Information Exposure From Consumer IoT Devices:

A Multidimensional, Network-Informed Measurement Approach

Jingjing Ren
Northeastern University

Daniel J. Dubois
Northeastern University

David Choffnes
Northeastern University

Anna Maria Mandalari
Imperial College London

Roman Kolcun
Imperial College London

Hamed Haddadi
Imperial College London

ABSTRACT

Internet of Things (IoT) devices are increasingly found in everyday homes, providing useful functionality for devices such as TVs, smart speakers, and video doorbells. Along with their benefits come potential privacy risks, since these devices can communicate information about their users to other parties over the Internet. However, understanding these risks in depth and at scale is difficult due to heterogeneity in devices' user interfaces, protocols, and functionality.

In this work, we conduct a multidimensional analysis of information exposure from 81 devices located in labs in the US and UK. Through a total of 34,586 rigorous automated and manual controlled experiments, we characterize information exposure in terms of destinations of Internet traffic, whether the contents of communication are protected by encryption, what are the IoT-device interactions that can be inferred from such content, and whether there are unexpected exposures of private and/or sensitive information (e.g., video surreptitiously transmitted by a recording device). We highlight regional differences between these results, potentially due to different privacy regulations in the US and UK. Last, we compare our controlled experiments with data gathered from an *in situ* user study comprising 36 participants.

1 INTRODUCTION

Consumer Internet of Things (IoT) devices are gaining popularity (expected to number 20 billion by 2020 [13]), offering services such as personal digital assistants, home security, and climate control. By combining rich sensors (e.g., cameras, microphones, motion sensors) and Internet connectivity, these devices have the potential to learn and expose extensive information about their users and their surrounding environment. Much of this information has major privacy implications, e.g., when devices surreptitiously record audio [9, 48] and user's TV viewing habits [32], then share this information over the Internet with device manufacturers and unknown third parties in different countries with different privacy regulations. As most of these devices lack any interfaces that indicate information exposure, there is an urgent need for research that provides transparency into such exposure at scale, and that identifies relevant privacy implications within different jurisdictions.

IMC '19, October 21–23, 2019, Amsterdam, Netherlands

ACM ISBN 978-1-4503-6948-0/19/10...$15.00
https://doi.org/10.1145/3355369.3355577

There are several key challenges that limit our understanding of information exposure from IoT devices and their privacy implications. *First*, IoT device ecosystems are generally closed, hence ground truth about information exposure is not readily available. In particular, for the vast majority of devices in our experiments, it is not feasible to modify the device firmware (e.g., to conduct taint tracking) and/or use man-in-the-middle techniques to decrypt TLS connections (e.g., to identify personal information exposed in plaintext). In the absence of ground truth, we must develop strategies to analyze information exposure; namely, we focus on using inferences based on information contained in (potentially encrypted) network traffic flows. *Second*, characterizing IoT information exposure at scale is cumbersome: it requires manually setting up large numbers of devices, using carefully controlled interactions with them, and capturing the salient network traffic they generate. Unlike in the mobile and web environments, where information exposure analysis is facilitated by existing emulation and automation tools, there is a need for new techniques for automating IoT device experiments, gathering their data, and analyzing them. *Third*, all previous studies focused on information exposure based on interactions with IoT devices from a research team in one jurisdiction. There is a need to understand how the same devices behave in jurisdictions with different privacy laws, and when used by larger numbers of users.

In this paper, we address these challenges by providing an in-depth analysis of information exposure from 81 devices located in our labs in the US and UK. Through rigorous automated and manual controlled experiments, we characterize information exposure in terms of destinations of Internet traffic, whether the contents of communication are protected by encryption, what are the IoT device interactions that can be inferred from such contents, and whether there are unexpected exposures of private and/or sensitive information (e.g., video surreptitiously transmitted by a recording device). Further, we determine whether there are regional differences between these properties, as the privacy regulations in the US (enforced by the FTC) and UK (GDPR) can have substantial impact on data collection. Last, we compare our controlled experiments with data gathered from an *in situ* uncontrolled experiment comprising 36 participants.

The highlights of our research findings include the following. Using 34,586 controlled experiments, we find that 72/81 devices have at least one destination that is not a first party (i.e., belonging to the device manufacturer), 56% of the US devices and 83.8% of the UK devices contact destinations outside their region, all devices expose information to eavesdroppers via at least one plaintext flow, and a passive eavesdropper can reliably infer user and device behavior from the traffic (encrypted or otherwise) of 30/81 devices.

Figure 1: US IoT lab (Northeastern), configured like a studio apartment that contains a large set of consumer IoT devices.

To understand how the devices behave outside of controlled experiments, we made the US lab devices available to 36 participants for six months in an IRB-approved user study, where such devices are used in a studio-apartment setting (see Fig. 1). At the end of the study, we found cases of devices exhibiting unexpected behaviors. For example, a video doorbell sends video recordings to its service provider based on movement sensors, without any notification or consent from recorded parties.

Our key research contributions include:

- Analysis of what we believe is the largest collection of popular consumer IoT devices to date.
- Semi-automated controlled experiments that enable device analysis at scale, plus six-months of uncontrolled experiments as part of an IRB-approved study.
- The first apples-to-apples comparison of device behavior in jurisdictions subject to different privacy laws.
- Using the above testbed features to analyze destinations of network traffic, measure what information is exposed to other parties over the Internet, evaluate how well device interactions can be predicted based on network traffic.
- Analyzing idle traffic to detect unexpected device activity.

To facilitate additional research into consumer IoT devices, and to help scale analysis to additional jurisdictions, we make our IoT measurement and analysis code and data from controlled experiments publicly available at `https://github.com/NEU-SNS/intl-iot`.

2 DEFINITIONS AND GOALS

This work measures network activity and corresponding information exposure (defined in §2.1) from popular consumer IoT devices. In particular, we focus on characteristics of the destinations of their IP traffic, whether such traffic is protected via encryption, and what are the potential privacy implications of this exposure. We define our key research questions and privacy concerns in §2.2.

2.1 Definitions

Information exposed by IoT devices. For the purpose of this study, we define three categories of information that can be exposed by IoT devices. This list is not exhaustive; rather, we focus on baseline information exposure detectable using controlled experiments and network traffic analysis.

- *Stored data.* This can include device identifiers and personally identifiable information given by the user during device activation, activity logs, device state, etc.
- *Sensor data.* This consists of information obtained by the sensors of an IoT device, e.g., motion detection, video surveillance footage, audio recording.
- *Activity data.* This comprises information about how a user interacted with a device (e.g., via an app on a mobile device or a button on the IoT device) and what functionality of the device has been used (e.g., toggling a light).

Parties to which information is exposed. When information is exposed by an IoT device, it is explicitly shared with the destination of its IP traffic and implicitly shared with any party passively observing its network traffic. We begin by defining *first, support, and third parties* based on the owner of the IP address being contacted by an IoT device.

- *First party.* Manufacturer of the IoT device or a related company responsible for fulfilling the device functionality.
- *Support party.* Any company providing outsourced computing resources, such as CDN and cloud providers.
- *Third party.* Any party that is not a first or support party. This includes advertising and analytics companies.

In addition to the destination of IP traffic, we consider *network eavesdroppers* that can passively observe information exposed by IoT devices, such as the device's Internet service provider (ISP).

Privacy concerns. A "non-first party" is defined as any support party, or any third party. To understand whether information exposure has privacy implications, we consider:

- any personally identifiable information (PII) contained in network traffic and exposed to a non-first party;
- any recordings of users (audio/video/image) or user activity (motion sensors, television viewing habits) exposed to non-first parties, or exposed to a first party in a way that is neither disclosed nor expected by an average user;
- any collection of network traffic that allows a non-first party to observe devices in a home, when they are used, and how they are used (e.g., for profiling users).

2.2 Goals

In this section, we define our key research questions concerning different information exposures.

RQ1: What is the destination of network traffic?
Communication with *third parties* can be a privacy concern because such parties can track information about users, possibly with the intent to monetize the data (e.g., through ads). Moreover, *support parties* that serve multiple IoT devices (including those from different manufacturers) may gain detailed visibility into activities within a home. Finally, data traversing international boundaries may be subject to different privacy laws, including lawful intercept regulations.

RQ2: To what extent is the traffic encrypted?

Use of encryption can prevent exposure of sensitive information to eavesdroppers, while lack of encryption could expose the identity of a device, interactions with the device, and other sensitive information.

RQ3: What data is sent in plaintext?
When plaintext network traffic contains sensitive data (see §2.1), we consider it a privacy concern. Examples include personally identifiable information (name, location, e-mail address), devices identifiers, and credentials that can be used for unauthorized access to a device and/or its data.

RQ4: What content is sent using encryption?
While traditionally considered to provide confidentiality, encryption alone does not prevent the exposure of sensitive information. For example, sensitive data may be exposed to third party via encryption, or an eavesdropper may reliably infer the devices types and activities based on encrypted traffic patterns and plaintext protocol information (e.g., TCP/IP headers, TLS handshakes).

RQ5: Does a device expose information unexpectedly?
This research question is inspired by recent reports indicating that smart speakers surreptitiously and continuously recorded audio and streamed it their providers [9, 48]. More generally, we focus on sensitive information (e.g., user recordings) that the user expects to be exposed only via explicit interaction (e.g., keywords for smart speakers, doorbell ringing for video doorbell). This becomes a privacy concern when it is exposed to any party—even a first party—when a user did not intentionally trigger this exposure.

RQ6: Does the device's location (jurisdiction, location of network egress) impact information exposure?
IoT devices may be permitted to expose more or less information depending on regional regulations (e.g., GDPR in the EU). Differences in exposure for the same devices located in, or whose network traffic egresses out of, different jurisdictions may indicate adaptation to local laws.

