Transparency in the consumer Internet of Things

DATA FLOWS AND DATA RIGHTS

UNIVERSITY OF CAMBRIDGE

Imperial College London

Report authors
Anna Ida Hudig (lead researcher), University of Cambridge
Chris Norval, University of Cambridge
Jatinder Singh, University of Cambridge

Contributors
Reuben Binns, University of Oxford
Richard Cloete, Harvard University
Hamed Haddadi, Imperial College London
Anna Maria Mandalari, University College London

This report presents independent research funded by a privacy innovation research grant from the Information Commissioner's Office (ICO). The views, opinions and interpretation of the legislation expressed by authors in this publication are those of the authors and do not necessarily reflect those of the Information Commissioner or the ICO. Where the authors have made references to compliance or non-compliance with the legislation for which the Commissioner is regulator, those views are solely those of the authors and do not represent a regulatory finding to that effect by the Commissioner or ICO.

March 2023

Executive Summary

The Internet of Things (IoT) is growing in prominence. Bridging the offline and online worlds, we increasingly see a range of IoT (or so-called "smart") devices marketed at consumers, to assist in various aspects of everyday life. Common examples of consumer IoT products include home security cameras, weighing scales, domestic white goods, streaming media devices, lighting systems, and wearables, to name but a few. These devices typically incorporate sensors, which perceive aspects of the physical environment. Some of these devices are able to influence their surroundings by emitting sound, or by turning lights or heating on or off or up or down, and so on. IoT devices are connected to the wider Internet, operating with and forming part of broader systems and services in which data is exchanged.

The consumer IoT naturally presents data protection considerations. Given that such devices are intended to be used for by people in domestic settings, there is real potential for such devices to capture personal and sensitive information about people's actions, behaviours, preferences, and interactions.

Currently, there is limited visibility over the nature of data processing that occurs in the consumer IoT. Details about the data collected by IoT devices, about how it is being used, and with whom it is being shared, can empower users and support accountability by improving our understanding of the general nature of the consumer IoT and its products, and whether the organisations behind these devices are behaving appropriately.

Research has observed that many IoT products transmit data to a range of locations, though the reasons and rationales for such transmissions can be unclear. Under data protection law, individuals have certain rights. In theory, the *right of access* and the *right to data portability*, which we collectively call *"transparency rights"*, allow users to obtain information about their personal data, how it is being processed and shared, and why. As such, there is an opportunity to use the transparency rights to systemically explore aspects of data processing by IoT product vendors, and to indicate the efficacy of data rights in practice.

Broadly, this research sought to shed more light on the current data practices of vendors in the consumer IoT. Taking a multidisciplinary, tech-legal approach, we focus on the data flows of consumer IoT products, conducting experiments with a selection of 43 IoT products from 11 different product categories. Specifically, we observed the data transmissions of these products as they were being used; exercised data transparency rights in an attempt to gain information directly from vendors about the nature of their data processing; and considered whether it is feasible and appropriate to block (restrict) certain data flows as a way of supporting users control their data.

Our resulting analyses highlight some (1) limitations of current vendor practices in handling data rights, and (2) ongoing challenges and uncertainties regarding the data flows of the consumer IoT. Some key findings include:

- **The overall response rate to the transparency rights requests was inadequate.** Despite a legal obligation to respond to rights requests, many vendors gave irrelevant responses, or no responses. The responses we did get were often incomplete. We faced numerous barriers in the process of exercising rights. Often, at least one follow-up was necessary before a response was forthcoming. Some staff handling the requests appeared to lack relevant knowledge of their data protection obligations. Vendors that had implemented automated systems for exporting personal data made it cumbersome to ask specific questions regarding their data practices. Some requests for identity verification seemed excessive in the given context, and caused delays to the process.
- **The form of rights responses was inconsistent and often lacking.** Responses to data access requests were often generic and prefabricated, and seldom tailored to the specific questions posed as part of the rights requests. Where data was returned, it came in various forms: some included files containing little to no data, or without sufficient information to support interpretation. Other responses consisted of extensive indexes or directories of data, including metadata.
- **Our analysis of responses reveals that IoT product usage entails the collection of a wide variety of information.** This included account registration data, device specifications, sensed data (e.g., heartbeat, motion, video, audio), user actions and events linked with timestamps (usage logs), and any inferences made. We observed that the vendors that did respond would often provide data about the registered account, or device data including device IDs, and sometimes activity logs. Data obtained by

sensors, and details of the inferences that were made, appeared infrequently in vendor responses, and vendors were generally vague about the reasons why they could not or did not return certain types of data.

