad Saidi, Anna Maria Mandalari, Roman Kolcun, Hamed Haddadi, Daniel J. Dubois, David Choffnes, Georgios Smaragdakis, and Anja Feldmann. A Haystack Full of Needles: Scalable Detection of IoT Devices in the Wild. In *ACM IMC*, 2020.

[2] Cisco. Cisco Annual Internet Report (2018âĂŞ2023) White Paper. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html, 2020.

[3] J. Ren, D. J. Dubois, D. Choffnes, A. M. Mandalari, R. Kolcun, and H. Haddadi. Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. In *ACM IMC*, 2019.

[4] O. Çetin, C. Gañán, L. Altena, T. Kasama, D. Inoue, K. Tamiya, Y. Tie, K. Yoshioka, and M. van Eeten. Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai. In *NDSS*, 2019.

[5] H. Haddadi, V. Christophides, R. Teixeira, K. Cho, S. Suzuki, and A. Perrig. Siotome: An edge-isp collaborative architecture for iot security. In *1st International Workshop on Security and Privacy for the Internet-of-Things (IoTSec)*, 2018.

[6] R. Perdisci, T. Papastergiou, O. Alrawi, and M. Antonakakis. IoTFinder: Efficient Large-Scale Identification of IoT Devices via Passive DNS Traffic Analysis. In *IEEE European Symposium of Security and Privacy*, 2020.

[7] D. Kumar and K. Shen and B. Case and D. Garg and G. Alperovich and D. Kuznetsov and R. Gupta and Z. Durumeric. All Things Considered: An Analysis of IoT Devices on Home Networks. In *USENIX Security Symposium*, 2019.

[8] B. Claise. RFC 3954: Cisco Systems NetFlow Services Export Version 9, 2004.

[9] Farsight Security. DNSDB. https://www.dnsdb.info/, 2017.

[10] C. Iordanou, G. Smaragdakis, I. Poese, and N. Laoutaris. Tracing cross border web tracking. In *ACM IMC*, 2018.

[11] A. M. Mandalari, D. J. Dubois, R. Kolcun, M. T. Paracha, H. Haddadi, and D. Choffnes. Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. In *Privacy Enhancing Technologies Symposium (PETS)*, 2021.