2.3 Non-Goals

Unmodified devices. We use only unmodified devices in our experiments. Modifying devices or their firmware may reveal ground truth about information exposure, but doing so is not scalable to large numbers of devices (and not feasible with many of them).

No use of MITM. We do not man-in-the-middle (MITM) TLS connections to reveal the plaintext content of encrypted traffic. Our preliminary work on this topic showed that using MITM most of the time fails and, when successful, it did not reveal much additional useful information. Worse, MITM attempts often affected device functionality and behavior (i.e., devices would malfunction due to TLS connection rejection). Since MITM affected the validity of our results by changing device behavior, we opted not to do so for this study.

No companion apps traffic. We capture all network traffic in our labs, including those from companion apps used to interact with IoT devices. However, since we found little additional information exposure beyond what has been found using prior techniques [23, 35, 37], we focus only on the traffic generated by IoT devices.

Incompleteness. We cannot identify *all* information exposed, privacy-related or otherwise, from the IoT devices in our tests. Further, we cannot quantify the privacy risk for the information exposure we measure, because risk is subjective and we often lack ground truth. Rather, we focus on information exposed, potential privacy implications, and case studies of unexpected exposure of sensitive data.

3 DATA COLLECTION METHODOLOGY

This section covers our data collection methodology. We describe the devices under test, the labs in which we conduct experiments, and the experimental methods we use to capture information exposure from the devices. Collectively, we conduct 34,586 repeatable experiments on 81 devices in two labs (one in the US at Northeastern University's Mon(IoT)r Lab and one in the UK at Imperial College London) over one month, and include six months of data gathered from an ongoing IRB-approved study involving 36 participants.

3.1 IoT Devices

Our analysis covers 81 IoT devices with IP connectivity: 46 purchased from US stores (*US devices*) and deployed in our US testbed, 35 purchased from UK stores (*UK devices*) and deployed in our UK testbed. There are 26 *common devices* across the two labs (i.e., a device with a given model name is in both labs). The devices belong to the following categories (summarized in Table 1): *cameras* (security cameras and video doorbells), *smart hubs* (home automation devices which act as bridges for non-IP IoT devices, such as Zigbee, Z-wave, and Insteon devices), *home automation* (smart lights, outlets, and thermostats), *TVs* (actual TVs and TV dongles), *audio* (smart speakers with voice assistant), and *appliances* (fridges, cleaning appliances, cooking appliances, weather stations).

We selected devices based on the following factors: First, we picked devices that cover a broad range of categories. Second, for each category of device we searched popular retail websites (e.g., Amazon) for availability and selected based on price, popularity, and customer ratings, much like we expect an average consumer would do when buying an IoT device. Third, we biased our selection toward devices that are both available to the US and UK market to enable direct comparison of device behaviors in different jurisdictions. Fourth, we acquired several wireless devices (light bulbs, smart locks, light switches) that require smart hubs to use Internet so we can test those hubs.

3.2 Testbeds

The devices are deployed in testbeds in the US and UK, using identical data-collection and experiment infrastructure.

Network. Both testbeds include a server that provides network connectivity and data collection. The server in each testbed is configured identically, consisting of a Linux server running Ubuntu 18.06 with two wired network interfaces (one for the Internet connectivity with a public IP address, one for the IoT devices under test using a private network address space), and two Wi-Fi adapters bridged with the wired IoT network for the wireless IoT devices under test (one for 2.4GHz devices and one for 5GHz devices). Multiple wired IoT devices are connected to the wired IoT network using a network switch. IoT devices communicate with the public Internet via a standard NAT implemented at the server.

To test the impact of egress IP address on information exposure, we configured VPN tunnels that connect the US lab to the UK lab and vice versa. Network traffic traverses these tunnels only during experiments labeled *VPN*.

Category	Cameras	Smart Hubs	Home Automation	TV	Audio	Appliances
Devices N_{US}=46 N_{UK}=35 $N_{US \cap UK}$=26 $N_{US \cup UK}$=81	🇺🇸 Amazon Cloudcam 🇺🇸 Amcrest Cam 🇺🇸🇬🇧 Blink Cam 🇺🇸🇬🇧 Blink Hub 🇬🇧 Bosiwo Cam 🇬🇧 D-Link Cam 🇺🇸 Lefun Cam 🇺🇸 Luohe Cam 🇺🇸 Microseven Cam 🇺🇸🇬🇧 Ring Doorbell 🇺🇸🇬🇧 Wansview Cam 🇬🇧 WiMaker Spy Camera 🇬🇧 Xiaomi Cam 🇺🇸🇬🇧 Yi Cam 🇺🇸 ZModo Doorbell	🇺🇸🇬🇧 Insteon 🇺🇸🇬🇧 Lightify 🇺🇸🇬🇧 Philips Hue 🇺🇸🇬🇧 Sengled 🇺🇸🇬🇧 Smartthings 🇺🇸 Wink 2 🇺🇸🇬🇧 Xiaomi	🇺🇸 D-Link Mov Sensor 🇺🇸 Flux Bulb 🇬🇧 Honeywell T-stat 🇺🇸🇬🇧 Magichome Strip 🇺🇸🇬🇧 Nest T-stat 🇺🇸 Philips Bulb 🇺🇸🇬🇧 TP-Link Bulb 🇺🇸🇬🇧 TP-Link Plug 🇺🇸🇬🇧 WeMo Plug 🇺🇸 Xiaomi Strip	🇺🇸🇬🇧 Apple TV 🇺🇸🇬🇧 Fire TV 🇺🇸 LG TV 🇺🇸🇬🇧 Roku TV 🇺🇸🇬🇧 Samsung TV	🇬🇧 Allure with Alexa 🇺🇸🇬🇧 Echo Dot 🇺🇸🇬🇧 Echo Spot 🇺🇸🇬🇧 Echo Plus 🇺🇸🇬🇧 Google Home Mini 🇬🇧 Google Home 🇺🇸 Invoke with Cortana	🇺🇸🇬🇧 Anova Sousvide 🇺🇸 Behmor Brewer 🇺🇸 GE Microwave 🇬🇧 Netatmo Weather 🇺🇸 Samsung Dryer 🇺🇸 Samsung Fridge 🇺🇸 Samsung Washer 🇬🇧 Smarter Brewer 🇺🇸 Smarter iKettle 🇺🇸🇬🇧 Xiaomi Cleaner 🇺🇸 Xiaomi Rice Cooker
Purpose	Devices offering or supporting a camera that can be accessed remotely such as smart cameras and doorbells.	Devices designed to integrate non-Wi-Fi wireless devices into an IP network.	Wi-Fi sensors and actuators such as switches, bulbs, movement sensors.	Smart TVs and devices designed to connect to TVs via HDMI.	Smart speakers offering a voice assistant.	Home appliances that offer remote control.
Interaction Experiments	Move in front of camera, watch remotely, record video, take picture, ring (doorbells).	Turn on/off, change brightness/color, move in front of movement sensor.	Turn on/off, change brightness/color, move in front of movement sensor, change temperature (thermostats).	Browse menu, voice command, change volume.	Voice command, change volume.	Start, stop, change temperature, view inside (fridge), voice/volume (fridge).

Table 1: IoT devices under test. From top to bottom: IoT devices by category, their common purpose within the category, and the interaction experiments we performed (if available) on all the devices within the category. Flags indicate the presence of the device in the US, UK, or both testbeds.

Data collection. All the traffic traversing the testbed's network gateway server is automatically captured using tcpdump, using different files for each MAC address to separate traffic from different devices. We specify *labels* (stored in additional pcap files) to isolate the traffic produced during specific interactions with the device (e.g., "turn on the smart light"). Our testbed supports labels that are manually or automatically added, depending on whether the corresponding experiment is manual or automated.

Ancillary devices for interaction. Our experiments involve manual and automated interactions. For IoT devices that require a companion app, we use Nexus 5X smartphones running Android 6.0.1. Similarly, for IoT devices that require a voice assistant, we use the Echo Spot (powered by Alexa).

A key enabling factor for the large scale of our measurement experiments is automation. We automate IoT device interactions via their apps using the Monkey Application Exerciser included in Android Studio. For interactions via voice assistants (e.g., Alexa, Cortana), we need a way to produce speech using the same command words and same voice across both labs. For this, we use the cloud-based Google voice synthesizer to generate voice commands based on text specified in automated experiments, and play those commands over a loudspeaker located near a voice assistant device (e.g., Echo Spot).

3.3 Experiments

We conduct controlled, idle, and uncontrolled experiments to analyze the information exposed by devices under different conditions. The controlled (power, interaction) and the idle experiments were performed both in the US and the UK testbed during April 2019, with device firmware and companion apps updated to the latest version. All the controlled and idle experiments were repeated over the VPN tunnel, thus giving UK connectivity to the US-based IoT devices, and US connectivity to the UK IoT devices. User accounts for all UK and US devices were created in the same country as the lab in which they were deployed. In total we performed 34,586 controlled experiments (20,777 using the US testbed, and 13,809 using the UK testbed), plus 112 hours of idle experiments.

Power experiments. During our preliminary study we found that most IoT devices exchange a considerable amount of traffic when they are powered on. Thus, our *power* experiments consist of powering on the device (previously disconnected from the AC power) and collecting network traffic for two minutes without any interaction. We manually repeat these experiments at least three times for each device. In total we have performed 487 power experiments across both labs.