- **Technical monitoring revealed that many IoT products transmitted data to a range of different destinations.** The data flow analyses and rights responses suggest that cloud and infrastructure providers are regularly among the recipients, though these only represent some of the organisations to which data flow was observed. We found that vendors would typically not disclose details about the identities and the locations of the recipients, nor the specific purposes for sharing the data, despite this information's being explicitly requested when exercising our rights.
- **Some IoT products involved data flowing to destinations across the world.** While most data was sent to IP addresses in the EU, UK, and US, we observed data flowing to addresses in 32 different countries worldwide, including countries with different standards of data protection. As one example, we observed a child smart watch sending data to China, Ireland, Japan, Russia and the UK.
- **Data transmissions often reflected product usage.** While conducting experiments with the products, we observed that certain interactions with the devices resulted in real-time data transmissions to external actors. This strongly suggests that certain aspects of product usage, such as data obtained by sensors, timestamped events or inferences, were being directly communicated to others.
- **Exercising data rights revealed little about the nature of data flows.** While we observed a range of transmissions during product usage, most vendors did not provide information relating to device usage or sensor streams in their responses to rights requests. For those that did respond with such data, we were able to better understand and corroborate the data flows we observed. However, generally, the responses by vendors (if, indeed, they did respond) shed little light on the nature and destinations of those data flows.
- **Vendors often do not clearly specify the periods for which they will retain personal data.** Typically, data was said to be retained until a user requests erasure, terminates their account, or when the data is no longer required (as described by the vendor's purposes). Vendors were generally vague as to the specific reasons for data processing, and this vagueness may be perceived as allowing prolonged periods of retention.
- **Users preventing (blocking) data flows to certain destinations has its limits.** It has been suggested that mechanisms that enable the restriction of data flows can empower users by giving them control. While this is a developing area that is currently being researched, we found that there are questions – both technical and non-technical – about the appropriateness and efficacy of such an approach.

Our findings reveal that consumer IoT products can involve substantial flows of data. However, the generally poor data rights practices of vendors hinder detailed understanding of what data is being shared, with whom, and why.

We therefore emphasise the need for interventions that make data rights more effective, along with other practical methods for improving transparency over data processing practices within the consumer IoT – be it through more education and support for vendors, through the development of best practices, or through the provision of means for better oversight and stronger enforcement. We argue that this is an area requiring urgent attention, as IoT products become increasingly common in our domestic lives.

Table of contents

Glossary 07

1 Introduction 08
- 1.1 Research aim 08
- 1.2 Scope 09
- 1.3 Report outline 09

2 Background 10
- 2.1 An overview of the consumer Internet of Things 10
  - 2.1.1 Key definitions in the IoT 10
  - 2.1.2 Key stakeholders in the IoT 11
- 2.2 Data protection law and data rights in the IoT 12
  - 2.2.1 A brief overview of data protection law 12
  - 2.2.2 The right of access and the right to data portability 15
  - 2.2.3 Data subject rights in practice 16
- 2.3 Exacerbating data risks 17
  - 2.3.1 Expanded scope for data processing 17
  - 2.3.2 Passive data collection 17
  - 2.3.3 Complexity 17
  - 2.3.4 Risks and data protection principles 17

