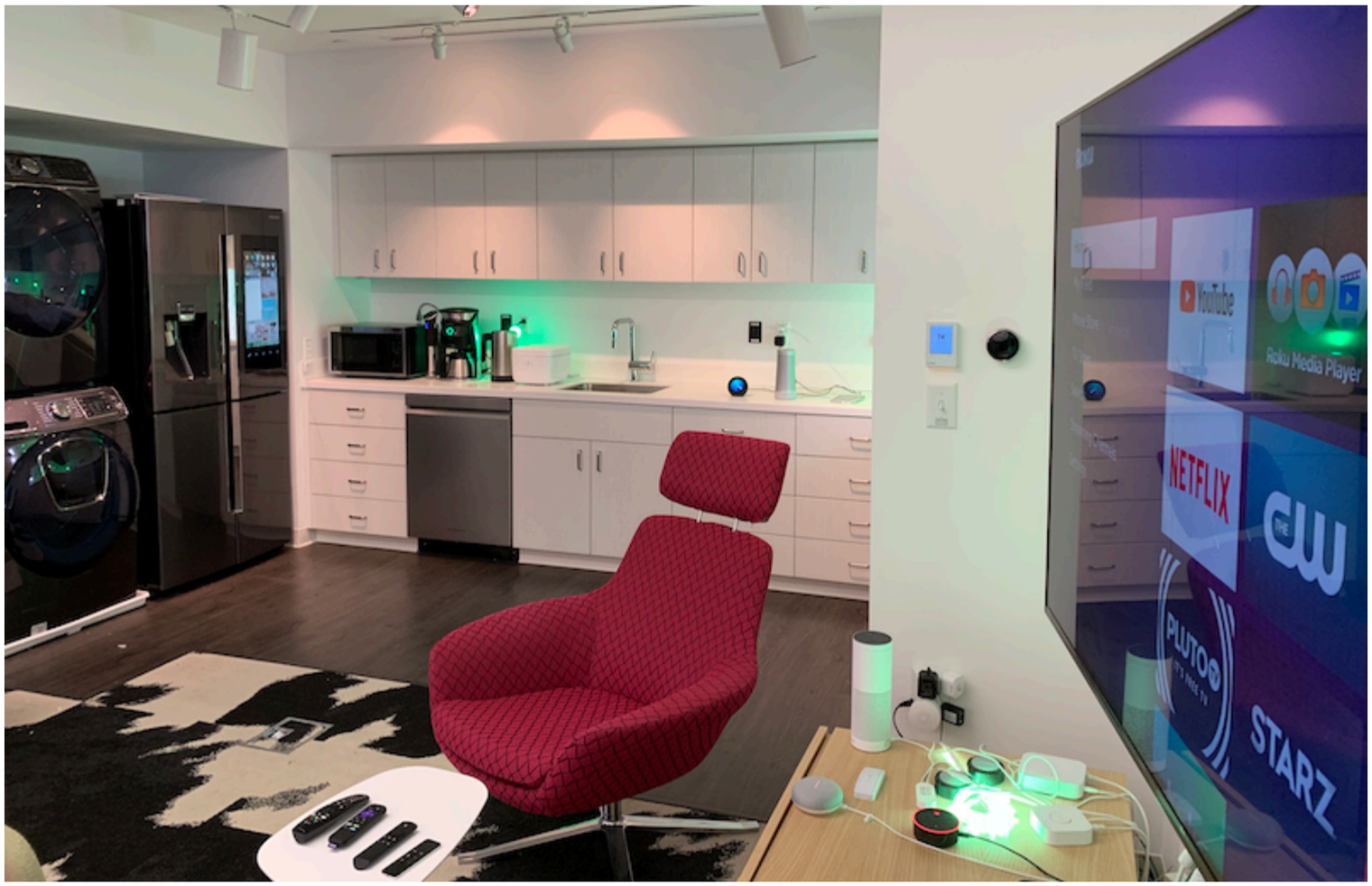


Information Exposure From Consumer IoT Devices: A Multidimensional Network-Informed Approach

Jingjing Ren, **Daniel J. Dubois**, David Choffnes

Anna Maria Mandalari, Roman Kolcun, Hamed Haddadi

Northeastern University

Imperial College London

Homeland Security Science and Technology

DATA TRANSPARENCY LAB

COMCAST INNOVATION FUND

EPSRC Pioneering research and skills

Motivation

7+ billion IoT devices deployed worldwide IOT ANALYTICS MARKET INSIGHTS FOR THE INTERNET OF THINGS

- Typical home IoT devices have access to private information

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

Bloomberg

Technology

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

IOT ANALYTICS
MARKET INSIGHTS FOR THE INTERNET OF THINGS

- 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)

DIGITAL TRENDS

A security flaw leaves Ring doorbells and cameras vulnerable to spying

Motivation

7+ billion IoT devices deployed worldwide

IOT ANALYTICS
MARKET INSIGHTS FOR THE INTERNET OF THINGS

- 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)

CR Consumer Reports

Electronics & Computers / Audio & Video / TVs / How To Turn Off Smart TV Snooping Features

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

IOT ANALYTICS
MARKET INSIGHTS FOR THE INTERNET OF THINGS

- 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?