Interaction experiments. To understand the information exposed while interacting with a device, we conduct *interaction* experiments. These consist of actively interacting with IoT devices and then labeling the captured traffic with the interaction name. For each of these experiments we first wait for the device to be powered on for at least two minutes (to avoid including power experiments traffic). After two minutes, and right before the interaction starts, we begin capturing the traffic and continue to do so for the entire duration of the interaction (the exact amount of time depends on the device and the interaction method, i.e., the duration of a physical/app/voice interaction), plus at least additional 5-15 seconds after the interaction has completed.

The type of interactions we consider for these experiments are the following: (*i*) local action, which consists of physically interacting with the device, or using voice commands (without using a voice assistant from a separate device). (*ii*) LAN app action, by using a companion app on a phone connected to the *same network* as the IoT device, thus allowing direct communication between the phone and the IoT device; (*iii*) cloud app action, by using a companion app on a phone connected to a *different network* than the IoT device, to force the IoT device to use cloud infrastructure to communicate; (*iv*) voice command action, by using voice commands to trigger the

Echo Spot's Alexa voice assistant, which subsequently interacted with the device according to the voice command.

The majority of interaction experiments that involve the use of a companion app or a voice command have been automated and repeated at least 30 times, resulting in a total of 32,030 automated interaction experiments across both labs. Experiments that involve physical interactions or whose interactions cannot be automated safely or reliably (e.g., devices in the appliances category having a heating element) have been performed manually and repeated at least three times, resulting in a total of 2,069 manual interaction experiments across both labs.

Idle experiments. Beyond the initial powering on phase, one might expect that an IoT device would have minimal information exposure when it is not actively being used. To test whether this intuition holds, we conduct *idle* experiments that capture the traffic of an IoT device when it is not actively used and is located in an environment isolated from human interaction. Our idle periods cover an average of 8 hours per night for one week for each lab.

Uncontrolled experiments (US only). These experiments consist of capturing all the (unlabeled) traffic generated by the IoT devices in the US testbed during an IRB-approved user study, where 36 user study participants are allowed to use the IoT devices for their intended purpose in a studio apartment setting. These participants are 5 faculty/postdoc and 31 students at Northeastern University's Cybersecurity and Privacy Institute in Boston. Participants may use the lab at any time (except during the period where we conducted idle experiments), and they may use any device in the room as they see fit. Commonly used devices include the refrigerator, laundry machines, and microwave, while the Alexa devices and motion-triggered devices are also frequently active. Collectively, we typically see about 20-30 lab accesses per day, with at least one active device interaction per access. A common interaction pattern is a person that enters the lab to put their food in the smart fridge (or to put their laundry in the washer), then they come again later to reheat it in the smart microwave (or to move their laundry to the dryer). These common interaction patterns do not trigger just the devices that the participants are actively using, but also smart cameras, smart doorbells, smart motion/contact sensor, and smart lights, which are always active and passively triggered by the simple presence of the participant.

We use a subset of the user study data collected between September 2018 and February 2019 (inclusive). We filter out any power and interaction experiments conducted during this period. The uncontrolled experiments were conducted on the US testbed only.

To measure regional differences in IoT traffic, we use two approaches. First, we compare information exposed by the common devices across the UK and US labs. Any differences in information exposure could be explained by factors such as differences in hardware/firmware sold in different markets, egress IP address, server selection based on IP address or location, and data protection regulations in each jurisdiction. Second, we use the VPN connection between labs to compare US devices egressing into the public Internet via a US IP address, and the same devices egressing via a UK IP (and vice versa for the UK lab devices). In this case, the hardware/firmware and jurisdiction are identical in each pair of scenarios, but the egress IP address (and server selection based on the IP) vary. We combine observations across these scenarios to help identify likely root causes for observed differences in information exposure.

4 DESTINATION ANALYSIS

In this section, we characterize IoT devices in terms of the destinations of their network traffic in order to answer the RQ1. Our focus is on which parties are contacted (as defined in §2.1), their geolocation, and what are the most common non-first-parties contacted.

4.1 Measuring Destinations

In this section we analyze the destination IP address of IoT device traffic according to whether the destination is a first party, third party, or support party.[1] In addition, we consider the geolocation of those destinations (in terms of inferred country), because traffic traversing international boundaries may be subject to different domestic surveillance regulations, and because different countries may impose different restrictions on content providers. We use the following approach to label the party and geolocation for a destination IP.

Second level domain name (SLD). For each flow from a device, we determine the SLD by first identifying whether the destination IP address corresponds to a DNS response for a request issued by the device. If so, we use the SLD for the corresponding DNS lookup; otherwise, we search HTTP headers (Host field) and/or TLS handshakes (Server Name Indication field) for the domain. If none of the above approaches yields a domain, we leave the IP's SLD unlabeled.

Identifying organization name. We identify the organization name for an SLD using either WHOIS data or common-sense matching rules (e.g., "Google" is the organization for google.com). If we could not identify an SLD for a domain, we set the organization to the owner of the IP address as reported by the corresponding regional registry (e.g., RIPE for European IPs).

Determining party type. If the IP's organization identified in the previous step matches the name, manufacturer, or related company to the IoT device, we classify it as a first party. If not, we manually search for public information about the party. If the company states on its website that it is specialized in providing connectivity (CDN) or cloud services (e.g., Amazon AWS), we consider the party a support party. In any other case we consider the party a third party.

Determining party country. We use the Passport [36] tool, which is able to infer the country containing a destination IP address by combining traceroute data with other IP geolocation sources. We do not use public geolocation databases alone, which we found to be highly inaccurate when manually validating results.

4.2 Destination Characterization

Table 2 shows the number of unique destinations for each type of experiment contacted by the US and UK devices. The first column of the table shows the type of experiment, the second column the type of party (support vs third), the remaining columns indicate values referring to a group of devices using the following notation that

[1] We ignore network traffic sent inside the LAN, as this is outside the set of privacy concerns we defined in the previous section.

Experiment	Party	US	UK	US∩	UK∩	VPN			
						US→UK	UK→US	US∩	UK∩
Idle	Support	40	38	28	23	44	38	31	20
	Third	4	3	4	3	3	2	3	2
Control	Support	85	82	71	66	76	71	66	57
	Third	7	4	5	4	7	3	5	3
Power	Support	39	39	29	29	33	35	21	24
	Third	4	3	3	3	4	3	4	3
Voice	Support	8	10	5	5	7	5	2	3
	Third	0	0	0	0	0	0	0	0
Video	Support	34	29	25	21	31	26	27	19
	Third	0	0	0	0	0	0	0	0
Total	**Support**	**98**	**87**	**74**	**70**	**93**	**84**	**72**	**69**
	Third	**7**	**5**	**5**	**5**	**7**	**5**	**5**	**5**

Table 2: Number of non-first parties contacted by devices, grouped by *experiment type* (first column), party type (second column) across lab location and network connectivity (right columns).

we will use for the remainder of the paper: "US" and "UK" headers represent all US and all UK devices, US∩ and UK∩ headers represent only the US and UK devices that are in common between the two labs, headers prefixed with "VPN" indicate that those devices are connected to the Internet using the VPN link (meaning that US devices reach the Internet via the public IP address of the UK testbed, and vice versa).

The table shows that control experiments lead to more communication to support and third parties when compared to other types of experiments. Among them, the power experiments represent the majority of the communications with third parties, likely due to devices establishing initial connections with destination parties. During the power experiments devices in the US contacted *netflix.com*, *tuyaus.com* (a Chinese IoT provider), *nuri.net* (a Korean ISP), and *facebook.com*. UK devices communicate with *nuri.net*, *netflix.com*, and *doubleclick.net*. During the control experiments there are two third parties contacted only by US-based devices: *omtrdc.net* (a tracking service owned by Adobe) and *mixpanel.com* (another tracking service), and one contacted only by the UK-based devices: *wowinc.com* (a US ISP). The latter is a Wansview camera, which we observed to contact IPs in many residential networks.

The total number of contacted parties (bottom rows) in the US is greater than the same for the UK. The Wansview camera contacts the largest number of destinations (52 unique destinations), followed by Samsung TV (30), Roku TV (15), and TP-Link plug (13).

Interestingly, when US devices are connected via the UK network (using VPN), they contact a lower number of parties. Among them: *branch.io* (tracking service), *fastly.net* (CDN), *edgecastcdn.net* (CDN owned by Verizon), and *hvvc.us* (cloud hosting). *Branch.io* is contacted by Fire TV, TP-Link plug and TP-Link bulb during the power experiment.

Table 3 shows the number of unique destinations per category in the two datasets. TVs (i.e., Samsung TV, LG TV, Roku, Fire TV) contact the largest number of third parties among all device categories.

Figure 2 shows the flow of traffic from the devices common to the US (left) and UK (right) labs, with the height of each band corresponding to the number of bytes transferred. Results are grouped by the device category (left middle and right middle) and terminate at destination countries/regions (center). A majority of device traffic terminates in the US for both the US and UK labs, likely due to reliance on infrastructure with limited geodiversity.

Category	Party	US	UK	US∩	UK∩	VPN			
						US→UK	UK→US	US∩	UK∩
Appliances	Support	5	3	-	-	8	5	-	-
	Third	1	0	-	-	1	0	-	-
Audio	Support	8	8	3	5	11	7	5	4
	Third	0	0	0	0	0	0	0	0
Smart Hubs	Support	9	9	5	4	9	7	5	3
	Third	0	0	0	0	0	0	0	0
Home Automation	Support	10	14	10	11	6	12	6	8
	Third	1	1	1	1	1	1	1	1
Cameras	Support	49	50	39	37	44	46	38	34
	Third	1	2	1	2	2	2	2	2
TV	Support	18	18	16	18	17	16	16	16
	Third	4	2	3	2	3	2	3	2

Table 3: Number of non-first parties contacted by devices, grouped by *device category* (first column), party type (second column). Right columns specify lab location and network connectivity.

