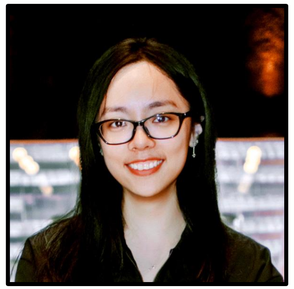
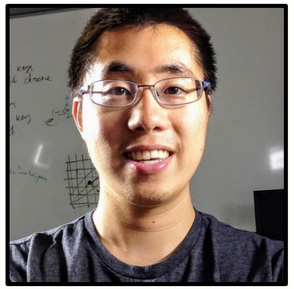
Can Allowlists Capture the Variability of Home IoT Device Network Behavior?






Weijia He, Kevin Bryson, Ricardo Calderon, Vijay Prakash, Nick Feamster, Danny Yuxing Huang, Blase Ur

THE UNIVERSITY OF CHICAGO

Security, Usability, & Privacy Education & Research

DARTMOUTH

NYU

Internet of Things Products














Security Shortcomings of Internet of Things Devices

What is the Mirai Botnet?

The Mirai malware exploits security holes in IoT devices, and has the potential to harness the collective power of millions of IoT devices into botnets, and launch attacks.

DDoS attacks using IoT bots have jumped five-fold in 12 months, says report

By **James Blackman** June 7, 2023

Internet of Things (IoT) | IoT

Kaspersky Reveals Alarming IoT Threats and Dark Web DDoS Boom

Blocklists vs. Allowlists

- Blocklist
  - ✕
  - ✕
  - ✕
  - ✕

  (All other traffic allowed)

Blocklists vs. Allowlists

- Blocklist

  ✗

  ✗

  ✗

  ✗

  (All other traffic allowed)

- Allowlist

  ✓

  ✓

  ✓

  ✓

  (All other traffic blocked)

Blocklists vs. Allowlists

- Blocklist
  - ✗
  - ✗
  - ✗
  - ✗

  (All other traffic allowed)

- Allowlist
  - ✓
  - ✓
  - ✓
  - ✓

  (All other traffic blocked)

Manufacturer Usage Description (MUD)

NIST SPECIAL PUBLICATION 1800-15

Securing Small-Business and Home Internet of Things (IoT) Devices:

Mitigating Network-Based Attacks Using Manufacturer Usage Description (MUD)

Can Allowlists Capture the Variability of Home IoT Device Network Behavior?

We Measured 24 Popular Internet of Things Products














Problem Formulation

- Specify **endpoints** with which a device may communicate
- Threat model: Remote attacker aims to compromise IoT devices, but cannot compromise vendor's backend infrastructure nor poison DNS

Terminology

- **Product:** Amazon Ring
- **Device:** One specific Amazon Ring

Crowdsourced Data

- IoT Inspector: Network flows from over 5,000 homes
  - Vendor and product (human-labeled), remote IP address & port
  - Identified hostnames using passive monitoring and reverse DNS
- Our subset: 24 products (3,461 devices)

Measuring the Feasibility of Allowlists

- Preliminary in-lab study of individual devices
- Analyze variability of network traffic in crowdsourced data
- Simulate allowlists based on crowdsourced data
- Verify functionality of crowdsourced allowlists in our lab

Measuring the Feasibility of Allowlists

- Preliminary in-lab study of individual devices
- Analyze variability of network traffic in crowdsourced data
- Simulate allowlists based on crowdsourced data
- Verify functionality of crowdsourced allowlists in our lab

Lab Study of Individual Devices

- Collect network traffic while exercising key functionality
- Later factory reset the device and enforced that allowlist
  - On the same network in our lab in the US
  - Tunneled through a VPN in Germany

Lab Study of Individual Devices

- Vary endpoint representation:
  - **Domain:** Second-level domain name (e.g., `amazon.com`)
  - **Hostname:** Fully qualified domain name
  - **Pattern:** Clustered and abstracted hostnames as regular expressions

Lab Study of Individual Devices

- Within the same US lab
  - 6 / 24 devices lost some functionality using domains
  - 10 / 24 devices lost some functionality using hostnames

Lab Study of Individual Devices

- Within the same US lab
  - 6 / 24 devices lost some functionality using domains
  - 10 / 24 devices lost some functionality using hostnames
- VPN to Germany
  - 11 / 24 devices lost some functionality using domains
  - 16 / 24 devices lost some functionality using hostnames