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

Search IoT traffic for private information exposure

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

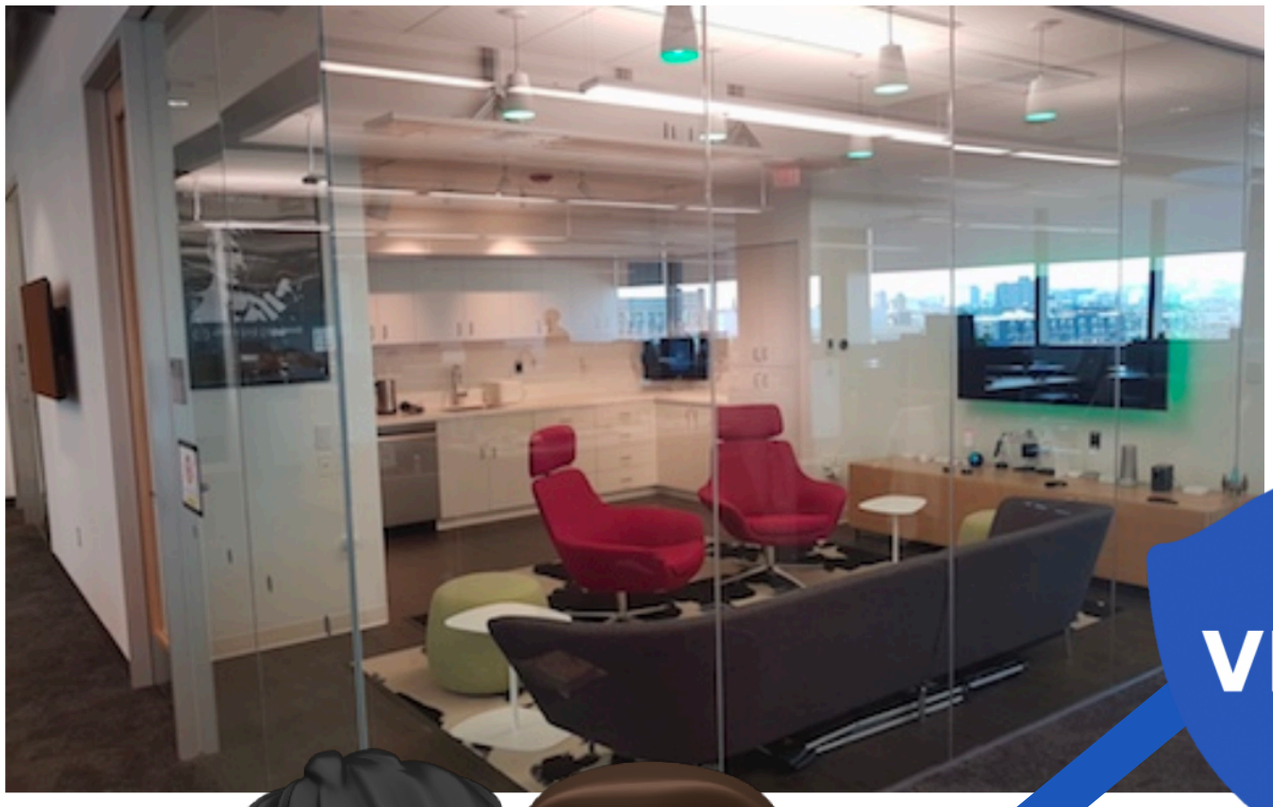
Difficult to perform IoT experiments and generalize

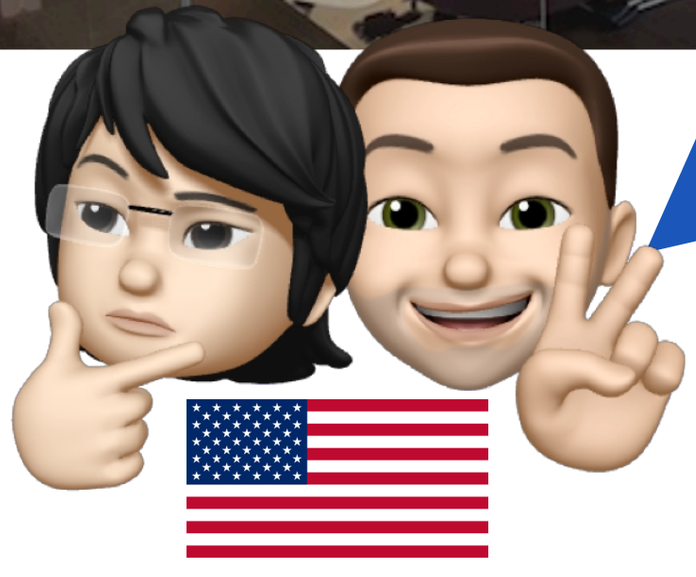
- Lack of automation and emulation tools
- Lack of standard testbed