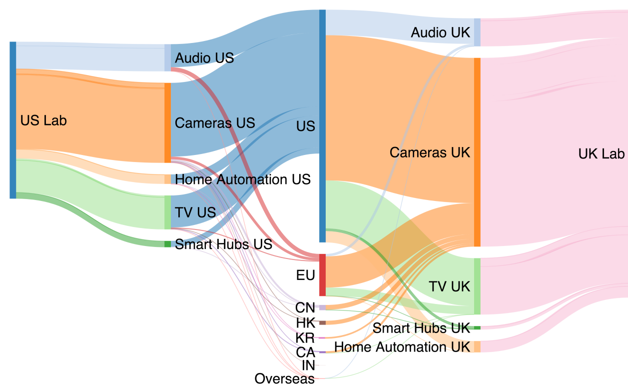

Figure 2: Volume of network traffic between the US (left) and UK (right) labs to the top 7 destination regions (center), grouped by category (middle left and right). Most traffic terminates the US, even for the UK lab; many devices send traffic to countries outside of their testbed's privacy jurisdiction.

Domain	US	UK	US∩	UK∩	VPN			
					US→UK	UK→US	US∩	UK∩
Amazon	31	24	16	17	31	24	16	17
Google	14	9	10	8	16	10	11	8
Akamai	10	6	6	5	10	6	6	5
Microsoft	6	4	1	1	6	4	1	1
Netflix	4	2	3	2	4	2	3	2
Kingsoft	3	3	1	1	4	3	1	1
21Vianet	3	3	1	1	4	3	1	1
Alibaba	3	4	2	2	4	4	2	2
Beijing Huaxiay	3	3	1	1	3	3	1	1
AT&T	2	0	1	1	0	0	0	1

Table 4: Organizations contacted by multiple devices.

US devices contact 13 overseas countries and several countries in the EU, while devices from the UK contact only 7 overseas countries including the US. From the oversea countries, most of the traffic is sent to China as many Chinese devices rely on the services provided by the *Alibaba Cloud*.

4.3 Commonly Contacted Destinations

Next, we analyze which non-first-party domains and companies are contacted by the largest number of devices. These parties can learn a great deal about the devices in a home and how they are used.

Even the knowledge of a significant subset of devices in a given home could be used to profile users for data-driven algorithms such as targeted advertising.

For this analysis, we focus on the destination *organization*, e.g., nest.com and google.com belong to the Google organization that owns both properties. A single device may contact multiple organizations, e.g., the Samsung TV contacts two non-first-party organizations: Netflix (a third party), and Amazon AWS (a support party), which hosts Netflix.

Table 4 lists the ten companies contacted by the largest number of devices. We find that 31 devices in the US and 24 in the UK contact at least one server that belongs to Amazon, largely due to their reliance on AWS for hosting servers. The next-most commonly contacted organizations are also hosting providers: Google, Akamai, and Microsoft. Thus, a substantial contributing factor to the amount of information exposed to non-first-parties is the use of supporting parties that provide computing resources.

Nearly all TV devices in our testbeds contacts Netflix *even though we never configured any TV with a Netflix account*. This, at the very least, exposes information to Netflix about the model of TV at a given location. The bottom half of the table (apart from AT&T) consists of Chinese counterparts to Amazon and Google. These servers are predominantly contacted by devices designed by Chinese companies.

Generally, connecting a device to the Internet via a VPN did not lead to significant differences. The only unexpected case was the US based Xiaomi Rice Cooker contacted Kingsoft only when connected via VPN, normally it contacts Alibaba cloud service.

4.4 Takeaways

When focusing simply on the destinations contacted by IoT devices, we identified a number of privacy concerns. Several non-first-party destinations (in particular Amazon, Google, and Akamai) receive information from many of our IoT devices, thus allowing them to potentially profile consumers. For example, these companies not only can learn the types of devices in a household, but also how/when they are used, simply by analyzing the network traffic from IoT devices to their cloud services (see §6.3). We observe substantial outsourcing of computing resources to cloud providers, particularly for camera devices; further, we found TVs comprise the largest fraction of third party communication (likely to customize content for users). Regarding regional differences, we note that US devices tend to contact more non-first parties, possibly due to more relaxed privacy regulations with respect to the EU. VPNs had minimal impact on the type of party contacted, with most differences likely being due to serving content using replicas closer to the VPN egress.

5 ENCRYPTION ANALYSIS

In this section, we analyze whether devices send data securely by studying the use of encryption by IoT devices in our testbed. Adoption of encryption would be easy to measure if most network traffic in the IoT environment used easily-identifiable protocols such as HTTP, HTTPS, and QUIC. However, standard protocol analysis tools (e.g., Wireshark's protocol analyzer) fail to classify nearly half (46%) of the network traffic originating from our testbeds. We address this problem using entropy analysis, then use it to characterize the volume of traffic sent securely by each device.

5.1 Identifying Encrypted Traffic

We start by using Wireshark's protocol analyzer to identify TLS (excluding handshakes) and QUIC traffic as encrypted. Wireshark does not identify other encrypted protocols. Certain unclassified network traffic contains encoded or compressed content (e.g., video, audio, gzip compression). We search for encoding-specific bytes in headers of such flows, and mark any traffic that contains them as unencrypted.

For the remaining flows, we do not have ground truth as to whether the traffic is encrypted. Therefore, we infer the use of encryption on those flows by measuring their byte entropy H, whose value is between 0 and 1 with higher value meaning byte sequences that are more similar to random.

By analyzing the entropy of some randomly sampled traffic (e.g., encrypted traffic consisting of HTTPS flows and HTTP flows containing encrypted payload, and unencrypted traffic consisting of HTTP flows containing textual payload), we found that the average entropy H for the payload of encrypted traffic is $H_{enc} = 0.85$ ($\sigma = 0.009, min = 0.80, max = 0.86$), while the entropy of unencrypted textual traffic is substantially lower but has higher variance, with average of $H_{unenc} = 0.25$ ($\sigma = 0.09, min = 0.12, max = 0.39$).

We conducted additional tests of encrypted and unencrypted content (namely, IMC 2019 web pages) to understand how entropy varies across encryption algorithms and cipher suites. The entropy of the unencrypted content is $H_{unenc} = 0.55$ ($\sigma = 0.07, min = 0.35, max = 0.62$). First, we used 14 cipher suites [2] available using python's TLS implementation, which resulted in $H_{enc} = 0.85$ ($\sigma = 0.02, min = 0.80, max = 0.87$). While HTTPS encryption is popular, other encryption schemes might yield substantially different entropy ranges. To investigate this, we used the python symmetric encryption library *cryptography/fernet* to encrypt the same set of IMC 2019 website content. We found the entropy of such encrypted content led to an average $H_{enc} = 0.73$ ($\sigma = 0.025, min = 0.67, max = 0.75$). Importantly, the differences in entropy values between TLS and fernet encryption are relatively small, and in both cases the gap between encrypted and plaintext entropy is large.

Based on these observations, we cannot identify a single threshold that will always classify encrypted and unencrypted payloads correctly. In light of this, we chose conservative thresholds for whether a connection is encrypted or not, with the goal of reducing false positives/negatives while relegating remaining cases to an "undetermined" class. Specifically, we classify traffic having entropy H>0.8 as *likely encrypted*, H<0.4 as *likely unencrypted*, and $0.4 \leq H \leq 0.8$ as *unknown*, which corresponds to undetermined encryption status.

We acknowledge that there are likely better alternatives for selecting thresholds. To facilitate future experimentation on this topic, we have made our datasets and code public so that others may investigate the suitability of different thresholds, including our current selections.

Note that the above analysis assumes substantial differences in entropy between unencrypted and encrypted content. We found

[2] This comprises all ciphers that successfully transmitted data; the full list is available in our public data.

Enc	Range	US	UK	US∩	UK∩	VPN US→UK	VPN UK→US	VPN US∩	VPN UK∩
✗	>75	0	0	0	0	0	0	0	0
	50-75	1	1	0	0	2	0	1	0
	25-50	4	1	1	1	3	2	0	1
	<25	41	31	24	24	41	31	24	24
✓	>75	7	7	5	5	4	5	3	3
	50-75	5	7	4	6	7	8	5	7
	25-50	10	5	5	4	12	5	7	5
	<25	24	14	11	10	23	15	10	10
?	>75	16	10	8	7	17	11	8	7
	50-75	11	6	5	5	11	5	5	4
	25-50	11	7	6	5	13	10	8	9
	<25	8	10	6	8	5	7	4	5

Table 5: Number of devices by encryption percentage in quartile groups across lab and network.

that this assumption does not hold for media content, e.g., in our experiments with recording videos from a mobile device we found that $H_{unenc} = 0.873$ ($\sigma = 0.0005, min = 0.871, max = 0.874$). Thus, for cases of media (audio/video) content that are not filtered out using recognized encodings, we use their network traffic patterns to identify and exclude them from this analysis using the technique described in Section 6.

5.2 Encryption Adoption

In this section we analyze the adoption of encryption by our US and UK devices. First, we show the fraction of the traffic that we recognize as unencrypted, encrypted, or unknown. Then, for unencrypted traffic we identify patterns among device categories and activities.

Overall adoption. Table 5 shows the number of IoT devices that have a particular fraction of their data unencrypted (first row ✗), encrypted (second row ✓), and unknown (third row "?"). Each subrow represents the fraction of the traffic we consider for the classification, using quartiles. The first two columns consider the set of all US and UK devices, the third and fourth columns consider the set of all US and UK devices in common in both testbeds. The remaining columns show the same data of the first four, but using VPN egress.

