Databox as a Platform for Monitoring IoT Devices at the Edge

Anna Maria Mandalari
Coseners
July 2019

Imperial College London

Privacy in a Connected World

- Mobile and Internet-of-Things (IoT) devices
  - Rich sensors
  - Ubiquitous connectivity
- Privacy when connected
  - Transparency
  - Control

Defining “Privacy” in This Work

- **Personal information**
  - Personally Identifiable Information (PII)
  - Multimedia
    - *e.g. private photos*
  - **Device activity**
    - *e.g. streaming videos*

- **Transport security**
  - Encrypted vs Plaintext
- **Destinations**
  - Third party: *e.g. trackers*
  - Other party: *e.g. cloud hosts*
  - Foreign jurisdictions

Goal

Is it possible to measure privacy exposure from IoT devices by analysing the network traffic they generate?

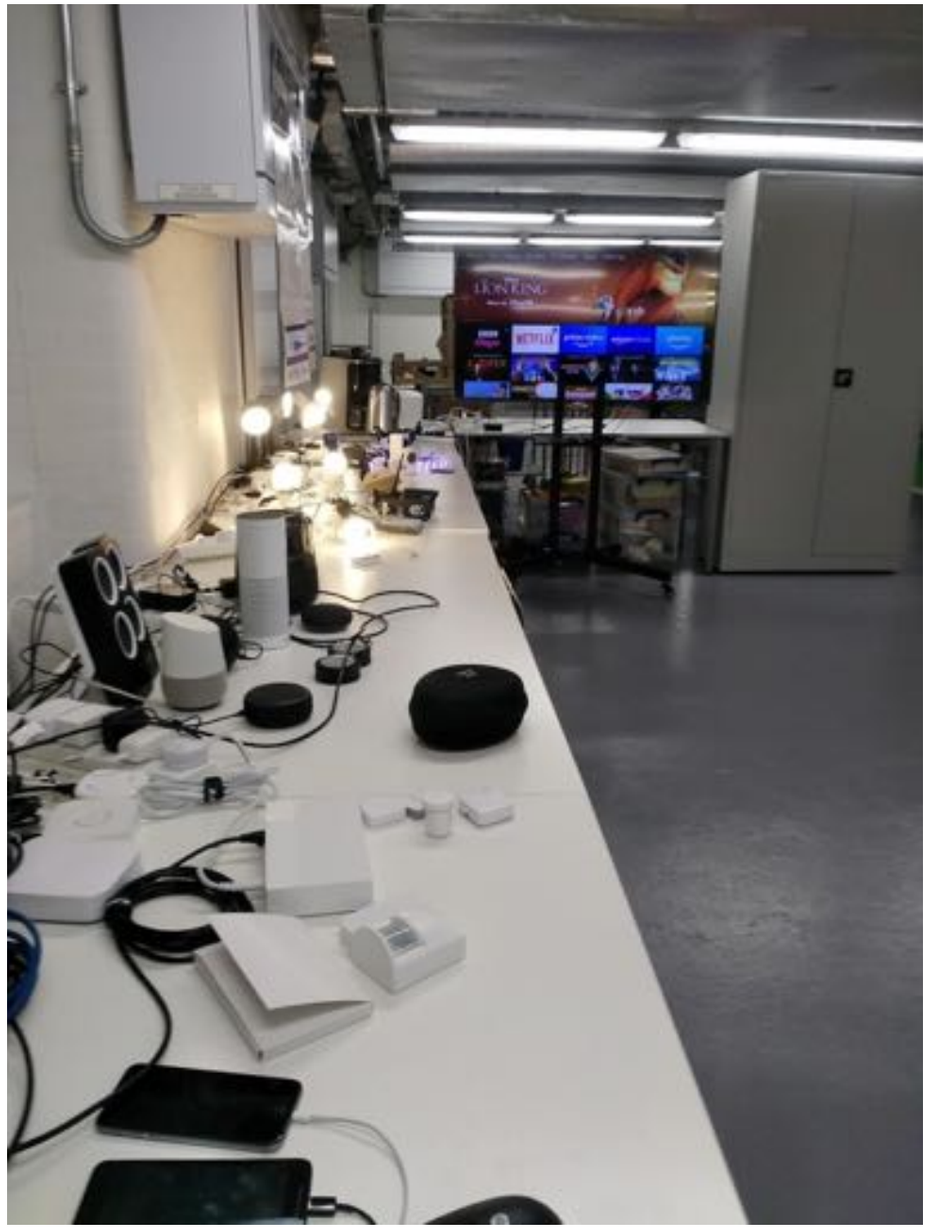
We aim to develop systems that:

- automatically identify personal information exposure
- analyse those corresponding privacy issues from multiple perspectives
- Give control back to the users and reshaping the IoT ecosystem

IoT Testbed

US: Northeastern University

UK: Imperial College London



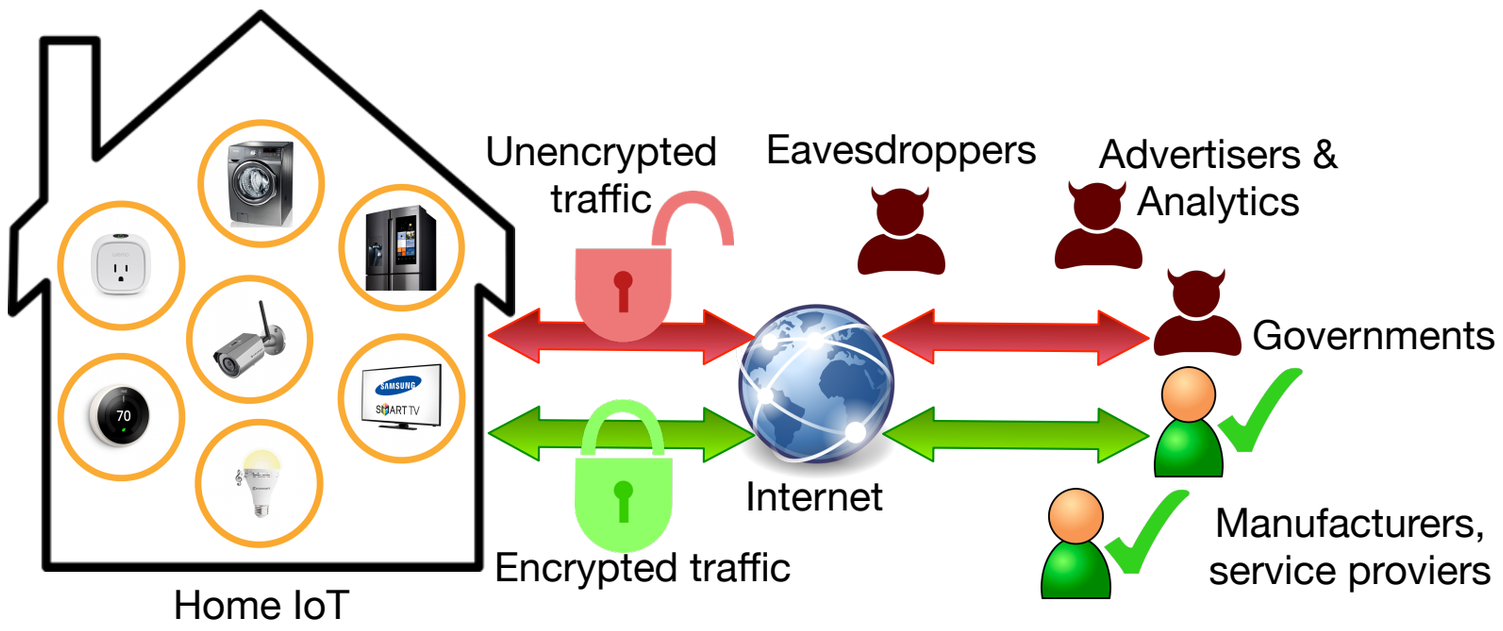
Threat Model

- Personal Information: Stored, Sensor, or Activity data
- Parties: First, Support, Third
- Privacy Concerns
  - Information goes to non-first parties/different jurisdictions
  - Information goes to first party unexpectedly
  - Activity data inferred by non-first parties

Design of Experiments

- Idle: during night
- Controlled Interaction
  - Manual (3 times)
  - Automated (>30 times) => *to detect when certain activity happens*
    - Text-to-speech to smart assistants (Alexa/Hi Google)
    - Monkey instrumented control from Android apps