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

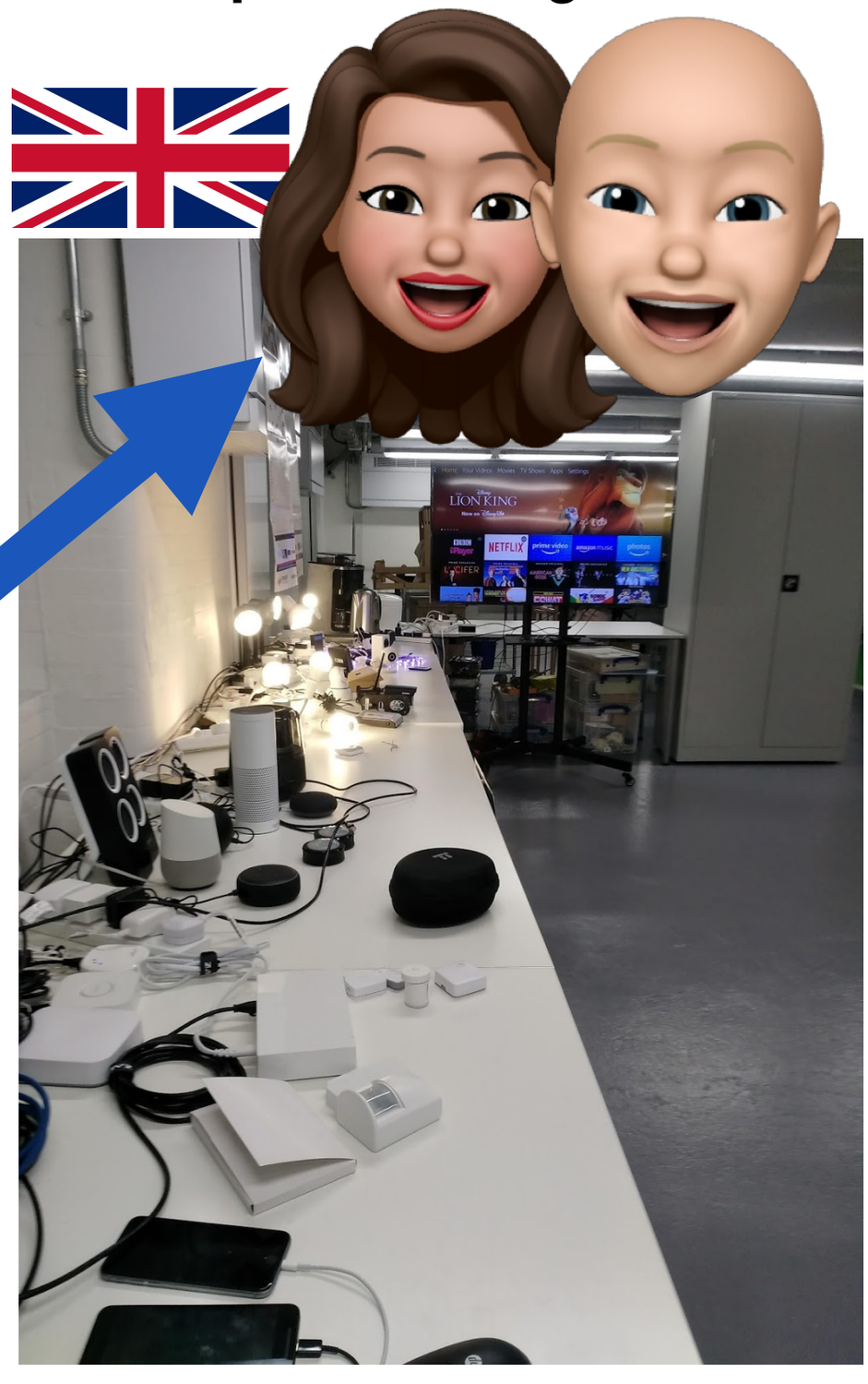
Testbeds

US: Northeastern University





UK: Imperial College London



VPN

Selecting Home IoT Devices

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

Flux Bulb
Xiaomi Strip
Philips Bulb
LG TV
Amazon Cam
Invoke Speaker
Amcrest Cam
Behmor Brewer
Lefun Cam
GE Microwave
Luohe Cam
Samsung Dryer
Micro7 Cam
Samsung Fridge
ZModo Bell
Samsung Washer
Wink2 Hub
Smarter iKettle
D-Link Sensor
Xiaomi Rice Cooker