The table shows some positive trends: no devices have more than 75% unencrypted traffic, and just one of them (in each testbed) have more than 50%, while 7 devices (in each testbed) have more than 75% encrypted traffic. We also observe some negative trends that reveal possible information exposure: 5 devices in the US and 2 in the UK send more than 25% unencrypted traffic, while all but 8 devices US and all but 10 devices in the UK send more than 25% unknown traffic. This last point motivates the need for better protocol analyzers to understand just how much of this unknown traffic is encrypted. If we consider the devices in common between the US and UK, and the VPN experiments we see similar trends, meaning that except for some isolated cases we will discuss in the next paragraph, there are no significant regional differences in the distribution of encrypted, unencrypted, and unknown traffic for the same devices.

Adoption by category. We now analyze information exposure in terms of unencrypted data exposed by devices according to their category. Table 6 shows the percentage of data exposed in aggregate by all devices in each category (rows), partitioned into different

Enc	Type	US	UK	US∩	UK∩	VPN US→UK	VPN UK→US	VPN US∩	VPN UK∩
✗	Appl-	7.1	0.3	0	0	7.6	0.3	0	0
	Audio	1.4	1.7	1.5	2.1	1.7	1.6	1.8	2.0
	Hubs	2.7	4.5	2.3	4.5	3.1	4.3	2.9	4.3
	Auto	7.1	6.0	9.8	7.9	9.3	4.5	11.4	6.1
	Cam	11.1	10.3	0.7	0.2	10.8	3.9	0.6	0.3
	TV	8.0	12.2	9.3	12.2	16.0	12.6	18.1	12.6
✓	Appl-	26.9	11.4	0	0	26.5	2.9	0	0
	Audio	61.2	61.8	57.6	54.0	54.1	54.7	49.6	43.6
	Hubs	24.9	18.2	19.3	18.2	28.9	21.4	24.8	21.4
	Auto	29.1	51.0	42.7	55.0	28.4	49.6	42.2	53.6
	Cam	9.6	14.6	22.6	25.4	8.9	13.3	21.4	23.1
	TV	61.2	73.8	64.4	73.8	40.7	59.7	41.3	59.7
?	Appl-	63.3	55.0	88.1	50.1	63.2	63.5	87.8	50.0
	Audio	36.0	36.5	39.2	43.8	41.1	43.8	44.8	54.4
	Hubs	71.9	77.2	77.8	77.2	67.5	74.2	71.6	74.2
	Auto	57.3	37.9	36.0	30.1	55.9	41.2	34.8	33.9
	Cam	76.8	69.4	70.8	64.4	77.5	76.5	71.1	65.7
	TV	30.7	13.9	26.3	13.9	43.3	27.7	40.6	27.7

Table 6: For each device category, the average percent of bytes sent unencrypted by corresponding devices.

experiment locations and egress configurations (columns). A key observation is that cameras collectively expose the largest fraction of unencrypted bytes (up to 11% in the US testbed), most of which come from the Microseven camera and Zmodo doorbell in the US, and the Spy camera in the UK, which are not common to both testbeds. The next largest fraction of unencrypted bytes (7.1% in the US and 6.0% in the UK) come from home automation devices (e.g., D-Link movement sensor) and appliances (mostly Samsung washer and dryer). On the other end, audio devices tend to use the most encryption (more than 60% on both testbeds), likely because they are built and designed by major corporations (e.g., Amazon and Google) known to have high security standards. Another important observation is that appliances, home automation devices, and smart hubs have the largest portion of unknown traffic. After a manual investigation we found that such devices have a larger fraction of proprietary protocols not known to Wireshark, which are often partly encrypted, making the entropy analysis inconclusive and motivating future investigation.

The Samsung TV and FireTV are the isolated cases showing a significant difference in encrypted traffic depending on whether they are connected to the Internet directly or via VPN (Table 7). We suspect this occurs because the TVs detect the device geolocation based on egress IP, and customize content displayed to the user (e.g., available streaming services and content) based on the inferred country. We have validated this hypothesis by manually observing what these TVs show when they are powered on and used: the content displayed clearly reflects the region corresponding the public IP address of the device and, in the case of VPN experiments, they advertise services and interactive content that is normally not available without using the VPN. Turning off the VPN and restarting the device is able to restore the previous content and behavior.

Encryption analysis by experiment type. We now investigate whether the experiment type has any impact on the fraction of (un)encrypted bytes sent. Our results, reported in Table 8, show video having the lowest fraction of encrypted bytes and voice interactions having the highest. This is similar to the category analysis

Device	US	UK	VPN	
			US→UK	UK→US
TP-Link Smartplug	***18.6***	*8.7*	***23.4***	*12.0*
TP-Link Bulb	**13.1**	**12.8**	***17.2***	***5.9***
Nest Thermostat	*11.6*	***15.8***	11.3	**10.5**
Smartthings Hub	*6.7*	*16.6*	*5.2*	*14.6*
Samsung TV	*7.1*	*4.5*	10.1	1.4
Echo Spot	*2.3*	*1.9*	2.6	2.1
Echo Plus	*1.8*	*2.9*	*1.8*	*2.4*
Fire TV	**0.8**	0.6	**5.2**	1.0
Echo Dot	0.7	2.6	*1.2*	*2.7*
Yi Camera	0.5	0.5	*0.9*	*0.5*
Samsung Dryer	**28.1**	-	**29.3**	-
Samsung Washer	**27.3**	-	**28.6**	-
D-Link Movement Sensor	**14.9**	-	**24.6**	-

Table 7: Average percentage of unencrypted bytes per device. The topmost devices are common devices to both testbeds, while the last three exist only in the US lab. Text decoration indicates statistically significant differences across the following categories: Bold indicates differences when using VPN; *Italic* means differences between US and UK.

Enc	Exp (#D) /%	US	UK	US∩	UK∩	VPN US→UK	VPN UK→US	VPN US∩	VPN UK∩
	Control (79)	5.4	4.8	4.2	5.0	6.6	4.3	5.4	4.5
	Power (78)	8.2	10.2	8.3	8.7	11.2	11.5	11.3	10.2
	Voice (18)	2.8	1.4	2.8	1.8	4.0	2.3	4.0	2.9
✗	Video (19)	6.9	2.7	0.8	0.4	7.6	1.9	0.7	0.3
	Others (78)	4.8	4.4	4.2	5.3	5.8	3.5	5.3	4.3
	Idle (81)	5.4	5.0	4.5	4.2	5.9	5.7	5.2	4.1
	Uncontrol (46)	3.9	-	-	-	-	-	-	-
	Control (79)	27.6	33.2	32.8	34.4	27.5	31.7	33.3	32.7
	Power (78)	33.0	42.6	38.2	43.5	27.1	35.2	32.5	37.2
	Voice (18)	58.7	67.4	53.2	62.9	48.0	60.8	44.8	55.2
✓	Video (19)	9.2	15.1	18.6	22.8	8.5	13.3	16.3	20.0
	Others (78)	29.0	34.0	33.2	33.9	30.0	33.7	35.2	33.1
	Idle (81)	26.3	21.8	32.4	25.6	18.7	17.9	22.2	19.7
	Uncontrol (46)	29.1	-	-	-	-	-	-	-
	Control (79)	66.8	60.4	62.6	59.0	65.7	63.6	60.9	62.3
	Power (78)	58.8	47.2	53.5	47.8	61.7	53.3	56.2	52.6
	Voice (18)	38.5	31.2	44.0	35.3	48.0	36.9	51.2	41.9
?	Video (19)	83.8	82.2	80.6	76.8	83.9	84.8	83.0	79.7
	Others (78)	65.9	59.3	62.1	58.7	63.9	62.2	59.0	61.9
	Idle (81)	68.2	73.1	63.1	70.2	75.4	76.3	72.6	76.2
	Uncontrol (46)	67.0	-	-	-	-	-	-	-

Table 8: Percent of bytes sent unencrypted, grouped by experiment type. Number of devices considered in each row is in parentheses in the first column.

for TVs and audio devices (i.e., video interactions and the camera category are analogs, as are voice interactions and audio devices). The other experiments, which do not have interactions directly mapped to specific categories, do not show a clear trend, meaning that the differences in encryption are mostly due to the device itself, and not to the type of experiment.

This observation also holds for common devices in different regions. However, we notice less significant trends that may require further investigation to understand the underlying reasons: power experiments are the easiest to classify using entropy analysis and show the highest percentages of both unencrypted (>8.2%) and unencrypted (>33.0%) traffic. We also observe differences when devices are connected via VPN (e.g., for encrypted and unencrypted traffic of video devices). These differences show no clear patterns and are likely due to changes in device behavior (e.g., similar to different content/functionality as seen with TVs) when the device detects a different region.

5.3 Takeaways

While unencrypted traffic is a minority of all traffic, we identified substantial information exposure via plaintext traffic for all devices, categories, interactions, and regions. Most differences in unencrypted traffic across device category and device interactions are due to specific devices, rather than being endemic of an entire category of device. We observed regional differences in the use of encryption, especially in devices in the TVs category, since they interact with different content providers depending on their detected region.

6 CONTENT ANALYSIS

In this section, we analyze the information exposed to other parties by IoT devices as defined in RQ3 (inferring unencrypted content) and RQ4 (inferring encrypted content). Specifically, we focus on the network traffic for two content types: textual PII contained in unencrypted network traffic (e.g., names, e-mail addresses, locations as identified), and device activity that can be inferred from encrypted network traffic (e.g., video is streaming from a video doorbell). For the latter, we build, evaluate and use a machine learning classifier that leverages network traffic statistics as features, and our ground-truth experiment labels for detection. We then analyze the potential privacy implications of such information exposure.

6.1 Identifying PII and Device Activity

We use the following techniques to identify PII and device activity in network traffic.