Measuring the Feasibility of Allowlists

- Preliminary in-lab study of individual devices
- **Analyze variability of network traffic in crowdsourced data**
- Simulate allowlists based on crowdsourced data
- Verify functionality of crowdsourced allowlists in our lab

Sources of Variability

- Load balancing and content distribution networks
  - `guc3-accesspoint-a-f002.ap.spotify.com`
  - `d37ju0xanoz6gh.cloudfront.net`
- Regionalization
- DNS
- Variable remote ports
  - Amazon Rings: `amazonaws.com` on 1,617 different remote ports

Measuring the Feasibility of Allowlists

- Preliminary in-lab study of individual devices
- Analyze variability of network traffic in crowdsourced data
- **Simulate allowlists based on crowdsourced data**
- Verify functionality of crowdsourced allowlists in our lab

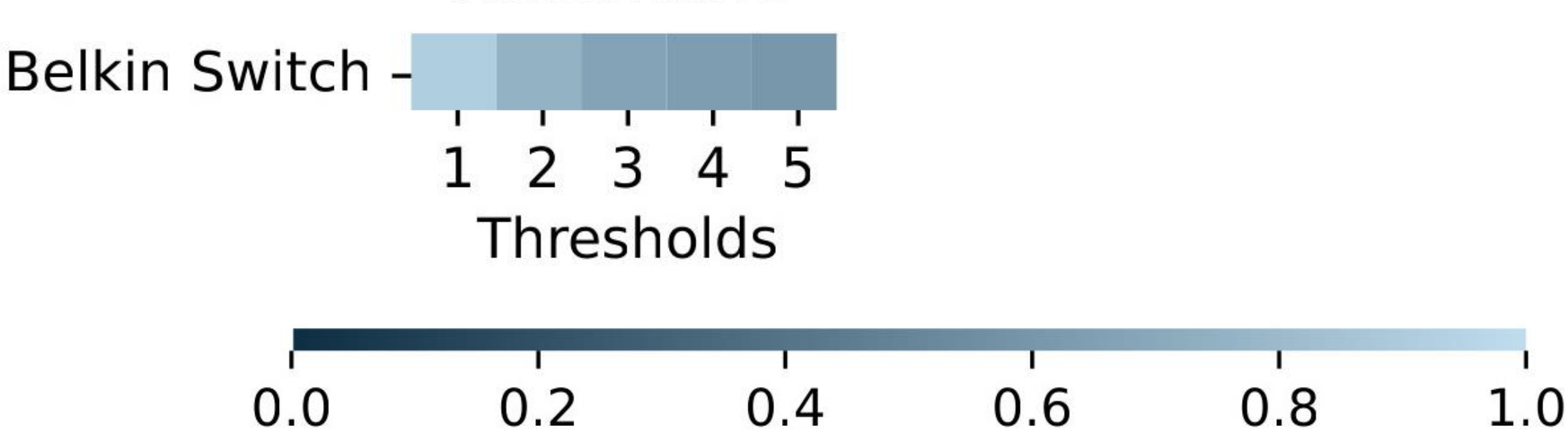
Retrospectively Simulate Crowdsourced Allowlists

- Metric: MFOF (Median Fraction of Observed Flows) captures the proportion of a device's flows observed in training data