3 Methodology 18
- 3.1 Selecting and obtaining IoT products 18
  - 3.1.1 Selecting IoT product types 18
  - 3.1.2 Selecting IoT products 19
- 3.2 Interactions with the IoT products 20
  - 3.2.1 Phases of experiments 20
  - 3.2.2 Recording experimental specifics 21
- 3.3 Research approach 22
  - 3.3.1 Data flow monitoring and analysis 22
  - 3.3.2 Data rights requests and analysis of responses 24
  - 3.3.3 Analysis of legal documents 25
  - 3.3.4 Blocking data flows 26
- 3.4 Limitations 26
  - 3.4.1 Device selection and purchase 26
  - 3.4.2 Data flow monitoring 26
  - 3.4.3 Device interactions 26
  - 3.4.4 Data rights requests 26
- Section 3: Summary 27

4 Vendor implementation of data transparency rights 28

4.1 Overview of response rates to data access and portability requests 28
4.1.1 Several vendors did not address any aspects of our data access or portability rights requests 28
4.1.2 Generic and specific follow-up emails were sometimes effective in eliciting responses 29

4.2 The process of interacting with vendors 29
4.2.1 Automated data rights processes can help or hinder 29
4.2.2 Employees handling requests may lack relevant knowledge and skills 30
4.2.3 Vendors sometimes requested more information 30
4.2.4 Other issues in the process of communicating with vendors 30

4.3 Form and content of the responses to data transparency rights requests 33
4.3.1 Data access: responses were generic and sometimes prefabricated 33
4.3.2 Data portability: copies of personal data were generally provided in mixed formats 33

Section 4: Summary 34

5 Data processing observations 35

5.1 Data types 35
5.1.1 Account registration data 35
5.1.2 Device details 36
5.1.3 Information directly perceived by sensors (sensor data) 36
5.1.4 Usage logs 37
5.1.5 Inferences 37

5.2 Data transmissions 37
5.2.1 IP addresses 37
5.2.2 Countries 40
5.2.3 Organisations 41

5.3 Patterns in data flows 45
5.3.1 Types of monitored flows 45
5.3.2 Transmissions and device usage 47
5.3.3 Responses and data flows 48

5.4 Data storage practices 57
5.4.1 Privacy policies 57
5.4.2 Vendor responses 57

Section 5: Summary 59

6 Blocking IoT data flows 60

6.1 Background on blocking 60

6.2 Blocking case studies 60
6.2.1 Case study 1: the light bulb 60
6.2.2 Case study 2: the baby monitor 61

6.3 Socio-technical considerations 61
6.3.1 Technical considerations 61
6.3.2 Legal considerations 61
6.3.3 Social considerations 62

Section 6: Summary 62

7 Conclusion and future work 63

7.1 Future work 64

Glossary

Term	Definition
Bluetooth-only devices	IoT devices that can only communicate via Bluetooth. Bluetooth is a short-range radio technology, meaning such devices tend only to interact directly with the companion app (on the mobile device), and depend on the companion app to communicate with others via the Internet.
Companion app	An application that runs on a mobile device which allows users to interact with an IoT product. The app may provide information about the device and its functions; details about data obtained by the device, such as recordings made by a video doorbell; and so on. A companion app is often, but not necessarily, issued by the manufacturer of the IoT product.
Data controller	The natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.
Data flow	See "network traffic".
Data processing	Any operation or set of operations, performed on personal data or on sets of personal data, by automated or manual means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction.
Data processor	A natural or legal person, public authority, agency, or other body which processes personal data on behalf of the controller.
Data subject	The identified or identifiable living individual to whom personal data relates.
Data transparency rights (or "data rights")	Refers to the *"right of data access"* (UK GDPR, Art 15) and the "*right to data portability"* (UK GDPR, Art 20).
Destination	The end point of a data transmission or communication channel.
Encryption	When communications are encrypted, this generally means that only the recipient and the sender can 'see' what is transmitted.
Inferences	Inferences are attributes, opinions, or characteristics of users inferred from the data. An example of an inference is a user profile.
Internet-enabled devices	IoT devices that can directly connect to and communicate via the Internet. Such communications can be with external actors, as well as the companion app.
IoT device	Refers to the physical IoT device, not including the companion app. This is usually the physical object with which one interacts.
IoT product	Product refers to a an IoT product *as a whole*, i.e., the IoT device and its companion app combined.
IoT product type	A category of IoT products, characterised by similarities in functionality and usage. Examples include fitness trackers, security cameras, smart lightbulbs, and so on.
IP address	An IP ("Internet Protocol") address serves as an identifier for a device on a network. This is used to establish connections through which communication takes place.
Network traffic ("traffic")	The data travelling over the Internet, from one destination to another, is collectively called *network traffic*. Network traffic involves what we refer to as "data flows" or "data transmission".
Personal data	Any information relating to an identified or identifiable natural person ("data subject").
Response	Response of a vendor to a data access and/or data portability request.
Rights request (or "request")	Data access or data portability request (also: "data transparency request").
Recipient	Any individual or other actor to whom data is disclosed. This could be a user, vendor, or any third party.
User (of IoT)	Any individual using or otherwise operating an IoT product. Usually, a user is also a data subject.
Vendor	The organisation that put the product on the market, and often is visibly associated with the product (e.g., via branding).