34,586 experiments



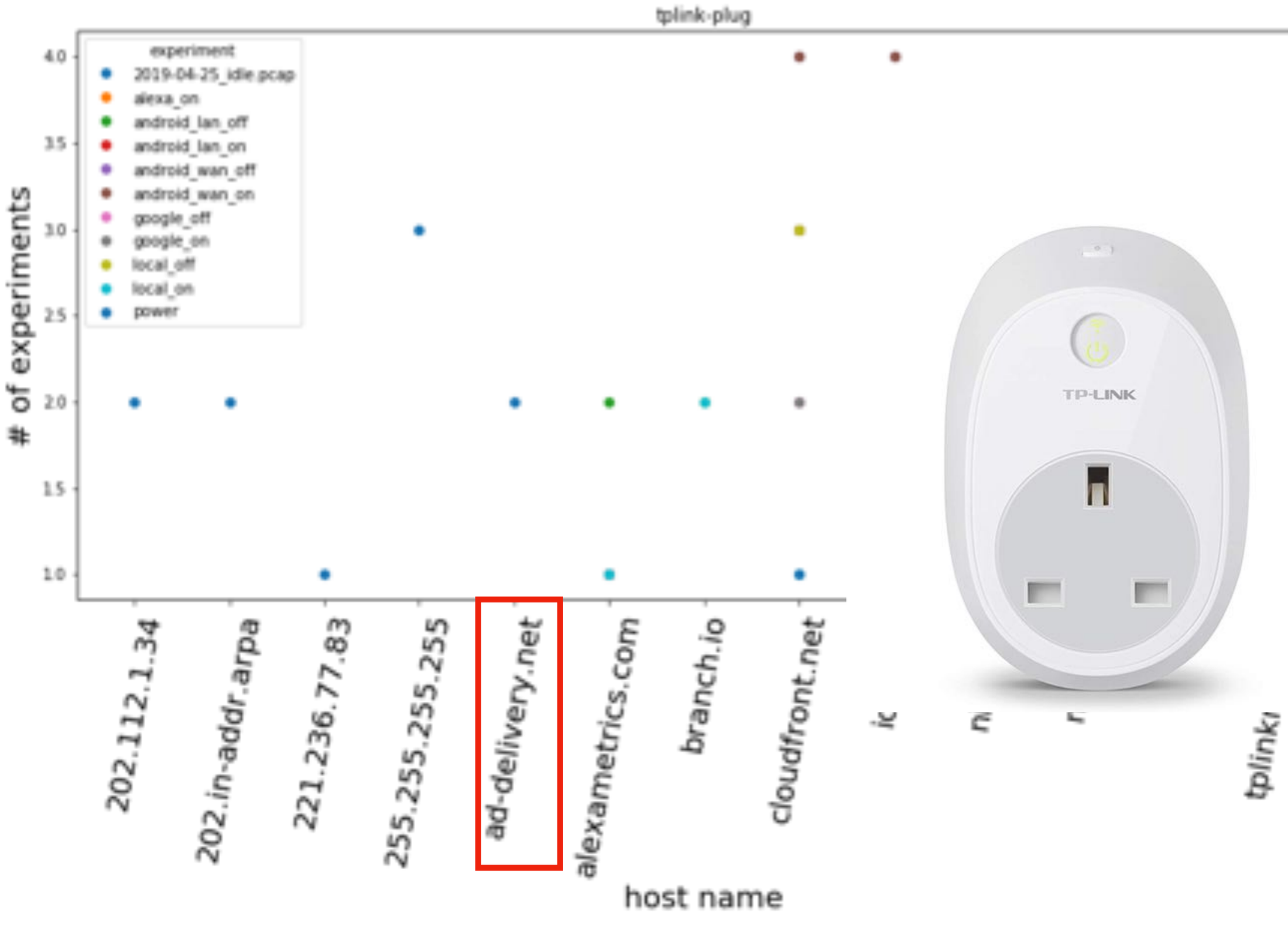
Value of the Study

- Difference between the two regions
- Unique dataset and set of experiments

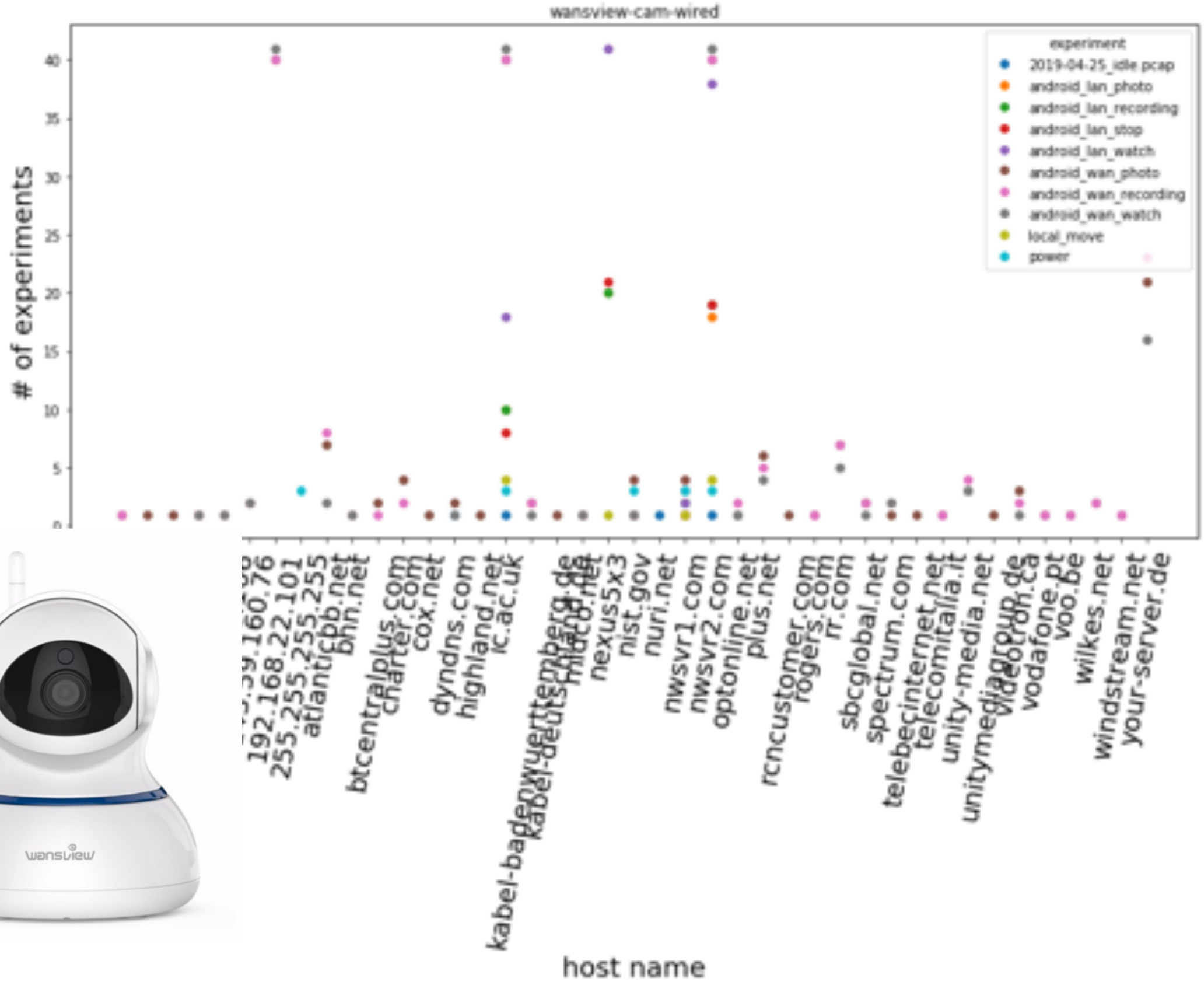
Data: Destinations

- Device
- IP
- Host
- Amount of traffic sent and received
- Lab
- Experiment
- Network

Destinations



Destinations



Who Are Contacted by Many Devices?