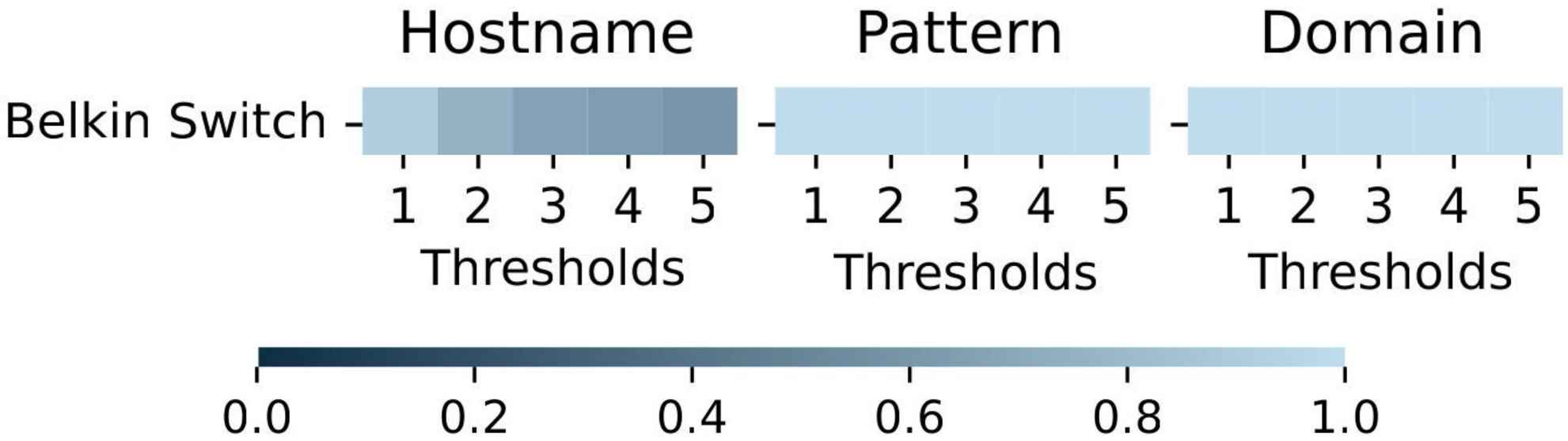
Retrospectively Simulate Crowdsourced Allowlists

- Metric: MFOF (Median Fraction of Observed Flows) captures the proportion of a device's flows observed in training data
- Vary host representation, sample size, thresholds, and more

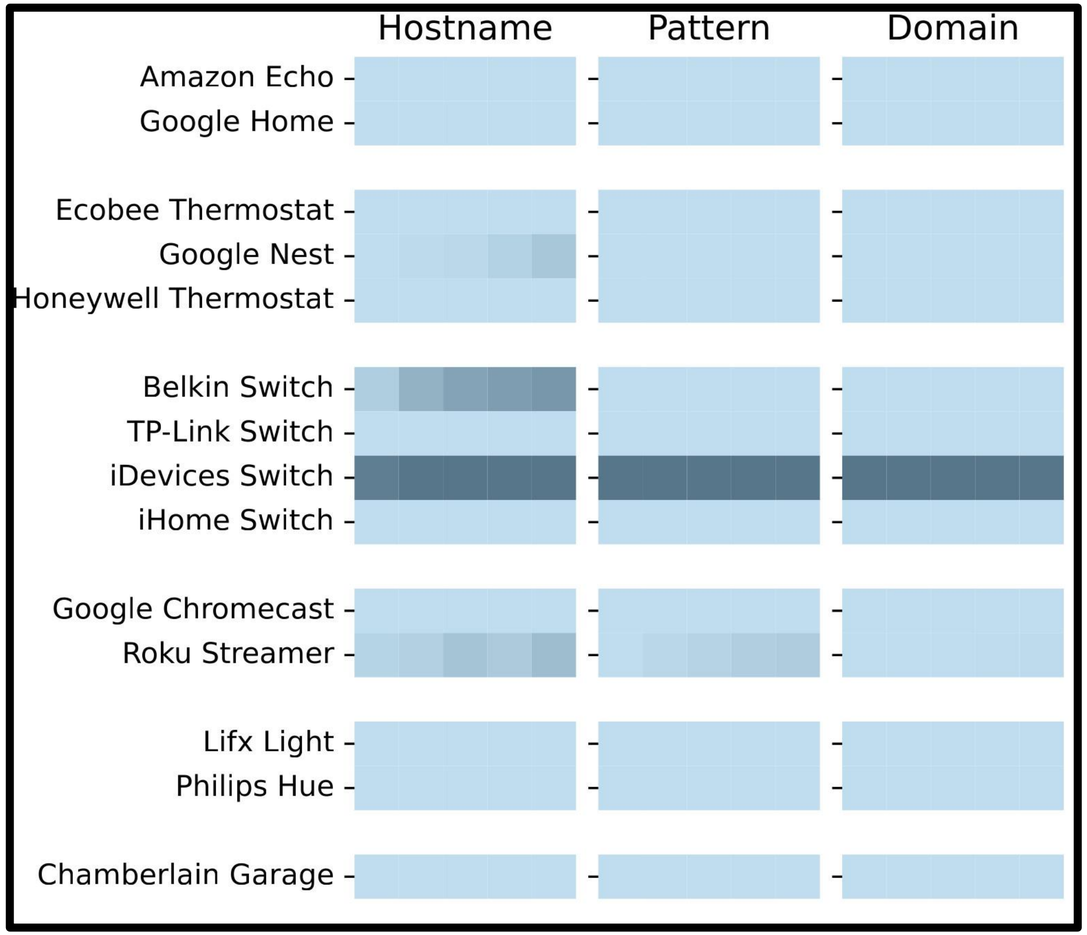
Variability of Behaviors in Crowdsourced Data