N=46

Blink Cam
TP-Link Bulb
Blink Hub
TP-Link Plug
Ring Doorbell
WeMo Plug
Wanswiew Cam
Apple TV
Yi Cam
Fire TV
Insteon Hub
Roku TV
Lightify Hub
Samsung TV
Philips Hue Hub
Echo Dot
Sengled Hub
Echo Spot
Smartthings Hub
Echo Plus
Xiaomi Hub
Google Home Mini
Magichome Strip
Anova Sousvide
Nest T-stat
Xiaomi Cleaner

N=26

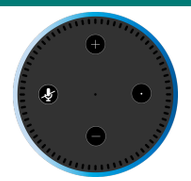
Bosiwo Cam
D-Link Cam
WiMaker Cam
Xiaomi Cam
Honeywell T-stat
Allure Speaker
Google Home
Netatmo Weather
Smarter Brewer

N=35

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







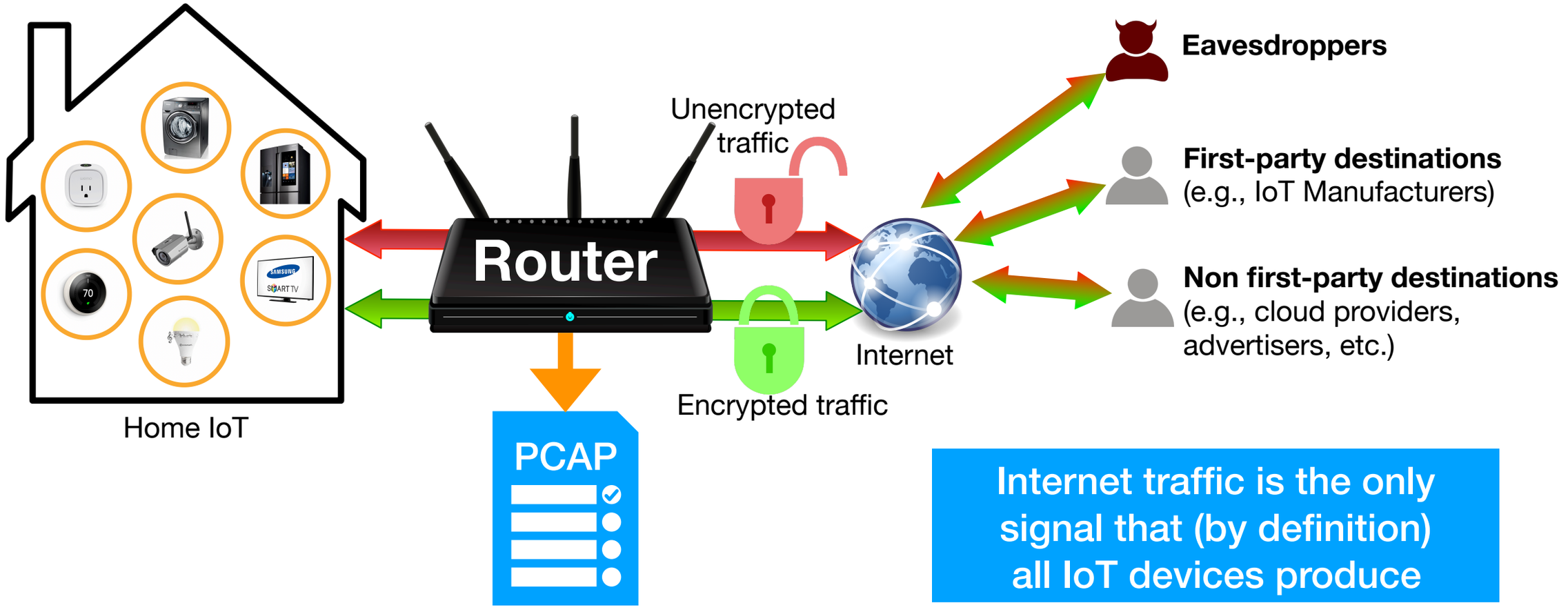
Design of Experiments

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

Activity	Description
Power	power on/off the device
Voice	voice commands for speakers
Video	record/watch video
On/Off	turn on/off bulbs/plugs
Motion	move in front of device
Others	change volume, browse menu

