Spy-oT

Understanding How Users Learn to Use Internet of Things Devices For Abusive Purposes

[Kieron] **Ivy Turk** They/She

Alice Hutchings She/Her

Department of Computer Science & Technology | Cambridge Cybercrime Centre

CONTENT WARNING

This is a **domestic abuse** talk

Feel free to leave, tune out, put headphones on etc at any time

Your mental health is more important than my talk

Threat modelling

- Existing threat models are designed for technical attacks against technical systems, e.g. to identify threats:
  - CIA: Confidentiality, Integrity, Availability
  - STRIDE: Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, and Escalation of privilege
  - LINDDUN: Privacy-centric threat modelling

  And to prioritise threats:
  - DREAD: Damage, Reproducibility, Exploitability, Affected users, and Discoverability
  - PASTA: Process for Attack Simulation and Threat Analysis

- But what about social attacks against technical systems?
  - Technology facilitated domestic abuse

The Human HARMS Model

Term	Definition	Examples
Harassment	Causing distress through interactions	Sending hateful messages or playing loud sounds
Access/Infiltration	Obtaining or extending access	Increasing own privileges, or adding an external user to a system
Restrictions	Reducing access of existing user	Removing legitimate user's access, or inhibiting specific functionality
Manipulation/Tampering	Controlling other users	Blackmailing users with information from the system, or creating fake evidence
Surveillance	Observing others without their knowledge	Using cameras and microphones to observe users, or investigating logs of past activity

Kieron Ivy Turk, Anna Talas, Alice Hutchings, *"Threat Me Right: A Human HARMS Threat Model for Technical Systems"*, Accepted for Publication at the Security Protocols Workshop 2025, Available at https://arxiv.org/abs/2502.07116

Adversarial modelling

- Adversarial modelling complements threat modelling.
- Define adversaries' goals, capabilities, and limitations.

- Dominant adversarial model for tech-abuse at the time of this paper was the 'UI-bound adversary' [1]
  - Varying technical abilities, but even advanced users are constrained to the UI
  - Not capable of advanced technical attacks.

- But are abusers 'bound' by the UI, or *enabled* by it?
- What types of attacks do highly skilled users attempt?

[1] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. *"A Stalker's Paradise": How intimate partner abusers exploit technology.* CHI '18

Investigating IoT Abuse

- Provide users with devices and work out abusive behaviours live
- Talk-aloud protocol to understand thought process
- Two 2-hour events: 16 + 9 Participants
    - Briefing -> Explore -> Debrief

Classifying IoT Abuse

Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment

Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment

Hey Smart Assistant,
Set an alarm at 3am
That plays Baby Shark
At full volume
For an hour before turning off

Classifying IoT Abuse

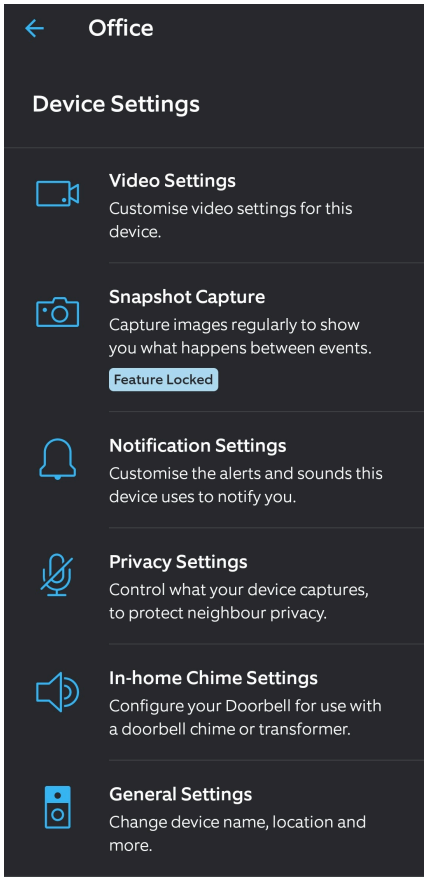
- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access

Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices

Classifying IoT Abuse

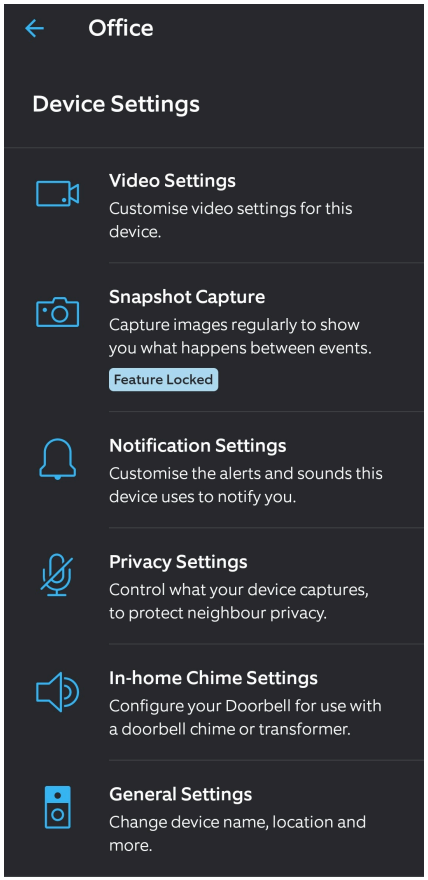
- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices



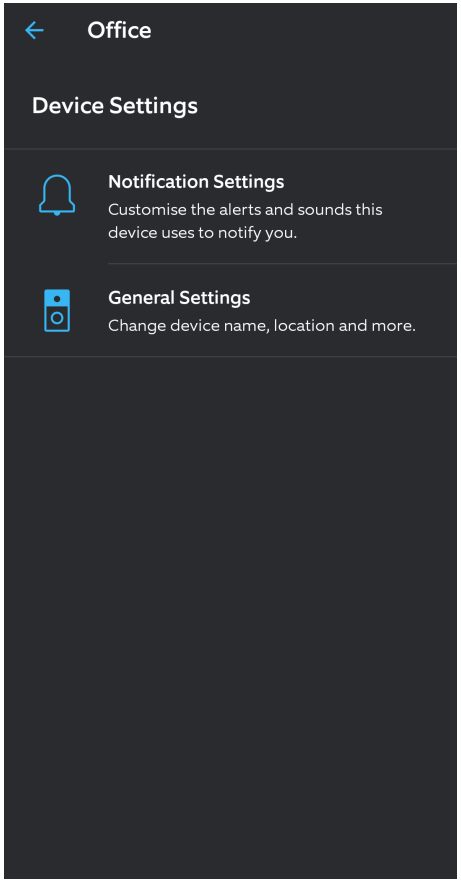
Admin user

Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices



Admin user



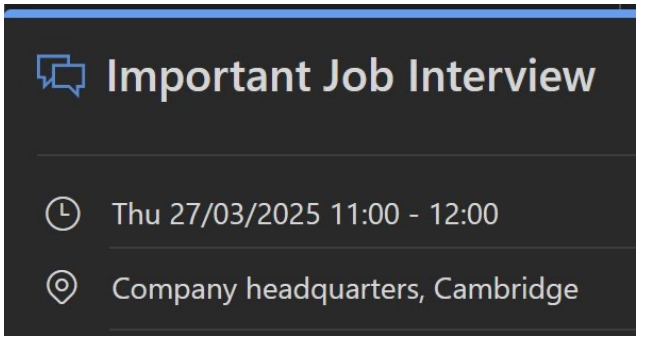
Guest user

Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices
- Manipulation
  - Gaslighting with calendars, logs