Textual PII in unencrypted traffic. To identify PII exposed in plaintext, we simply search for any PII known (in various encodings) in each device's network traffic. For the purpose of this analysis PII includes device identifiers (e.g., MAC address, UUID, etc.), and any personal information given at registration time (e.g., names, email address, home address, phone number, username, password, etc.).

Device activity inference (encrypted or unencrypted). To infer the device activity based on network traffic (regardless of whether it is encrypted), we train a random forest machine learning classifier using experiment labels and network traffic for each device interaction. Examples of device activity that is contained in our labels include power on a device, issue voice command, view video stream. The set of features we use to train our classifier are *timing statistics of the traffic* with respect to packet sizes and inter-arrival times. The statistical proprieties we consider as features are the following: *min*, *max*, *mean*, *deciles of the distribution*, *skewness*, and *kurtosis*. We focused on features that avoid dependencies on text- or size-based features that can easily vary across deployment location (e.g., due to different hostnames selected as part of location-based server redirection), while still yielding high accuracy under cross-validation.

We observed that experiments from certain devices contain network traffic resulting directly from the experiment interaction, as well as network traffic that is unrelated to the experiment (e.g., time synchronization via NTP). Thus, we leverage multiple interactions (30 automated tests, 3 manual tests) for each interaction type, to mitigate the effects of such traffic noise. The large numbers of tests in the automated cases also provide enough samples to apply cross-validation, thus allowing us to evaluate the accuracy of our method.

6.2 Textual Unencrypted Content

We found limited identifiable content, and even less PII, in unencrypted traffic. This is good news, particularly compared with prior work that identified substantial amounts of PII exposed in plaintext in other contexts (e.g., mobile apps and web sites [25, 37]).

Nonetheless, we found notable cases of PII exposure. This included various forms of unique identifiers (MAC address[3], UUID, device ID), geolocation at the state/city level, and user specified/related device name (e.g., John Doe's Roku TV). A notable case that we found in our US lab is the Samsung Fridge sending MAC addresses unencrypted to an EC2 domain, which is a support party in the best case. The implication is that it is now possible for an ISP to track this device.

In both our labs we found that Magichome Strip is sending its MAC address in plaintext to a domain hosted on Alibaba. Interestingly, the Insteon hub was sending its MAC address in plaintext to an EC2 domain, but only from the UK lab. We did not find similar behavior in the US lab. Interestingly, each time the Xiaomi camera detected a motion, its MAC address, the hour and the date of the motion (in plaintext) was sent to an EC2 domain. We also noted that a video was included on the payload.

We investigated the content of plaintext communication that did not contain PII, and found that common cases included queries related to device actions (e.g., turn on/off the device or issue a command to the device). We also identified large unencrypted file transmissions that contained firmware updates and/or metadata pertaining to initial device set up.

6.3 Device Activity Inference

In this section, we describe how we trained machine learning classifiers to estimate how much device activity can be inferred based on network traffic. Note that we do not claim (or attempt) to produce the most performant classifiers according to metrics such as accuracy or F1 score. Rather, we use such metrics to understand whether device activities are *inferrable* as explained below.

To reliably infer device activity, we first validate our machine learning classifiers using 7/3 split cross validation (i.e., train on randomly selected 70% of the data and test on the 30% remaining data, and we repeat the process for 10 times to get the average metrics). Then we use the F1 score, defined as the harmonic mean between precision and recall, as a quality metric to evaluate the effect of false positives and false negatives for the detection of the activities of a device, where $F1 = 0$ is the worst score and $F1 = 1$ is the best score. We calculate the F1 score for the prediction of each activity of the device (defined as the F1 score for the activity), and the F1 score across all activities for each device (defined as the F1 score for the device). We consider an activity or device as *inferrable*, when its F1 score is greater than 0.75.

Predictable devices per category. Table 9 shows the number of devices whose actions are mostly all inferrable using our classifier.

[3]Note that MAC addresses can be used to identify a device vendor, which in some cases may uniquely identify the device.

Category (#D)	US	UK	US∩	UK∩	VPN US→UK	VPN UK→US	VPN US∩	VPN UK∩
Appliances (10)	2	0	0	0	1	0	0	0
Audio (11)	3	1	2	0	3	3	2	3
Cameras (17)	8	6	3	3	10	6	4	3
Home Auto (16)	0	1	0	0	1	1	1	0
Smart Hubs (14)	1	0	1	0	1	0	1	0
TV (8)	5	3	3	3	5	4	3	3

Table 9: Number of inferrable devices (F1 score > 0.75), grouped by category. Total number of devices per category in parenthesis in the first column.

Exp (#D)	US	UK	US∩	UK∩	VPN US→UK	VPN UK→US	VPN US∩	VPN UK∩
Power (75)	41	30	20	22	42	32	21	22
Voice (17)	10	6	6	4	9	6	5	5
Video (19)	11	7	5	4	11	7	5	4
On/Off (45)	9	5	5	3	9	7	6	5
Movement (19)	9	5	4	4	8	5	3	4
Others (52)	19	14	13	10	20	15	12	10

Table 10: Number of inferrable activities (number of devices with such an activity in parentheses), aggregated by activity group. We consider an activity for a device is inferrable when its F1 score is >0.75.

Cameras have the largest fraction of inferrable devices, followed by television devices, and audio devices. We believe this is due to the fact that cameras, TVs, and audio devices produce the most traffic during interactions, and this provides more samples to better train our classifier. If we look at the regional comparison between common devices in Table 9 we can see that in most cases we have some differences in the number of inferrable devices across regions (for example, 2 audio devices are inferred in the US and 0 in the UK). We observe this pattern of slight differences also if we compare each lab with and without VPN connectivity (for example, 2 appliances can be inferred in the US lab without VPN, and 1 with the VPN).

Devices with reliably inferrable activities. Table 10 shows the number of devices whose activities we can reliably infer. We find that the "Power" activity is the most inferrable, due to the unique traffic patterns that characterize it, followed by video and movement activities (due to the large amount of data that interacting with cameras produce). Each of these types of activities can be considered sensitive, as they indicate presence and activity in a home or other deployment space—information that can be readily inferred by a network eavesdropper. Regarding the regional comparison, Table 10 also shows that there are differences in the number of inferrable devices across regions (for example, 41 power experiments are inferrable in the US and 30 in the UK). Similarly to the previous case, we also observe differences in inferrable devices with and without VPN (for example, we can infer 9 devices offering movement activities without VPN, and 8 with the VPN).

6.4 Takeaways

We analyzed both unencrypted and encrypted content in this section. First, we found very limited sensitive or personal information exposed in plaintext—a welcome observation given the sensitivity of data potentially exposed by such devices. Second, we found that even when devices use encryption, the timing patterns of their network traffic permits reliable identification of the interactions that caused the network traffic. Put another way, an eavesdropper can reliably learn a user's interactions with a device across a wide range of categories, opening the potential for profiling and other privacy-invasive techniques. There are differences in inferrability across regions, a topic that warrants further investigation as part of future work.

7 UNEXPECTED BEHAVIOR

In this section we use the activity prediction methodology developed in §6 to detect unexpected behavior in idle and uncontrolled experiments.

7.1 Measuring Unexpected Behavior

We define unexpected behavior as cases when a device generates network traffic corresponding to an interaction that either did not occur, or was not intended by the user. To identify which network traffic corresponds to an interaction with an IoT device, we use our device activity inference approach on the traffic generated by *idle experiments* and *uncontrolled experiments*. In idle experiments there is no interaction with devices, so any inferred activity reveals possible privacy concerns due to any user monitoring activity. For our IRB-approved user study, we have some ground truth about user/device interactions (e.g., by asking users, watching users recordings, inspecting logs, and notifications produced by the device). We can compare such ground truth to what was inferred, thus allowing us to determine whether the detected activity is expected or not.

To identify activities from unlabeled network traffic, we must divide network traffic from a device into units amenable to classification. Choosing a value that is too small provides too little data for classification; a value that is too large may merge traffic together from multiple activities, and thus inhibit correct classification. For this study we use an empirically derived *traffic unit* of classification: a sequence of packets containing inter-packet interval greater than 2 seconds. To focus our analysis only on the most significant predictions, we use only the most accurate models based on cross-validation in §6; namely, only those with an F1 score > 0.9. During idle experiments, our model identified activities for 21% and 69% *traffic units*, depending on the location of the device and its network egress.

7.2 Idle Experiments

Table 11 shows the number of reliably predictable activities we have predicted during around 30 hours of idle experiments (details in the first row of the table). The most commonly detected activities are: "power" activities across all devices, "menu" activities from TVs (i.e., the act of navigating the home menu screen of the TV), and "move" activities for some cameras (i.e., the act of moving in front of a camera). We can also notice several less common and/or specialized activities (such as "view inside the fridge").

The large number of "power" activities is due to devices that frequently disconnect and reconnect to the Wi-Fi network (which we verified using DHCP server logs). When a device reconnects, it performs a new handshake with its cloud services similarly to when it is powered on. Thus, we do not consider power activities as unexpected or suspicious.

We believe that "menu" activities are explained by TVs that occasionally refresh the content of their menu page (e.g., showing new on-demand content available for viewing), much the same actions taken when the user navigates to the "menu". We do not consider this unexpected or suspicious.

The "move" activities are sometimes triggered when there is no movement in front of a camera and/or movement sensor. In the case of a camera, the device records a video when it is not triggered by movement, which is both unexpected and potentially suspicious. We do not have a root cause for this behavior. A possibility is that despite our best efforts to isolate each camera from external stimuli, there is nonetheless something in the lab that is triggering them (e.g., background noise from other IoT devices). Another possibility is that the devices capture such video intentionally.