Data Collection Methodology



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

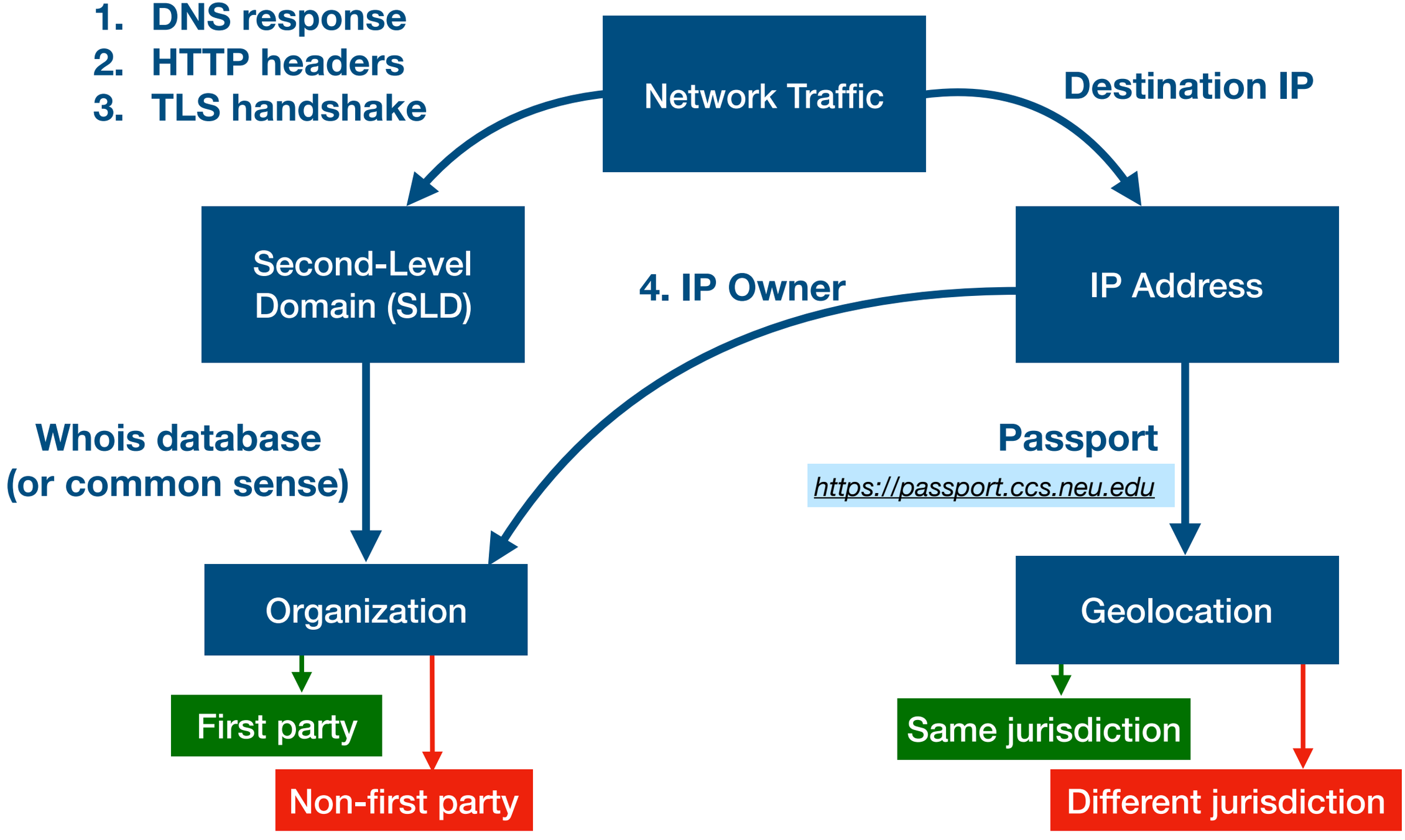
- Monitor all traffic at the **router**
  - per-device
  - per-experiment

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



What Non-First Parties Are Contacted?