Organization	US 46	UK 35	US Common 24	UK Common 24
Amazon	31	24	17	17

Nearly All TV contacts Netflix w/o preconfigured

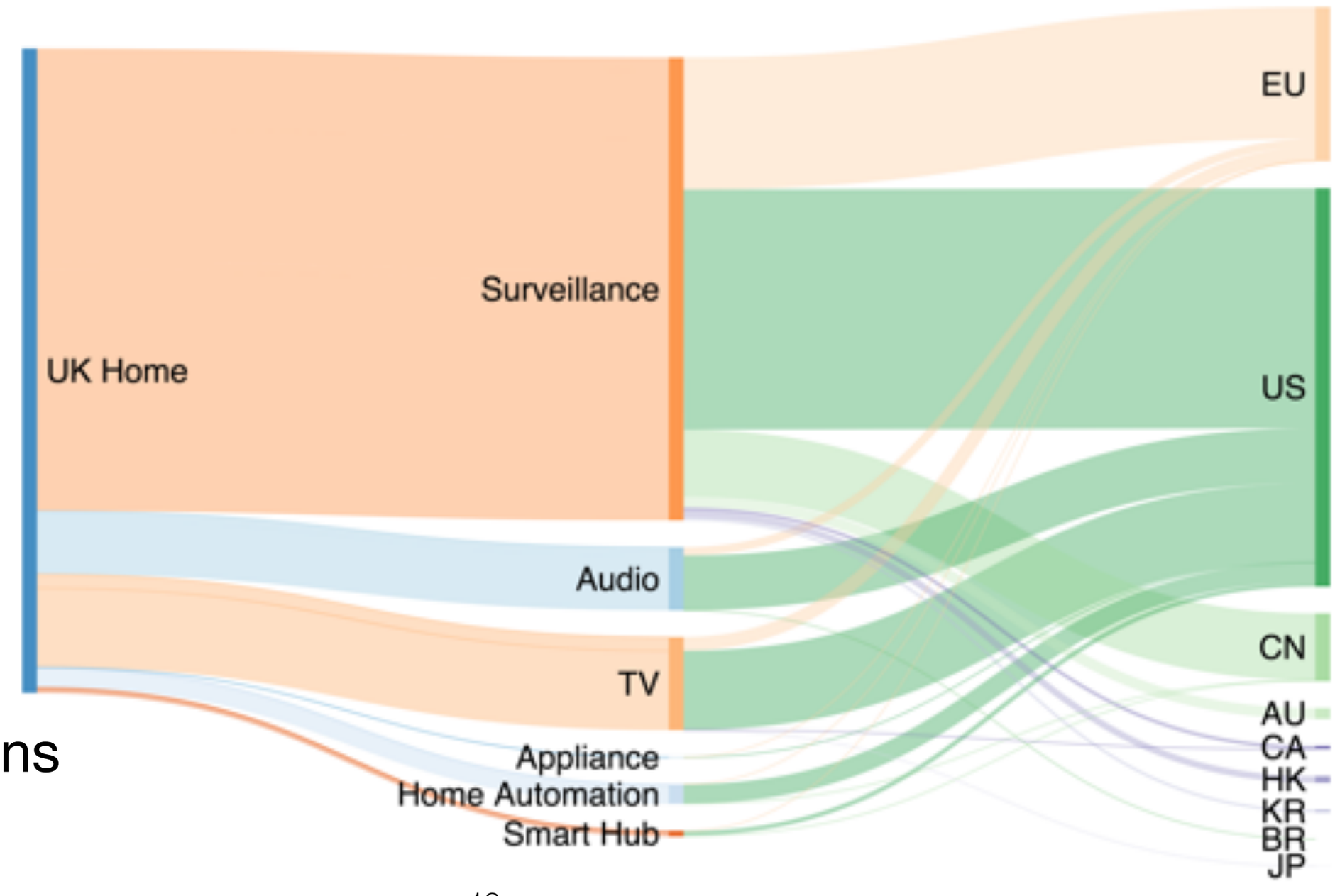
High reliance on AWS, followed by Google, Microsoft for hosting

Chinese cloud providers

- Non-first party organizations receive information from many IoT devices
  - US devices tends to contact more

