Information Exposure From Consumer IoT Devices: A Multidimensional Network-Informed Approach

Jingjing Ren, Daniel J. Dubois, David Choffnes
Anna Maria Mandalari, Roman Kolcun, Hamed Haddadi

Northeastern University
Imperial College London
Motivation

7+ billion IoT devices deployed worldwide

• Typical home IoT devices have access to private information

They may listen to you (e.g., smart speakers)

---

Amazon Workers Are Listening to What You Tell Alexa

A global team reviews audio clips in an effort to help the voice-activated assistant respond to commands.
Motivation

7+ billion IoT devices deployed worldwide

- Typical home IoT devices have access to private information

They may listen to you (e.g., smart speakers)  They may watch you (e.g., smart doorbells)

A security flaw leaves Ring doorbells and cameras vulnerable to spying
Motivation

7+ billion IoT devices deployed worldwide

- Typical home IoT devices have access to private information

They may listen to you (e.g., smart speakers)
They may watch you (e.g., smart doorbells)
They may know what you watch (e.g., smart TVs)

How to Turn Off Smart TV Snooping Features

Smart TVs collect data about what you watch with a technology called ACR. Here’s how to turn it off.
Motivation

7+ billion IoT devices deployed worldwide

- Typical home IoT devices have access to private information

They may listen to you (e.g., smart speakers)  They may watch you (e.g., smart doorbells)  They may know what you watch (e.g., smart TVs)

- They can (by definition) access the Internet and therefore may expose private information

- Lack of understanding on what information they expose, on when they expose it, and to whom

- Lack of understanding of regional differences (e.g., GDPR)
IoT Privacy Exposure in a Smart Home

**Goal 1:** What is the destination of IoT network traffic?

**Identify destinations:** First-party, Non first-party, Eavesdroppers

**Geolocate destinations:** same vs. different privacy jurisdiction

**Goal 2:** What information is sent?

**Search IoT traffic** for private information exposure

E.g., video from cameras, audio from smart speakers, user activities, ...

**Goal 3:** Does a device expose information unexpectedly?

Information exposure we expect vs. information exposure we observe
Challenges for Measuring IoT Privacy

Difficult to measure exposed information for IoT

- Closed systems
- MITM fails most of the time

Our contribution: information inference from traffic patterns

Difficult to perform IoT experiments and generalize

- Lack of automation and emulation tools
- Lack of standard testbed

Our contribution: a testbed for running repeatable semi-automated IoT experiments at a scale (software and data available online)
Testbeds

US: Northeastern University

UK: Imperial College London
Selecting Home IoT Devices

- **Criteria**: category; features; popularity; US & UK markets

---

**Amazon Cam**
- **Flux Bulb**
- **Xiaomi Strip**
- **Philips Bulb**
- **LG TV**
- **Invoke Speaker**

**Amcrest Cam**
- **Behmor Brewer**
- **GE Microwave**
- **Samsung Dryer**

**Lefun Cam**
- **Samsung Fridge**
- **Samsung Washer**

**Luohe Cam**
- **Smartr iKettle**
- **Xiaomi Rice Cooker**

**Micro7 Cam**
- **D-Link Sensor**

---

**Blink Cam**
- **Blink Hub**
- **Ring Doorbell**
- **Wanswiew Cam**
- **Yi Cam**
- **Insteon Hub**
- **Lightify Hub**
- **Philips Hue Hub**
- **Sengled Hub**
- **Smartthings Hub**
- **Xiaomi Hub**
- **Magichome Strip**
- **Nest T-stat**

**ZModo Bell**
- **Wink2 Hub**

---

**Wanswiew Cam**
- **TP-Link Bulb**
- **TP-Link Plug**
- **WeMo Plug**
- **Apple TV**
- **Fire TV**
- **Roku TV**
- **Samsung TV**
- **Echo Dot**
- **Echo Spot**
- **Echo Plus**
- **Google Home Mini**
- **Anova Sousvide**
- **Xiaomi Cleaner**

---

**Bosiwo Cam**
- **D-Link Cam**
- **WiMaker Cam**
- **Xiaomi Cam**
- **Honeywell T-stat**
- **Allure Speaker**
- **Google Home**
- **Netatmo Weather**
- **Smarter Brewer**

---

N=46
N=26
N=35

---

20 Cameras 13 Smart Hubs 15 Home Automation 9 TVs 11 Speakers 13 Appliances 81 Total
Design of Experiments

34,586 experiments (92.6% automated)

- **Controlled interactions**
  - Manual (repeated 3 times)
  - Automated (repeated 30 times)
    - Text-to-speech to smart assistants (Alexa/Google/Cortana/Bixby)
    - Monkey instrumented control from Android companion apps

- **Idle**: ~112 hours

- **Uncontrolled interactions (US Only)**
  - IRB-approved user study
  - 36 participants, 6 months Sep/2018 to Feb/2019

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>power on/off the device</td>
</tr>
<tr>
<td>Voice</td>
<td>voice commands for speakers</td>
</tr>
<tr>
<td>Video</td>
<td>record/watch video</td>
</tr>
<tr>
<td>On/Off</td>
<td>turn on/off bulbs/plugs</td>
</tr>
<tr>
<td>Motion</td>
<td>move in front of device</td>
</tr>
<tr>
<td>Others</td>
<td>change volume, browse menu</td>
</tr>
</tbody>
</table>
Data Collection Methodology

- Monitor all traffic at the **router**
  - per-device
  - per-experiment

Internet traffic is the only signal that (by definition) all IoT devices produce
Research Questions

• What is the destination of IoT network traffic?