- Number of devices contacting non-first party organizations

High reliance on cloud and CDN providers

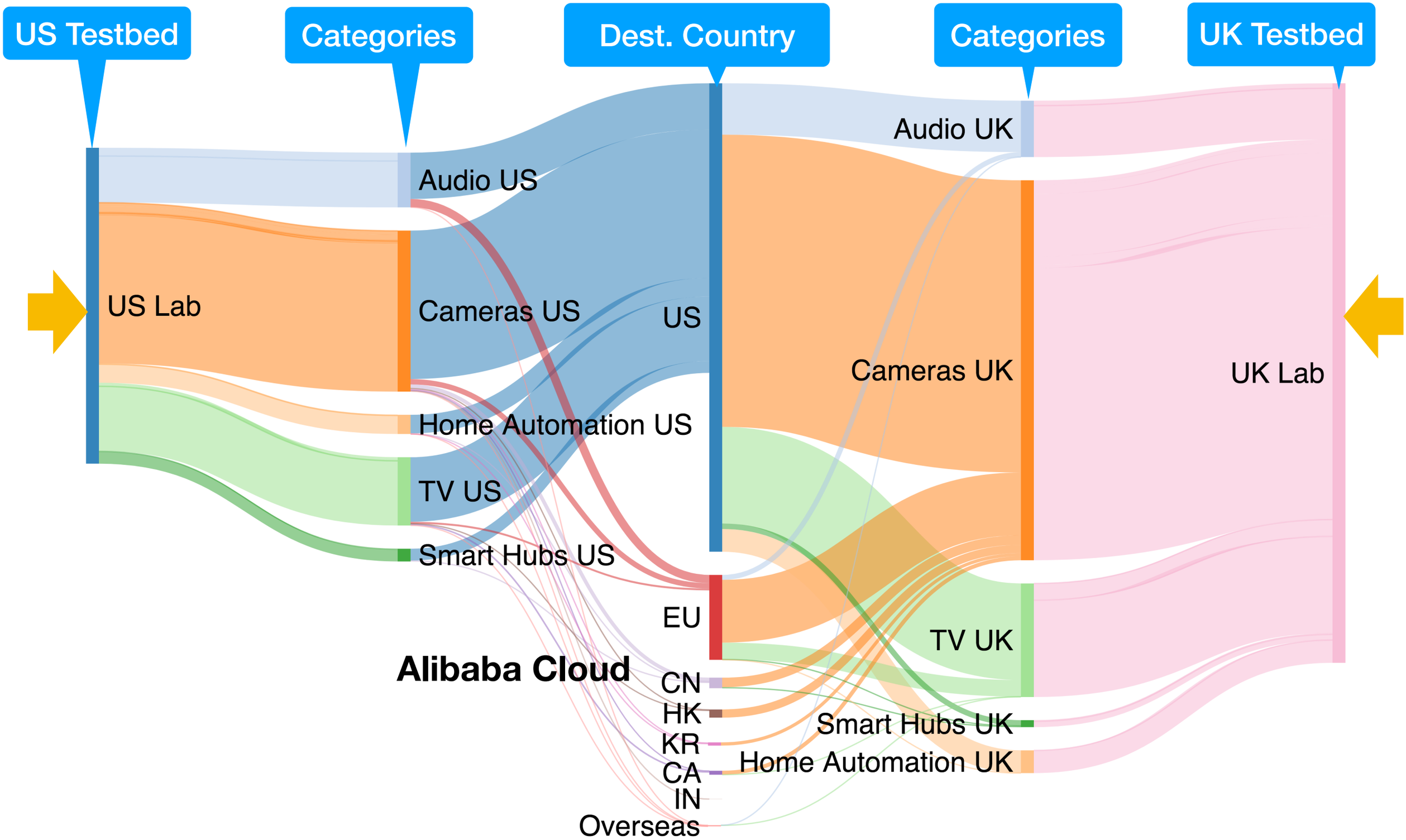
Nearly all TVs contact Netflix w/o it being logged in or used

Chinese cloud providers

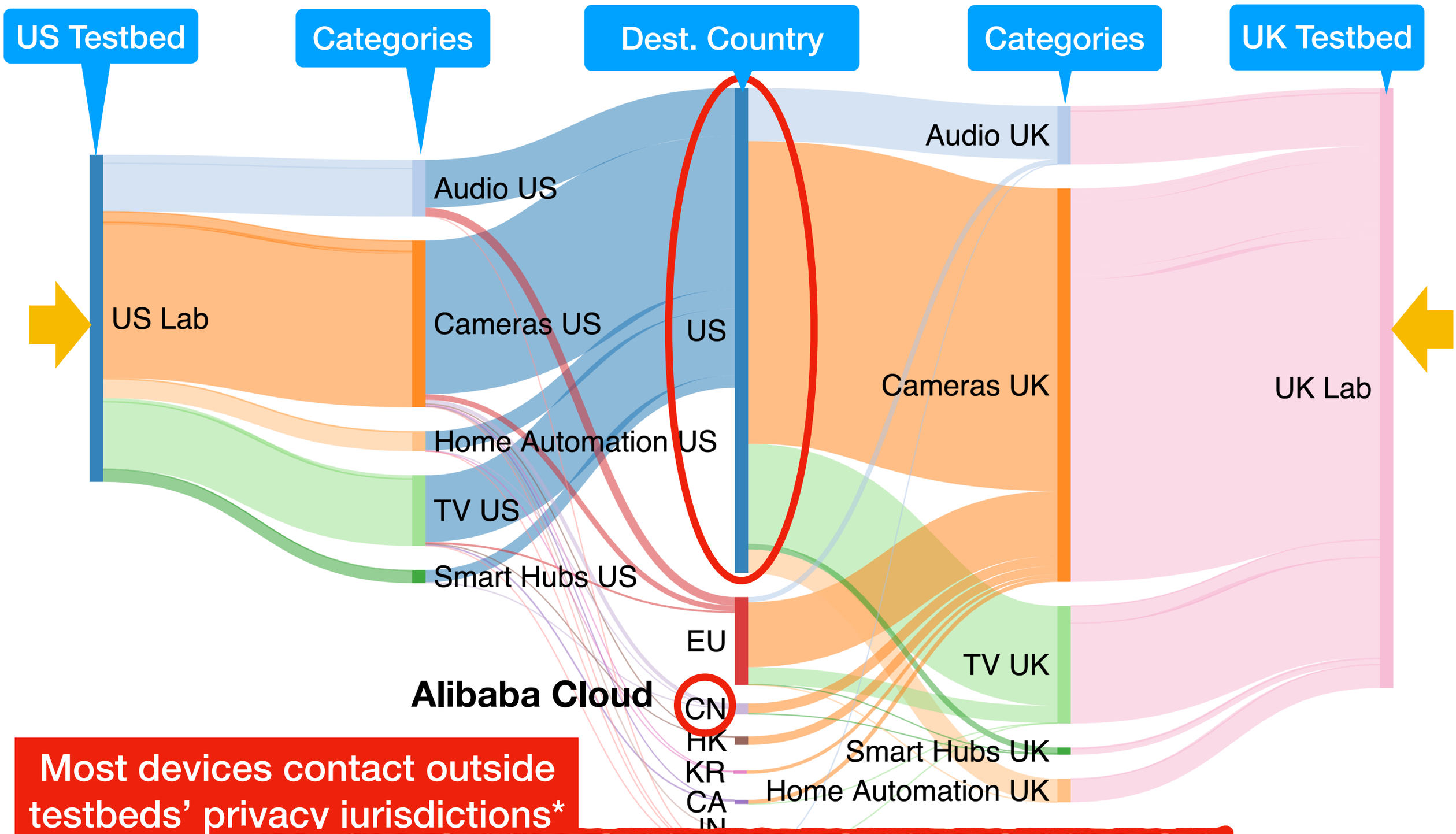
Organization	US 46	UK 35	US Common 24	UK Common 24
Amazon	31	24	16	17
Google	14	9	10	8
Akamai	10	6	6	5
Microsoft	6	4	1	1
Netflix	4	2	3	2
Kingsoft	3	3	1	1
21Vianet	3	3	1	1
Alibaba	3	4	2	2
Beijing Huaxiay	3	3	1	1
AT&T	2	0	1	1