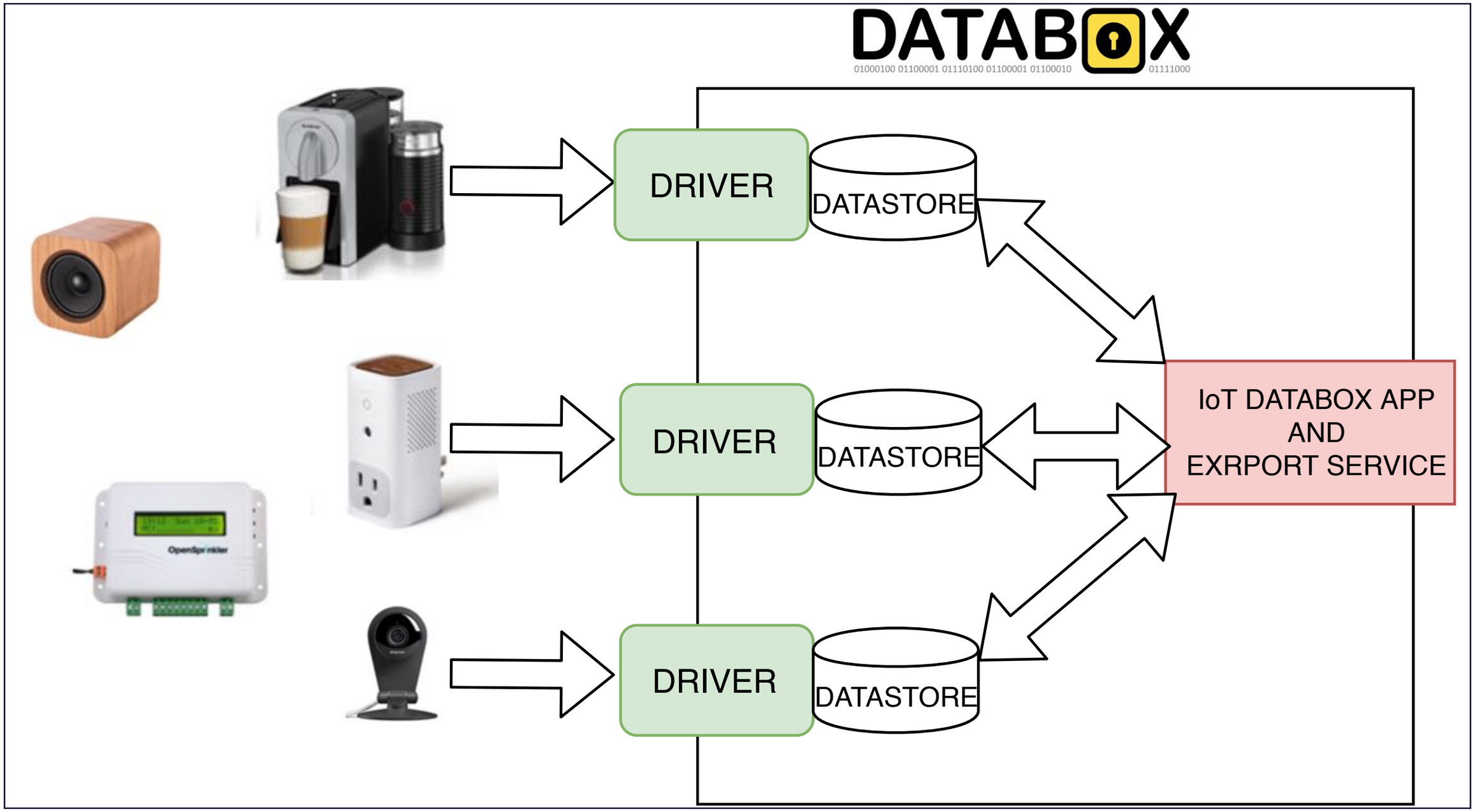
Geolocation

- Band width = bytes of traffic
  - UK Lab: 8 overseas (UK) + US
  - Overseas: Mostly China *(Alibaba Cloud)*