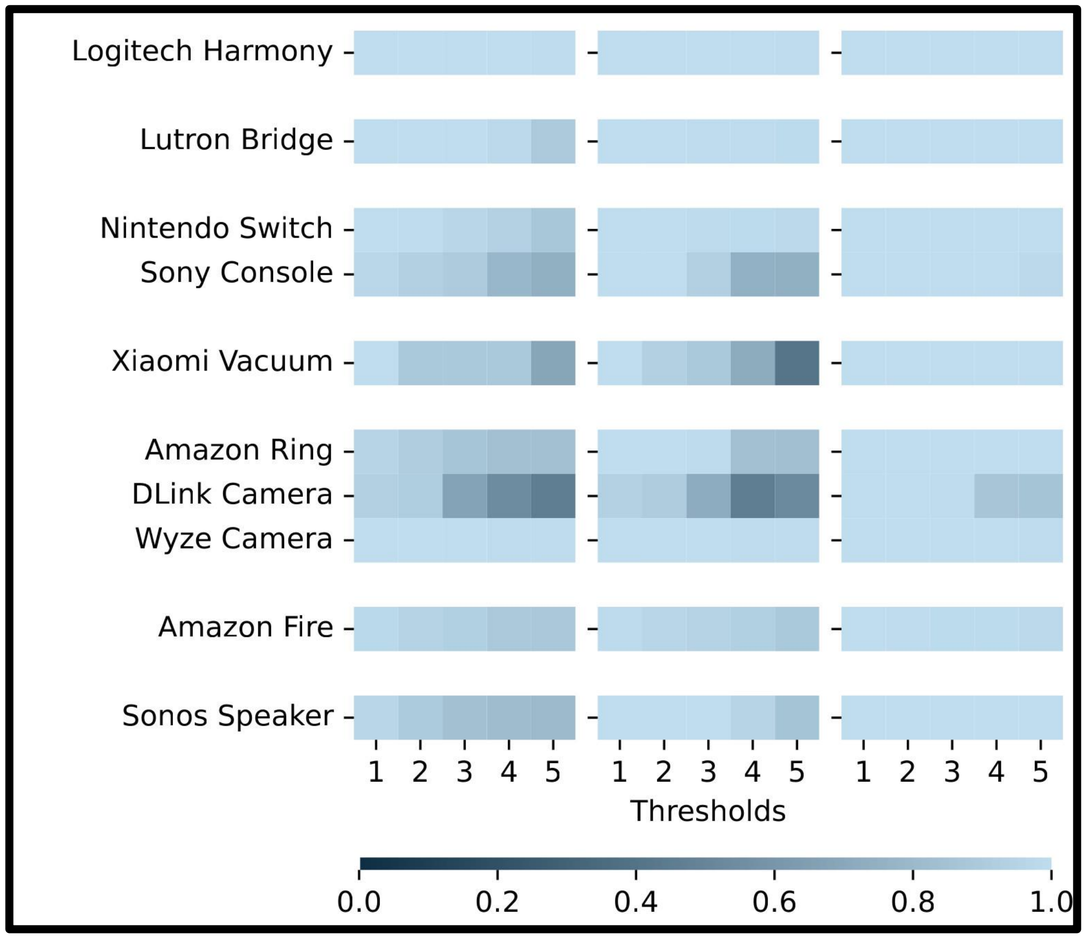


Variability of Behaviors in Crowdsourced Data

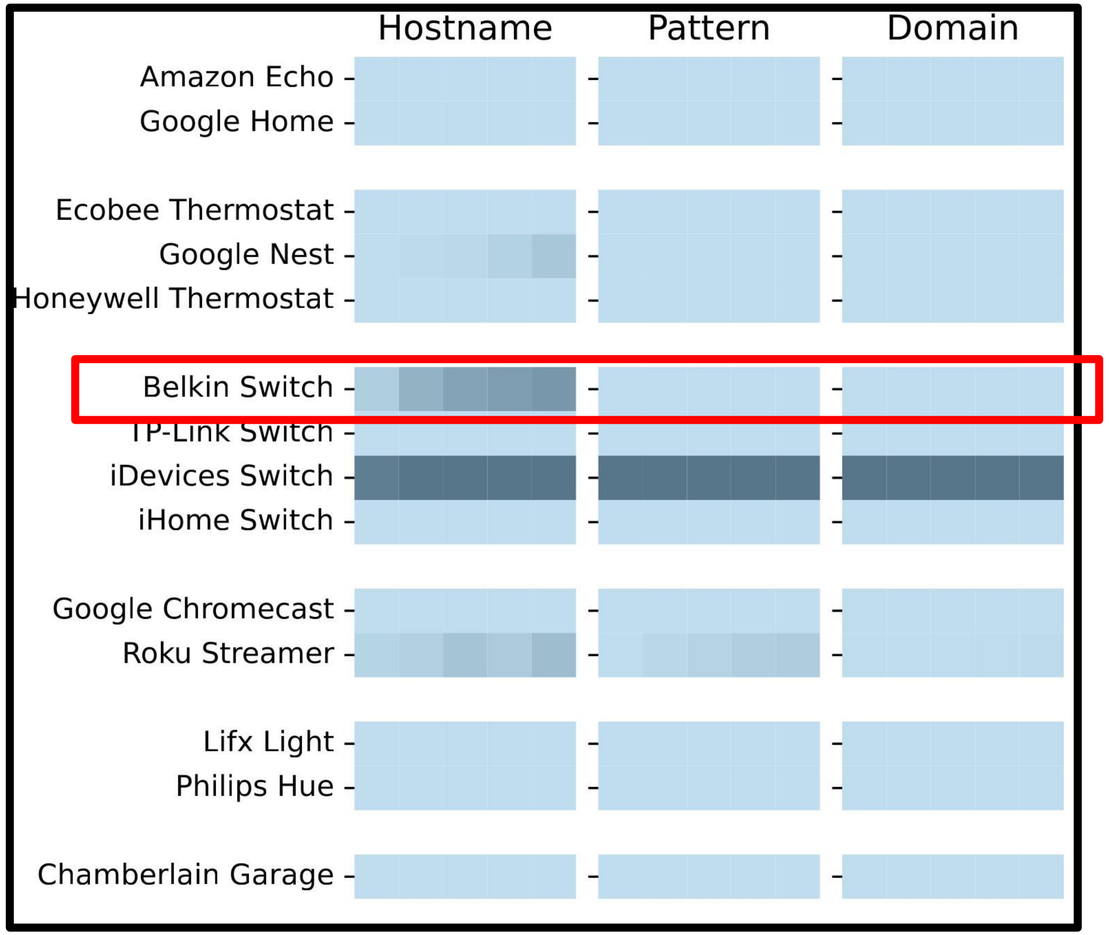