Regional differences

Destination Characterization



Destination Characterization



Most devices contact outside testbeds' privacy jurisdictions*

BBC WORLD NEWS **Would you recognise yourself from your data?**

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

PII: MAC Address unencrypted!

Insteon Hub

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

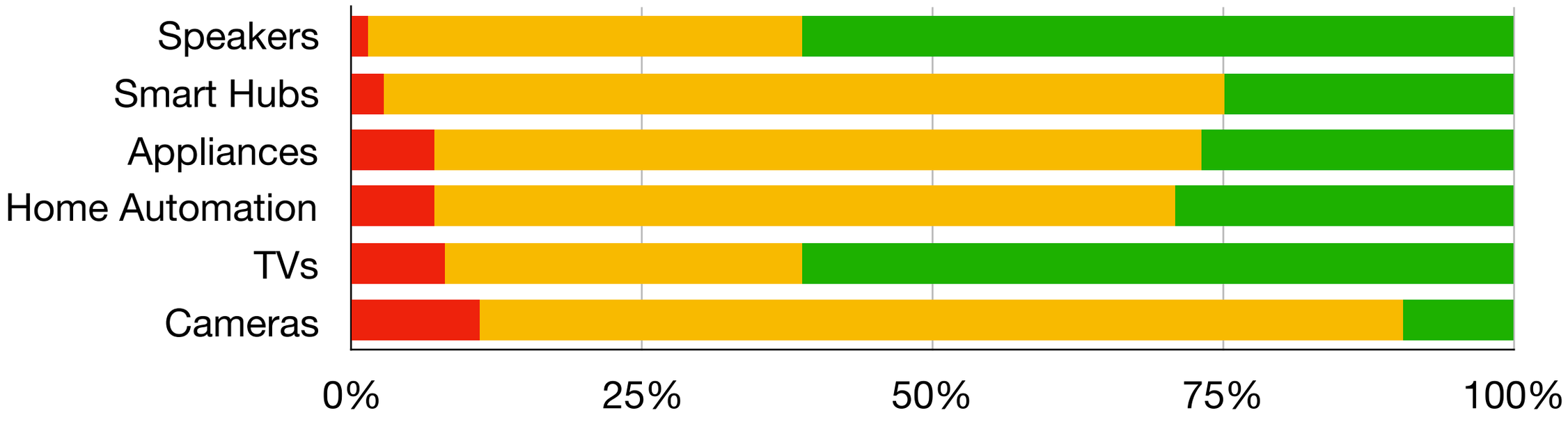
Xiaomi Camera

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 Unencrypted Unknown Encrypted traffic by device category (US)



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

Can We Infer User Activity from Network Traffic?