- **Takeways**
  - Many devices contact outside testbeds' privacy jurisdictions

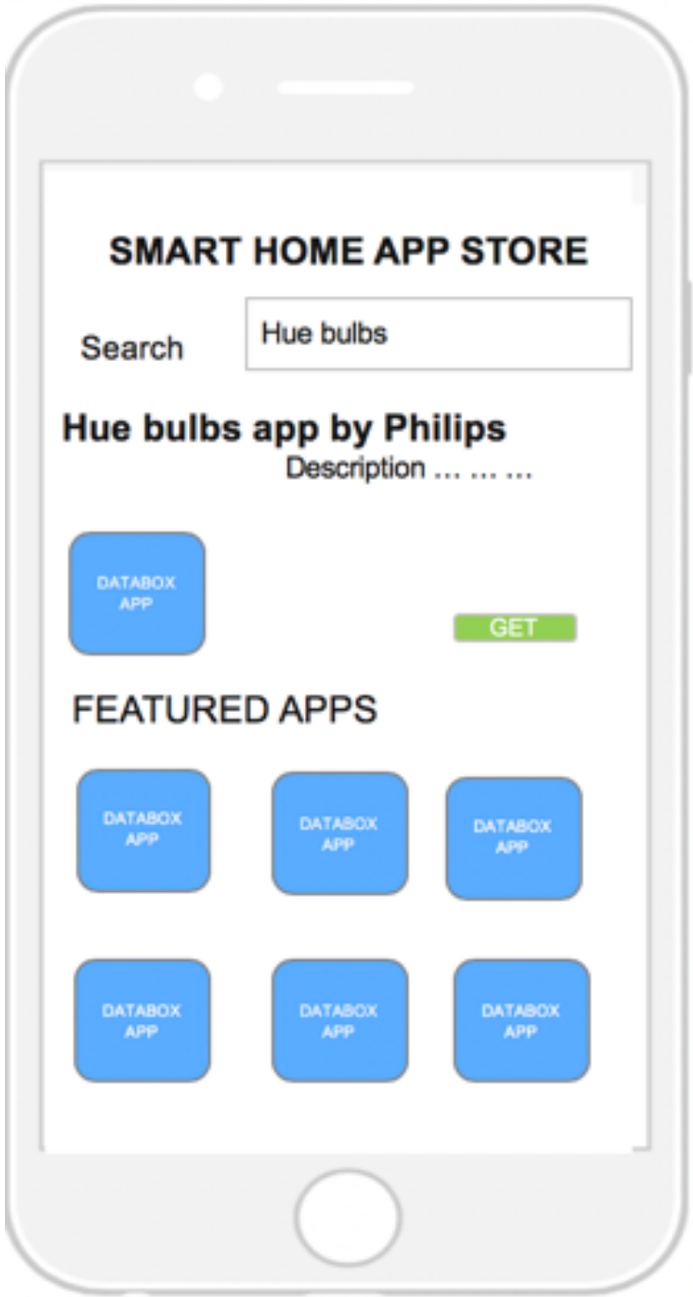
Solutions: Databox!



Solutions: IoT Databox App

A user-side application running locally allowing to monitor the IoT devices, interact with them, understand their operation.