We could not identify root causes for the remaining, rarer, cases of detected activity. As with any inference technique, ours is potentially subject to false positives and false negatives. Combined with our lack of ground truth, we cannot say for certain whether these cases are unexpected or suspicious.

Device	Activity	US	UK	VPN US→UK	VPN UK→US
TOTAL HOURS	-	28	31	26.75	27
Zmodo Doorbell	local move	1845	-	-	-
Wansview Camera	local move	114	130	1	-
Wansview Camera	power	4	2	151	-
Roku TV	local menu	11	-	3	-
Roku TV	android lan remote	1	1	-	43
Roku TV	power	1	3	3	1
Ring Doorbell	power	1	2	3	3
Google Home Mini	local voice	3	-	-	-
Google Home Mini	power	-	3	163	5
Samsung Fridge	local voice	6	-	-	-
Samsung Fridge	local viewinside	3	-	-	-
Fire TV	android lan menu	6	-	6	-
Fire TV	local voice	-	-	12	13
Echo Spot	local volume	5	-	-	-
Echo Dot	power	2	-	3	-
Echo Dot	local volume	-	-	256	-
Apple TV	local menu	17	68	12	9
Apple TV	local voice	-	2	1	3
Sous Vide Cooker	power	-	65	-	37
Osram Lightify Hub	power	-	1	4	2
Google Home	power	-	4	-	3
Echo Plus	volume	-	-	-	15
Echo Plus	power	-	-	-	3
LG TV	local off	-	-	17	-
LG TV	local voice	-	-	4	-
LG TV	android lan menu	-	-	3	-
Invoke with Cortana	local voice	-	-	4	-
Invoke with Cortana	local volume	-	-	4	-
Netatmo Weather Station	android wan graphs	-	-	-	20

Table 11: Number of detected activity instances in Idle experiments (using highly accurate models only, F1 score >0.9). Activities with fewer than 3 detected instances are not shown.

Regional comparison. By comparing the US and UK columns of Table 11 we can see both similarities and differences in detected behavior from idle experiments. Differences in power experiments (such as the Sous Vide Cooker) are explained by different reliability of the Wi-Fi in the two labs (confirmed using DHCP logs). However, we could not identify root causes for other differences, and might be due to non-determinism among common devices across the labs.

There is a noticeable difference in results when using a VPN egress, particularly smart speakers (both from Amazon and from Google) and one of the cameras (Wansview). While we were unable to identify a root cause for these differences, it is an interesting area for future work.

7.3 Uncontrolled Experiments

We use the same approach to analyze six months of data from our uncontrolled experiments generated as part of our user study in the US testbed. Most of the detected activities correspond to interactions that we confirm are popular/commonplace in our lab (e.g., using the microwave, opening/closing the fridge, moving in the lab). However, due to the large numbers of detected activities in such categories, it is difficult to manually verify them all.

Instead we focus on the most sensitive and unexpected activities flagged by our classifier, then manually attempted to trigger the same (unexpected) behavior. We describe the most interesting cases observed below.

Ring doorbell. We observed that the Ring doorbell performs a video recording action every time a user moves in front of it. However, this is unexpected behavior: the app used to set up the device does not warn the user that the doorbell performs such recording in real time, the doorbell offers no indication that recording is occurring, and the only disclosure is in fine print as part of the privacy policy [38]. Upon discovering this barely documented feature, we logged into our account to see the video, and learned that we must pay an additional monthly fee to access such recordings. We have not identified any way to turn off this feature.

Zmodo doorbell. The Zmodo doorbell uploads camera snapshots when the device is first turned on, and also when anyone moves in front of the device. This feature is undocumented, and we were unable to prevent such snapshots from being taken, nor we were able to access them.

Alexa voice assistant. A number of our user study participants complained that Alexa-enabled devices are frequently triggered during normal conversations, as if the keyword had been spoken. Upon investigation, we found that the default Alexa wake-up keyword "Alexa" is frequently triggered by many other unrelated words—much more so than the other voice assistants in our testbeds. A notable example is a sentence beginning with "I like [*s*-word]" such as "I like Star Trek." We understand that this may be a limitation of voice recognition technology, but it is still a potential privacy exposure since Amazon devices typically recognize false activations *after* sending the whole sentence to their servers (i.e., after such sentence has been permanently stored).

7.4 Takeaways

While we can only rarely positively identify unexpected behavior, such cases certainly exist, and we identified many cases by observing idle traffic. We further found notable cases of devices unexpectedly sending audio or video in our uncontrolled experiments. Our findings highlight that concerns about information exposed by IoT devices is warranted, as is further investigation into more accurate device-activity classifiers and the root causes for the inferred behavior.

8 RELATED WORK

We now review work related to IoT information exposure.

Traffic characterization. Prior efforts characterize IoT traffic from different points of view. Several focus whether encryption is used, and if so, how it is misused. Alrawi et al. [3] analyze traffic from 45 consumer IoT devices to identify unencrypted traffic and vulnerability to MITM attacks for encrypted traffic. Jia et al. [19] find weaknesses communication channels between devices and companion apps, and between those entities and their cloud service. Sivaraman et al. [43] measure encryption and authentication protocol weaknesses on a set of 24 devices.

Amar et al. [4] characterize network traffic for a handful of home IoT devices in a idle state. Sivaraman et al. [42] characterize destinations as first party and third parties, and identify protocols used by a set of 28 devices. The IoT Inspector tool [1] visualizes and characterizes volumes and destinations of IoT traffic on the same IP network.

Our study extends prior work by considering significantly more devices (81), devices that are usually ignored due to their size and cost (e.g., large appliances such as fridge, TV, washer and dryer), large number of experiments (34,586) due to our high level of automation, and our regional analysis across two testbeds located in different privacy jurisdictions.

Unexpected behavior detection. Unexpected behavior can more generally be classified as anomaly detection, a topic of significant prior work. Prior work focuses on intrusion detection systems that detect attacks by using device search engines [40], vulnerability repositories [46], and machine learning to identify known and unknown devices [20, 22, 29, 33, 42, 45]. Alternatively, related work focuses on a *policy enforcement* approach to detecting anomalies. For example, the recent Manufacturer Usage Description (MUD) [24] IETF standard proposal allows manufacturers and other parties to specify the expected behavior for IoT devices. Ongoing work addresses the problem of automatically verifying and/or enforcing the compliance of a given IoT device according to its specification, as a means to detect and block unexpected behavior [14–16]. Other approaches for detecting unexpected behavior use machine learning and other statistical techniques on the device traffic to infer the type of the IoT devices and their activities, and to profile their users [2, 41, 44]. Our contribution on unexpected behavior detection is inspired by these last approaches, but our machine learning approach uniquely considers interaction method (e.g., the use of the device locally vs through its companion app vs through a voice assistant) and includes larger numbers of training experiments.

Other relevant work. Other related work tackles information exposure for different categories of IoT devices: home/office consumer devices [5, 27], medical devices [49], smart toys [10, 39, 47] as well as home automation solutions [26, 31]. Several companies offer commercial solutions to address security issues in IoT, including privacy exposure [6–8, 11, 12, 17, 18, 21, 28, 30, 34]; however, as closed solutions little is known about their approach and effectiveness.

To summarize, our work shares many of the goals and techniques proposed in prior traffic characterization and anomaly detection work; however, our primary and differentiating goal is to increase the *scale*, *experimental rigor*, and *geographic diversity* of such analyses. We accomplish this goal by analyzing a broader range of devices over different geographic locations and privacy jurisdictions, broader range of interactions/experiments, and supplement controlled experimental data with uncontrolled experiments from a 6-month of naturally occurring user interactions with our devices.

9 CONCLUSION

Using the largest known set of controlled experiments (34,586) comprising 81 devices in the US and UK, along with uncontrolled experiments consisting of an IRB-approved user study, we are the first to quantify such information exposure across different networks, geographic regions, and interactions with devices.

We observe several promising practices: most of the devices use encryption or other encodings that protect users' PII, thus resulting in an overall minimal PII exposure in plaintext. However, even if the traffic is encrypted and without relying on MITM or any kind of IoT device modification, our analysis identifies notable cases of information exposure: 57.45 % (50.27 %) of the overall destinations contacted by the US (UK) IoT devices are third or support parties, and 56 % of the US devices and the 83.8 % of the UK devices contact destinations outside their region. We further find significant fractions of traffic use encryption or are otherwise unclassified, while for most devices plaintext traffic is rare (with notable exceptions), we also identify that encryption does not hide the kinds of interactions that cause a device to generate network traffic, in many cases allowing an eavesdropper to infer devices in consumer network and how they are used. We identify unexpected activity from devices that capture audio and video, and we find several notable cases where exposure differs based on the location. This study is a first step toward understanding information exposure by consumer IoT devices at scale. To facilitate analysis and reproducibility at even larger scales, our experiment infrastructure, code, and data are publicly available at `https://github.com/NEU-SNS/intl-iot`.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers and our shepherd Mobin Javed for their constructive feedback. The research in this paper was partially supported by grants from DHS S&T (contract #FA8750-17-2-0145), the Data Transparency Lab, Comcast Innovation Fund, and the EPSRC (Databox EP/N028260/1, DADA EP/R03351X/1, and HDI EP/R045178/1).

REFERENCES

[1] IoT Inspector. `https://iot-inspector.princeton.edu/`, 2019. (Accessed on 05/08/2019).

[2] Acar, A., Fereidooni, H., Abera, T., Sikder, A. K., Miettinen, M., Aksu, H., Conti, M., Sadeghi, A.-R., and Uluagac, A. S. Peek-a-Boo: I see your smart home activities, even encrypted! *arXiv preprint arXiv:1808.02741* (2018).