Hypothesis:

Eavesdroppers may infer **activity information** even from encrypted traffic

- Interaction method (local, app, or voice?)
- Functionality (e.g., toggling a light)

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

Feasibility of a solution: use supervised machine learning

ML APPROACH

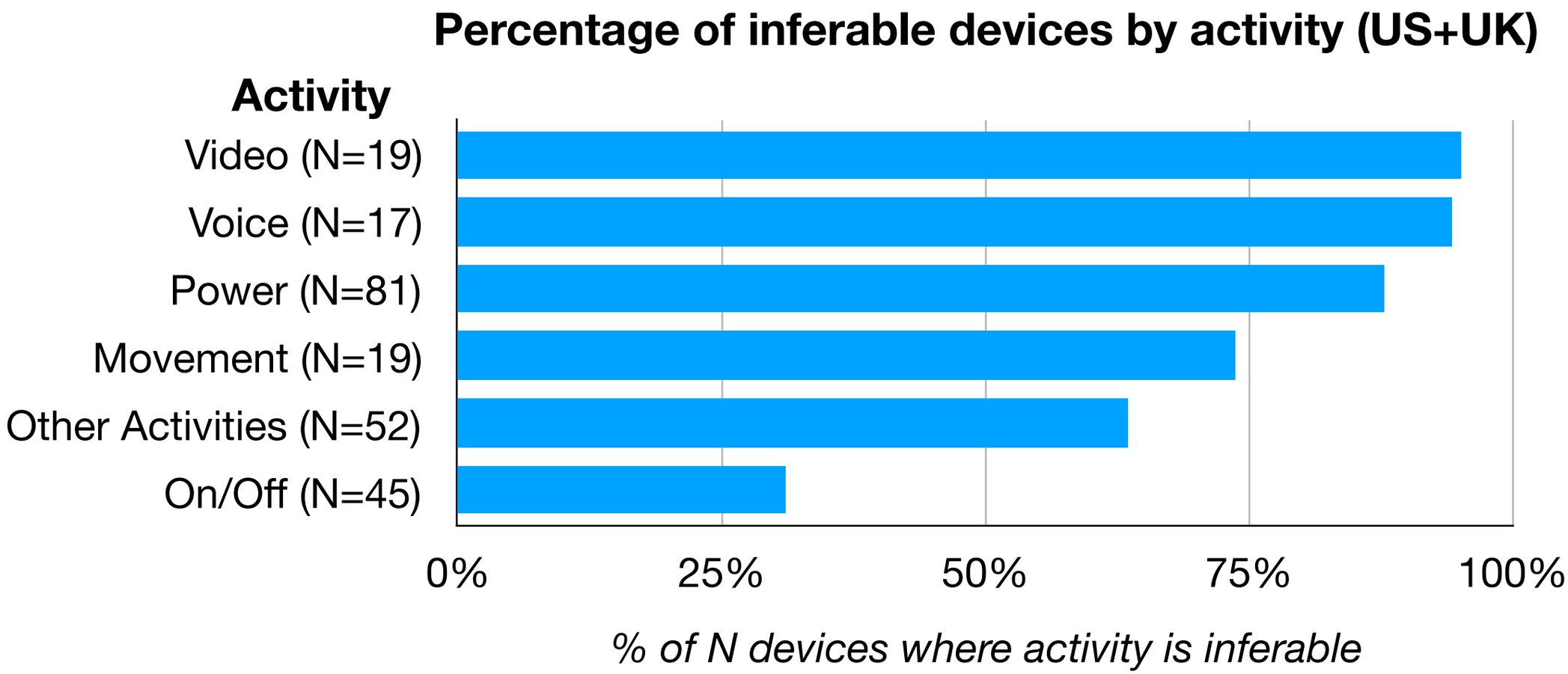
- Random Forest Tree Classifier
- Features (*assuming encrypted*):
  - packet size, inter-arrival times
  - min, max, mean, deciles, …

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 doorbells

Video recording on detected motion (cannot be disabled)

Popular smart TVs

Contact **Netflix**, **Google**, and **Facebook** unexpectedly

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 S**tar Trek")

FINANCIAL TIMES

Smart TVs sending private data to Netflix and Facebook

Researchers from North
that a number of smart
the streaming dongles R
as location and IP addre

CR Consumer Reports

Electronics & Computers / Audio & Video / Smart Speakers / Smart Speakers That Listen When They Shouldn't

Smart Speakers That Listen When They Shouldn't

Researchers at Northeastern University say the Amazon Echo may respond to dialogue other than its wake word

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")

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

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 FT FINANCIAL TIMES BBC WORLD NEWS CR Consumer Reports
  - Working with manufacturers to understand information exposure LG zmodo amazon
  - **Testbed/analysis framework and data are publicly available**

    https://moniotrlab.ccis.neu.edu/imc19/