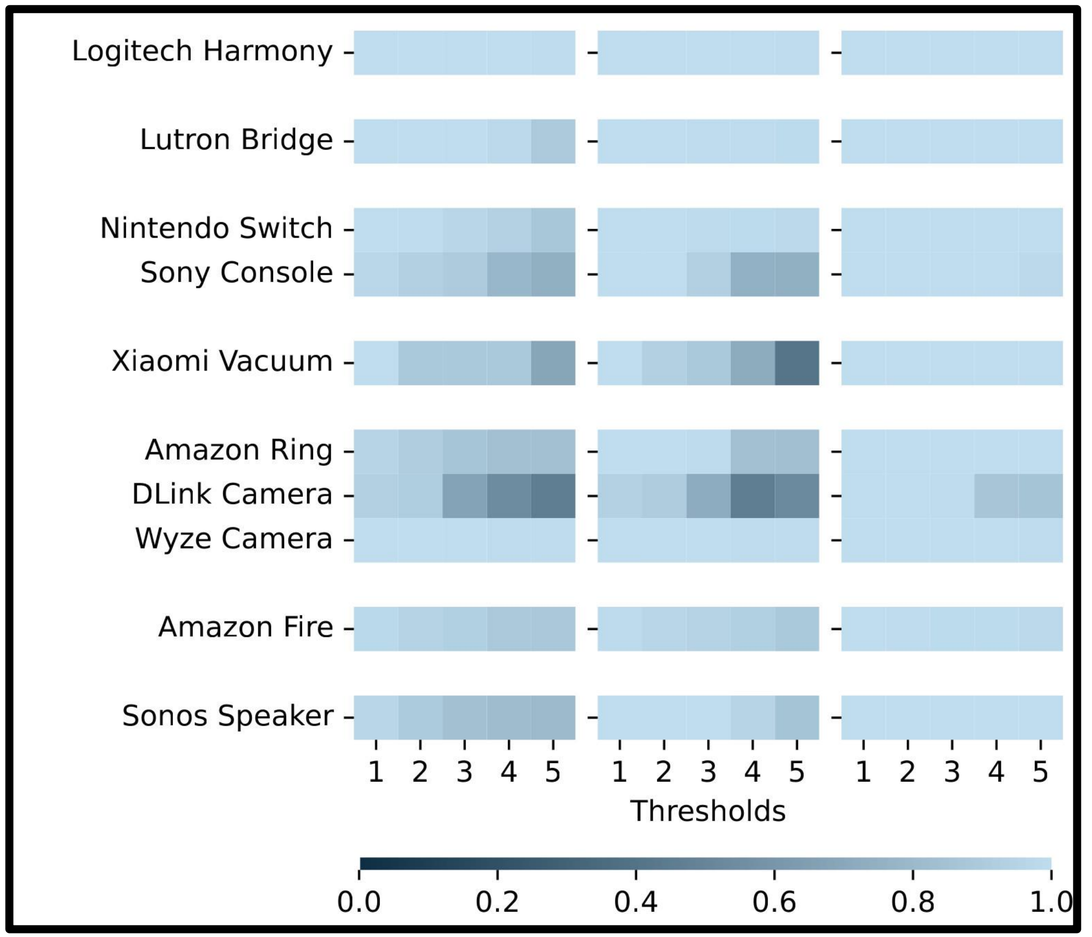
Variability of Behaviors in Crowdsourced Data