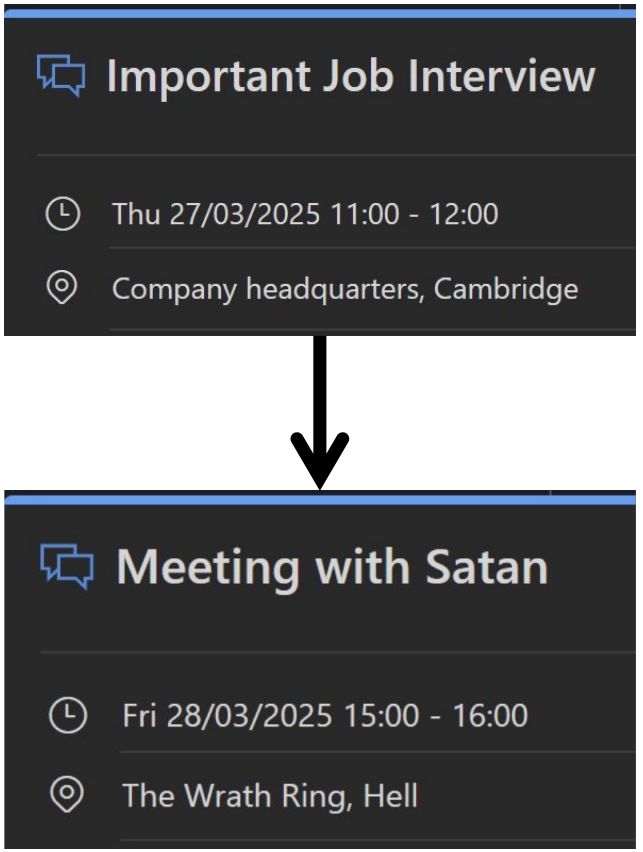
Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices
- Manipulation
  - Gaslighting with calendars, logs



Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices
- Manipulation
  - Gaslighting with calendars, logs



Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices
- Manipulation
  - Gaslighting with calendars
- Surveillance
  - Cameras, Audio, History and logs

Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices
- Manipulation
  - Gaslighting with calendars
- Surveillance
  - Cameras, Audio, History and logs

Classifying IoT Abuse

- Harassment
  - Messaging
  - Interacting with Environment
- Access & Infiltration
  - Share Access
  - Force Access
- Restriction
  - Access control hierarchies
  - Destroy devices
- Manipulation
  - Gaslighting with calendars
- Surveillance
  - Cameras, Audio, History and logs

Miscellaneous “Technical” Attacks

- NFC cloning of smart lock tags
- Upload malware through the charging port
- Build malicious app based on API
- Disable encryption and use WireShark
- Use an EMP pulse generator to disable the device
- Use microphone to work out room architecture

Miscellaneous “Technical” Attacks

- NFC cloning of smart lock tags
- Upload malware through the charging port – No data line
- Build malicious app based on API – No API, cloning and modifying apps requires high skill level
- Disable encryption and use WireShark – Not feasible to disable encryption for arbitrary app
- Use an EMP pulse generator to disable the device – Ran the study IRL not in a movie
- Use microphone to work out room architecture – Need multiple high-quality microphones and lots of time

How Participants Discover Abuse

Discovery Method	As Initial Approach	At Any Stage
Interacting with UI	23	40
Interacting with Physical Device	19	31
Hypothesis-Driven	3	12

How Participants Discover Abuse

Discovery Method	As Initial Approach	At Any Stage
Interacting with UI	23	40
Interacting with Physical Device	19	31
Hypothesis-Driven	3	12

Takeaways:

- [Most] Technical attacks were not achievable
- Possible misuses were largely discovered through interaction, not pre-existing goals
- The provided functionality **enables abuse**

Functionality-Enabled Adversary

- "UI-Bound Adversary" [1]
  - Only able to use interface
  - Technical attacks not possible
  - *Should focus tech-abuse interventions on user interface, not stopping hackers*

- "Functionality-Enabled" Adversary [This paper!]
  - Discovers misuse ideas from provided features
  - Uses and abuses provided features to cause harm
  - Users focus on easy-to-execute attacks
  - *Should focus on possible misuses of provided features and "abusability" of system*

[1] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. *"A Stalker's Paradise": How intimate partner abusers exploit technology.* CHI '18

Conclusions

- Need to understand how abusive actions learned as well as the abuses themselves
- Most users discover misuses through interaction not ideation
- Can model abusers as “functionality-enabled” in addition to “UI-bound”

More info in the paper about:

Codebook of Attacks Discovered

Functionality-Enabled Adversary Example

Attack Feasibility Analysis

Possible IoT interventions and limitations

Contact us at
{kieron.turk, alice.hutchings}@cl.cam.ac.uk

Possible discussion questions

- What role should companies play in identifying opportunities for abuse and creating more abuse-resistant technologies?
- Threat modelling allows us to consider what the potential misuses of technology might be. Is this enough? Do developers and designers have the necessary tools to design out harms?
- How do the adversary models differ across the Spy-oT and Thunderclap papers?
- The papers both challenge hidden assumptions: that peripherals are trustworthy or that legitimate users are benign. How can threat models better surface and question such assumptions?