Solutions: IoT Databox App



Solutions: IoT Databox App

1. Detecting and Limiting Unnecessary Communication for Consumer IoT Devices

2. Do IoT devices communicate with destinations which are not critical or essential to their operations?

3. Are there any patterns or trends on the unnecessary destinations among different devices?

4. Regional differences

Threat Model

- Personal Adversary: unnecessary destinations

- Victim: IoT device user

- Scope: consumer IoT

Challenges

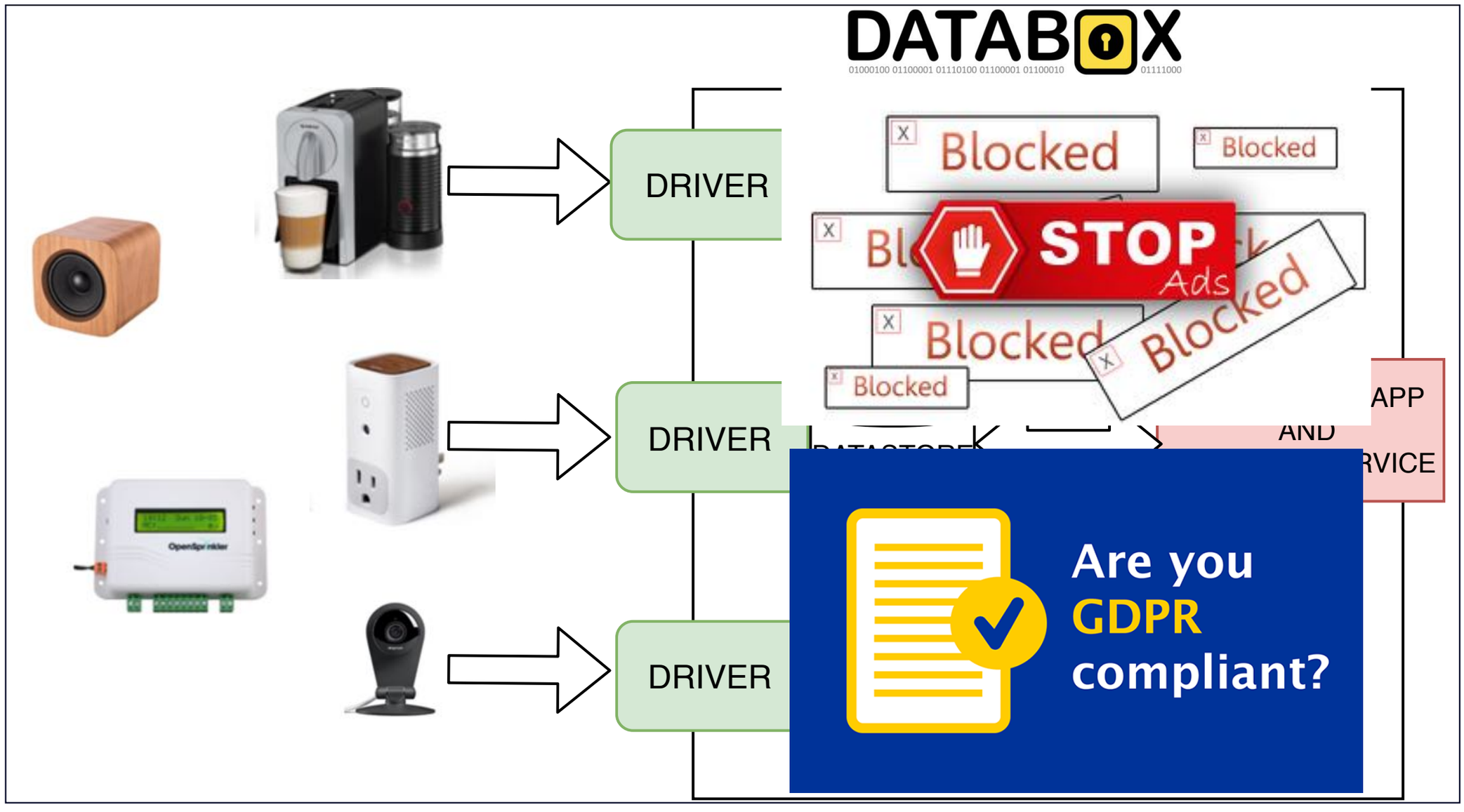
- Self-validating experiments

- Combinatorial problem: blocking a destination may make other destinations appear/ disappear.

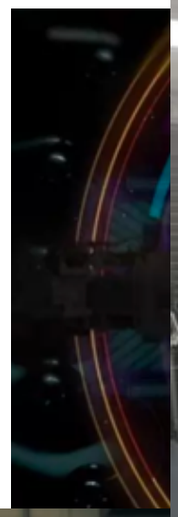
- Blocking update servers

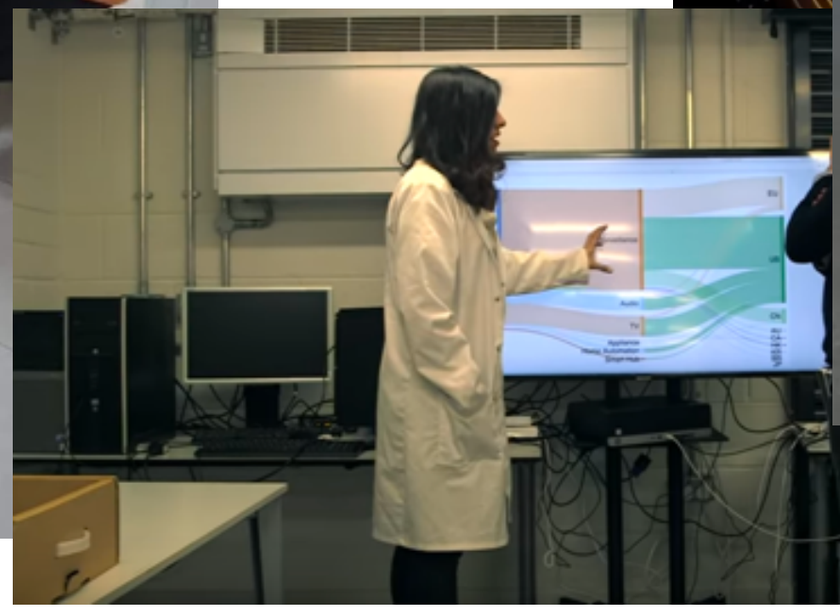
But this is just the begging…

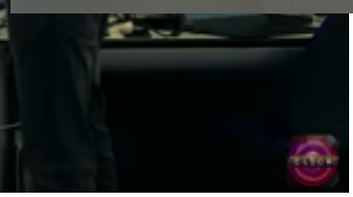
- Add other functionalities to the app



BBC World News







BBC News Technology: Would you recognise yourself from your data?

https://www.bbc.com/news/technology-48434175

BBC Click: Who has my data?

https://www.bbc.co.uk/iplayer/episode/m0005cx6/click-gdpr-one-year-on

https://www.youtube.com/watch?v=32gV9AEQCII