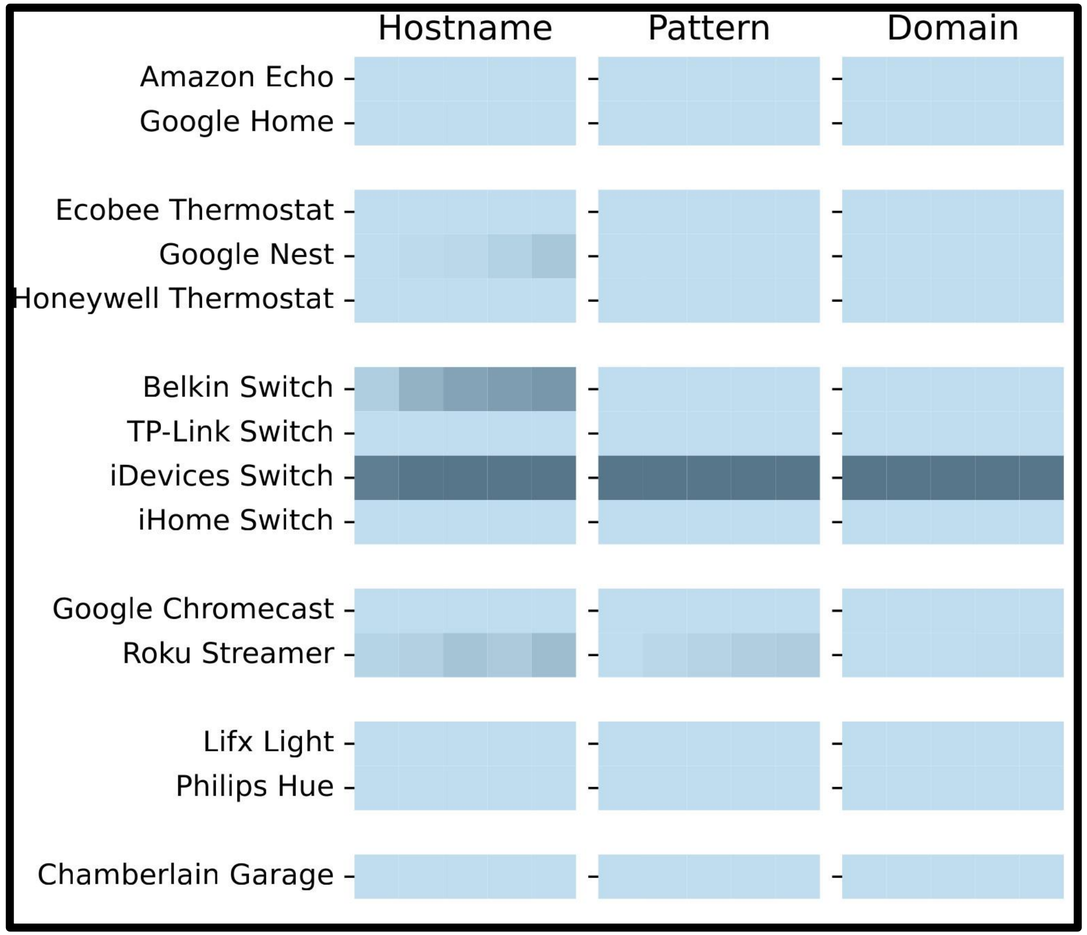


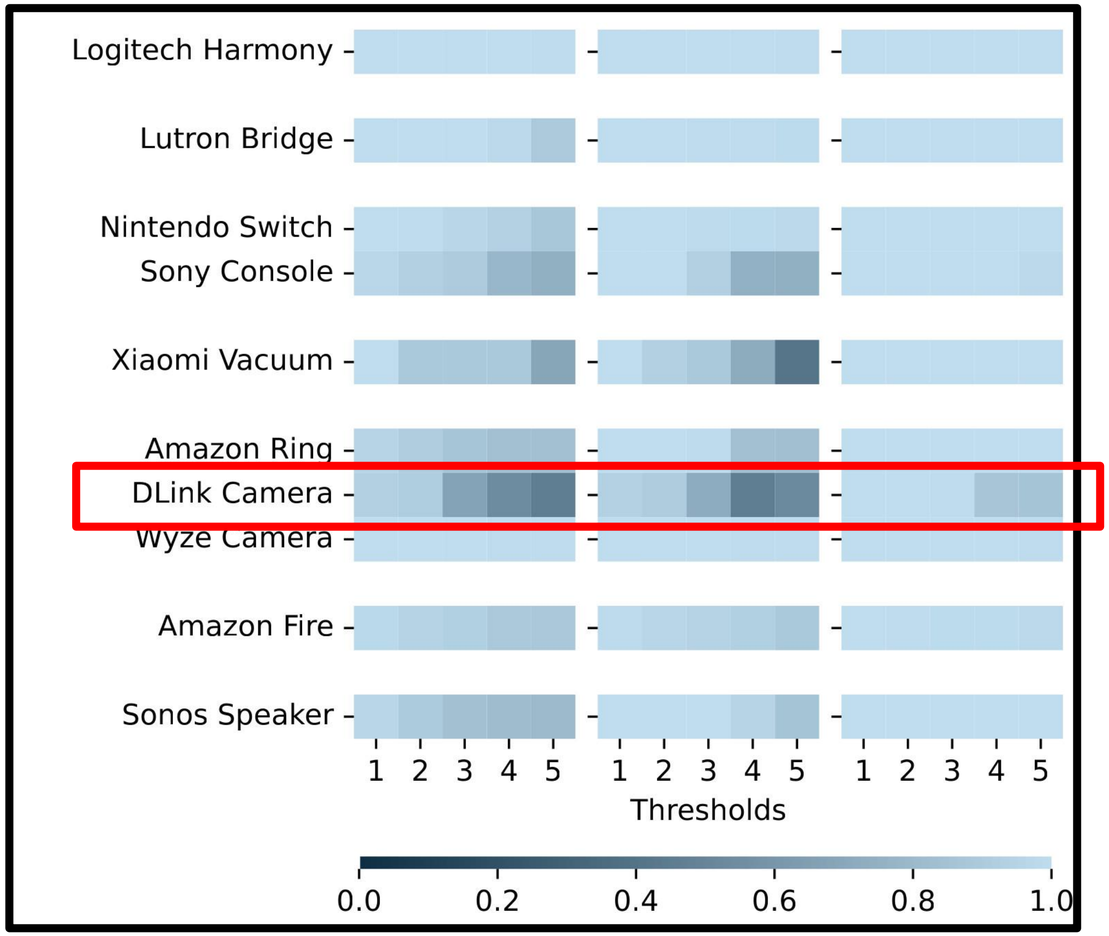
Variability of Behaviors in Crowdsourced Data