• What information is sent?

• Does a device expose information unexpectedly?
What Is the Destination?

1. DNS response
2. HTTP headers
3. TLS handshake

Network Traffic

Destination IP

Second-Level Domain (SLD)

IP Address

Whois database (or common sense)

4. IP Owner

Organization

Passport

https://passport.ccs.neu.edu

Geolocation

First party

Non-first party

Same jurisdiction

Different jurisdiction
What Non-First Parties Are Contacted?

- Number of devices contacting non-first party organizations

<table>
<thead>
<tr>
<th>Organization</th>
<th>US 46</th>
<th>UK 35</th>
<th>US Common 24</th>
<th>UK Common 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>31</td>
<td>24</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Google</td>
<td>14</td>
<td>9</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Akamai</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Microsoft</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Netflix</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Kingsoft</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21Vianet</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Alibaba</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Beijing Huaxiay</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AT&amp;T</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

High reliance on cloud and CDN providers

Nearly all TVs contact Netflix w/o it being logged in or used

Chinese cloud providers

Regional differences
Destination Characterization

Categories

US Testbed

Categories

Dest. Country

Categories

UK Testbed

Alibaba Cloud

US Lab

Cameras US

Home Automation US

TV US

Smart Hubs US

Audio US

US

Cameras UK

TV UK

Smart Hubs UK

Home Automation UK

Overseas

Audio UK

EU

CN
HK
KR
CA
IN
UK Lab
Most devices contact outside testbeds’ privacy jurisdictions*
Research Questions

• What is the destination of IoT network traffic?
• What information is sent?
• Does a device expose information unexpectedly?
Unencrypted Information Leakage

- MagicHome LED
- Samsung Fridge
- Insteon Hub
- Xiaomi Camera

PII: MAC Address
unencrypted!

PII: MAC Address and Timestamps
unencrypted (plus evidence of a video stream) each time motion is detected!

Other unencrypted content
- Device toggle actions (e.g., on-off)
- Firmware updates
- Metadata pertaining to initial device set up
How Much Traffic Is Encrypted?

Percentage of traffic by device category (US)

- **Unencrypted traffic**: we can analyze exposed information directly
- **Rest of the traffic**: can we *infer* information?

- Speakers
- Smart Hubs
- Appliances
- Home Automation
- TVs
- Cameras
Can We Infer User Activity from Network Traffic?

Hypothesis:
Eavesdroppers may infer activity information even from encrypted traffic

Idea: Given the traffic patterns of an activity, look for similar patterns

Feasibility of a solution: use supervised machine learning

**ML APPROACH**
- Random Forest Tree Classifier
- Features (assuming encrypted):
  - packet size, inter-arrival times
  - min, max, mean, deciles, ...

**ML EVALUATION**
- 10-fold cross validation
  - Iterated 10 times
- F1 score (val=[0,1]):
  - 0 is the worst, 1 is the best
Device Activity Inference

We consider an activity inferable when F1-score is >0.75

Significant amounts of device activities are inferable

- Inferable activities can be exploited by eavesdroppers (e.g., ISP)
- But they also offer an opportunity for researchers to audit device behavior
Research Questions

• What is the destination of IoT network traffic?

• What information is sent?

• Does a device expose information unexpectedly?
Cases of Unexpected Behavior

Popular doorbells

Video recording on detected motion (cannot be disabled)
Cases of Unexpected Behavior

Popular smart TVs

Contact Netflix, Google, and Facebook unexpectedly

Popular doorbells

Video recording on detected motion (cannot be disabled)

FINANCIAL TIMES

Smart TVs sending private data to Netflix and Facebook

Researchers from Northeastern University and Imperial College London found that a number of smart TVs, including those made by Samsung and LG, and the streaming dongles Roku and Amazon’s FireTV were sending out data such as location and IP address to Netflix and third-party advertisers.
Cases of Unexpected Behavior

Popular doorbells
- Video recording on detected motion (cannot be disabled)

Popular smart TVs
- Contact Netflix, Google, and Facebook unexpectedly

Alexa-enabled devices
- Frequently falsely triggered (e.g. "I like Star Trek")

FINANCIAL TIMES
Smart TVs sending private data to Netflix and Facebook

Researchers from Northeastern University say the Amazon Echo may respond to dialogue other than its wake word.
Cases of Unexpected Behavior

- Other notable cases of activities detected when **idle**
  - Cameras reporting **motion** in absence of movement
  - Devices spontaneously **restarting** or reconnecting

Popular doorbells
- Video recording on detected motion (cannot be disabled)

Popular smart TVs
- Contact Netflix, Google, and Facebook unexpectedly

Alexa-enabled devices
- Frequently falsely triggered (e.g. "I like Star Trek")
Conclusion

• First step towards more large-scale IoT measurements:
  • 81 devices, 2 countries, 34K experiments

• Main results:
  • 57% (50%) of destinations of the US (UK) devices are not first-party
  • 56% (84%) of the US (UK) devices have at least one destination abroad
  • 89% (86%) of the US (UK) devices are vulnerable to at least one activity inference
  • Activity inference can be used to identify unexpected activities

• Impact:
  • Press coverage
  • Working with manufacturers to understand information exposure
  • Testbed/analysis framework and data are publicly available
    https://moniotrlab.ccis.neu.edu/imc19/