[3] Alrawi, O., Lever, C., Antonakakis, M., and Monrose, F. Sok: Security Evaluation of Home-based IoT Deployments. In *Proceedings of the IEEE Symposium on Security and Privacy (S&P)* (2019).

[4] Amar, Y., Haddadi, H., Mortier, R., Brown, A., Colley, J. A., and Crabtree, A. An analysis of home iot network traffic and behaviour. *CoRR abs/1803.05368* (2018).

[5] Apthorpe, N., Reisman, D., Sundaresan, S., Narayanan, A., and Feamster, N. Spying on the Smart Home: Privacy Attacks and Defenses on Encrypted IoT Traffic. *arXiv preprint arXiv:1708.05044* (2017).

[6] Avast. Avast Smart Life. Accessed May 8, 2019. `https://www.avast.com/en-us/iot`.

[7] Bitdefender. Bitdefender Box. Accessed May 8, 2019. `https://www.bitdefender.com/box`.

[8] Bullguard. Dojo by Bullguard. Accessed May 8, 2019. `https://dojo.bullguard.com/dojo-by-bullguard/`.

[9] Burke, S. Google admits its new smart speaker was eavesdropping on users. Accessed May 8, 2019. `http://money.cnn.com/2017/10/11/technology/google-home-mini-security-flaw`.

[10] Chu, G., Apthorpe, N., and Feamster, N. Security and Privacy Analyses of Internet of Things Children?s Toys. *IEEE Internet of Things Journal 6*, 1 (2019), 978–985.

[11] Cujo. Cujo Smart Firewall. Accessed May 8, 2019. `https://www.getcujo.com/smart-firewall-cujo/`.

[12] F-Secure. F-Secure Sense. Accessed May 8, 2019. `https://www.f-secure.com/en_US/web/home_us/sense`.

[13] Gartner Inc. Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016. Accessed May 8, 2019. `https://www.gartner.com/newsroom/id/3598917`.

[14] Hamza, A., Gharakheili, H. H., Benson, T. A., and Sivaraman, V. Detecting Volumetric Attacks on loT Devices via SDN-Based Monitoring of MUD Activity. In *Proceedings of the 2019 ACM Symposium on SDN Research* (New York, NY, USA, 2019), SOSR '19, ACM, pp. 36–48.

[15] Hamza, A., Ranathunga, D., Gharakheili, H. H., Benson, T. A., Roughan, M., and Sivaraman, V. Verifying and Monitoring IoTs Network Behavior using MUD Profiles. *arXiv preprint arXiv:1902.02484* (2019).

[16] Hamza, A., Ranathunga, D., Gharakheili, H. H., Roughan, M., and Sivaraman, V. Clear As MUD: Generating, Validating and Applying IoT Behavioral Profiles. In *Proceedings of the 2018 Workshop on IoT Security and Privacy* (New York, NY, USA, 2018), IoT S&P '18, ACM, pp. 8–14.

[17] HighIoT. Akita. Accessed May 8, 2019. `https://akita.cloud/`.

[18] IoT Defense. RATtrap. Accessed May 8, 2019. `https://www.myrattrap.com/`.

[19] Jia, Y., Xiao, Y., Yu, J., Cheng, X., Liang, Z., and Wan, Z. A Novel Graph-based Mechanism for Identifying Traffic Vulnerabilities in Smart Home IoT. In *IEEE INFOCOM 2018* (April 2018), pp. 1493–1501.

[20] Jose, A. C., Malekian, R., and Ye, N. Improving home automation security; integrating device fingerprinting into smart home. *IEEE Access 4* (2016), 5776–5787.

[21] Keezel. Keezel. Accessed May 8, 2019. `https://keezel.co/`.

[22] Kurtz, A., Gascon, H., Becker, T., Rieck, K., and Freiling, F. Fingerprinting Mobile Devices Using Personalized Configurations. In *Proceedings on Privacy Enhancing Technologies (PETS '16)* (2016), pp. 4–19.

[23] Le, A., Varmarken, J., Langhoff, S., Shuba, A., Gjoka, M., and Markopoulou, A. AntMonitor: A System for Monitoring from Mobile Devices. In *Proc. of Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data* (2015).

[24] Lear, E., Droms, R., and Romascanu, D. Manufacturer Usage Description Specification. *IETF RFC8520* (2019).

[25] Leung, C., Ren, J., Choffnes, D., and Wilson, C. Should you use the app for that?: Comparing the privacy implications of app-and web-based online services. In *Proc. of IMC* (2016).

[26] Liu, H., Li, C., Jin, X., Li, J., Zhang, Y., and Gu, D. Smart solution, poor protection: An empirical study of security and privacy issues in developing and deploying smart home devices. In *Proc. of the 2017 Workshop on Internet of Things Security and Privacy* (New York, NY, USA, 2017), IoT S&P '17, ACM, pp. 13–18.

[27] Loit, F., Sivanathant, A., Gharakheilit, H. H., Radford, A., and Sivaramant, V. Systematically evaluating security and privacy for consumer iot devices. In *Proc. of the 2017 Workshop on Internet of Things Security and Privacy* (New York, NY, USA, 2017), IoT S&P '17, ACM, pp. 1–6.

[28] Luma Home. Luma Home WiFi System. Accessed May 8, 2019. `https://lumahome.com/`.

[29] Marchal, S., Miettinen, M., Nguyen, T. D., Sadeghi, A.-R., and Asokan, N. AuDI: Towards autonomous IoT device-type identification using periodic communications. *IEEE Journal on Selected Areas in Communications* (2019).

[30] McAfee. McAfee Secure Home Platform. Accessed May 8, 2019. `https://securehomeplatform.mcafee.com`.

[31] Mi, X., Qian, F., Zhang, Y., and Wang, X. An empirical characterization of ifttt: Ecosystem, usage, and performance. In *Proc. of IMC* (New York, NY, USA, 2017), IMC '17, ACM, pp. 398–404.

[32] Michéle, B., and Karpow, A. Watch and be watched: Compromising all smart tv generations. In *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC '14)* (2014), IEEE, pp. 351–356.

[33] Msadek, M. N., Soua, R., and Engel, T. IoT Device Fingerprinting: Machine Learning based Encrypted Traffic Analysis. In *The IEEE Wireless Communications and Networking Conference (WCNC)* (2019).

[34] Norton. Norton Core. Accessed May 8, 2019. `https://us.norton.com/core`.

[35] Razaghpanah, A., Vallina-Rodriguez, N., Sundaresan, S., Kreibich, C., Gill, P., Allman, M., and Paxson, V. Haystack: In Situ Mobile Traffic Analysis in User Space. *arXiv preprint arXiv:1510.01419* (2015).

[36] Rehman, M. A., Choffnes, D., and Goldberg, S. Passport. `https://passport.ccs.neu.edu/`, 2017. (Accessed on 05/08/2019).

[37] Ren, J., Rao, A., Lindorfer, M., Legout, A., and Choffnes, D. R. ReCon: Revealing and Controlling Privacy Leaks in Mobile Network Traffic. In *Proc. of MobiSys* (2016).

[38] Ring. Privacy notice. `https://shop.ring.com/pages/privacy`, March 2018.

[39] Shasha, S., Mahmoud, M., Mannan, M., and Youssef, A. Playing with danger: A taxonomy and evaluation of threats to smart toys. *IEEE Internet of Things Journal* (2019), 1–1.

[40] Shodan. The search engine for internet-connected devices. Accessed May 8, 2019. `https://www.shodan.io/`.

[41] Singh, A., Murali, S., Rieger, L., Li, R., Hommes, S., State, R., Ormazabal, G., and Schulzrinne, H. HANZO: Collaborative Network Defense for Connected Things. In *2018 Principles, Systems and Applications of IP Telecommunications (IPTComm)* (Oct 2018), pp. 1–8.

[42] Sivanathan, A., Habibi Gharakheili, H., Loi, F., Radford, A., Wijenayake, C., Vishwanath, A., and Sivaraman, V. Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics. *IEEE Transactions on Mobile Computing* (2018), 1–1.

[43] Sivaraman, V., Gharakheili, H. H., Fernandes, C., Clark, N., and Karliychuk, T. Smart IoT Devices in the Home: Security and Privacy Implications. *IEEE Technology and Society Magazine 37*, 2 (June 2018), 71–79.

[44] Takbiri, N., Houmansadr, A., Goeckel, D. L., and Pishro-Nik, H. Matching Anonymized and Obfuscated Time Series to Users? Profiles. *IEEE Transactions on Information Theory 65*, 2 (2019), 724–741.

[45] Thangavelu, V., Divakaran, D. M., Sairam, R., Bhunia, S. S., and Gurusamy, M. DEFT: A Distributed IoT Fingerprinting Technique. *IEEE Internet of Things Journal 6*, 1 (2019), 940–952.

[46] US-CERT. CVE: The Standard for Information Security Vulnerability Names. Accessed May 8, 2019. `http://cve.mitre.org/`.

[47] Valente, J., and Cardenas, A. Security and privacy of smart toys. In *Proc. of the 2017 Workshop on Internet of Things Security and Privacy* (New York, NY, USA, 2017), IoT S&P'17, ACM, pp. 19–24.

[48] Warren, Tom. Amazon explains how Alexa recorded a private conversation and sent it to another user. Accessed on 05/25/2018. `https://www.theverge.com/2018/5/24/17391898/amazon-alexa-private-conversation-recording-explanation`.

[49] Wood, D., Apthorpe, N., and Feamster, N. Cleartext data transmissions in consumer iot medical devices. In *Proc. of the 2017 Workshop on Internet of Things Security and Privacy* (New York, NY, USA, 2017), IoT S&P '17, ACM, pp. 7–12.