Variability of Behaviors in Crowdsourced Data





Measuring the Feasibility of Allowlists

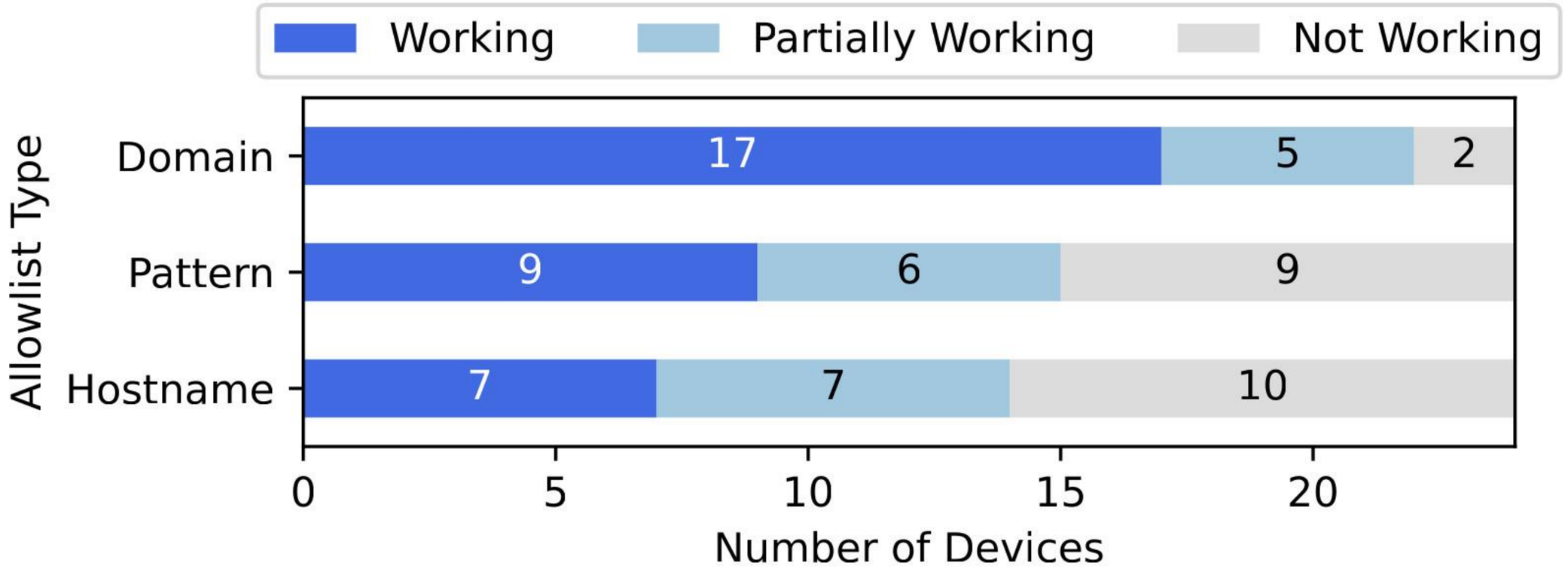
- Preliminary in-lab study of individual devices
- Analyze variability of network traffic in crowdsourced data
- Simulate allowlists based on crowdsourced data
- Verify functionality of crowdsourced allowlists in our lab

Apply Crowdsourced Allowlists in Our Lab

- Using three-year-old IoT Inspector data

Apply Crowdsourced Allowlists in Our Lab

- Using three-year-old IoT Inspector data



Can Allowlists Capture the Variability of Home IoT Device Network Behavior?

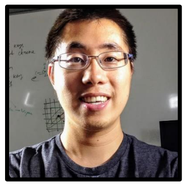
- For some products, yes, years-old crowdsourced data works!
  - Including for likely unseen events (e.g., factory resets)
- Challenges: new endpoints, load balancing, regionalization
- Challenge for streaming devices: varied endpoints
- Security shortcomings: use of cloud services, malicious devices

Can Allowlists Capture the Variability of Home IoT Device Network Behavior?

- For some products, yes, years-old crowdsourced data works!
  - Including for likely unseen events (e.g., factory resets)
- Challenges: new endpoints, load balancing, regionalization
- Challenge for streaming devices: varied endpoints
- Security shortcomings: use of cloud services, malicious devices






Weijia He, Kevin Bryson, Ricardo Calderon, Vijay Prakash, Nick Feamster, Danny Yuxing Huang, Blase Ur

THE UNIVERSITY OF CHICAGO

Security, Usability, & Privacy Education & Research

DARTMOUTH

NYU