1. Introduction

The Internet of Things, or *"IoT",* comprises physical devices that are connected to a network, working to "bring the physical world online". These devices incorporate sensors that can perceive information from the physical environment, and/or actuators that can effect changes to the physical environment, such as controlling or switching off the light, or closing the curtains. As devices are network-enabled, they can interact with a range of online services, software applications, other devices, etc. This involves the devices sending or receiving data – including data from sensors, results of computations, summaries, inferences, commands to take particular actions, and so on – in order to deliver particular functionality.

The term *"Internet of Things"* reflects the vast range of physical devices that could be brought online, connected via the Internet. Though the IoT has a range of applications, our focus here is on consumer-oriented IoT products marketed as "smart" or "intelligent" for household or individual use.

The market for smart devices for consumers is growing rapidly, projected to double in size between 2022 and 2027.[1] A 2021 survey showed that 28% of consumers in the US have at least one smart device, with 11% having more than three IoT devices.[2] The *"consumer IoT"*, the focus of this report, broadly includes products such as smart lighting, home security (camera) systems, fitness trackers and other wearables, thermostats, kitchen appliances and other white goods, smart vacuum cleaners, and so on.

What these products have in common is that they are marketed to consumers. While IoT products promise efficiency and convenience, their integration into everyday life, often in highly personal settings, also raises concerns. Because consumer IoT products operate in domestic environments, such as within our homes and on our bodies, there is real potential for personal, intimate, and sensitive information about ourselves and our lives to be collected and used at scale.

Currently, however, there is uncertainty and opacity regarding the data processing that occurs in the consumer IoT. To better understand and assess the potential issues, risks, and challenges regarding data in the IoT, we require information on the nature and behaviour of IoT products, and how they and their vendors process data. Not only is better understanding important for accountability when things go wrong, it is also crucial for interventions that work towards improved protection of the rights and interests of the individuals using these products.

1.1 Research aim

Our research focuses on issues of transparency in the consumer IoT. Specifically, we focus on the data flows associated with IoT products, as a means for uncovering, and providing insight into, the data processing undertakings of consumer IoT vendors.

For this purpose, we experiment with a selection of 43 IoT products, across 11 different product types, employing primarily two research methods. The first method entails technical mechanisms for tracking, capturing, analysing, and blocking the data transmissions of the IoT products. The second method concerns the exercise of rights under data protection law, specifically the *"right of access"* (UK GDPR Art 15) and the *"right to data portability"* (UK GDPR Art 20), collectively referred to in this report as data *"transparency rights"*. These transparency rights allow individuals to request information about a product vendor's personal data processing practices, and obtain copies of the personal data the vendor holds. We use these methods and mechanisms in combination, taking a multidisciplinary (tech-legal) approach that attempts to build better understanding of the data flows in the consumer IoT.

We report our findings from three areas:

1. We describe our **experiences in exercising our data rights with IoT product vendors,** noting the ways vendors responded, as well as the barriers we encountered. This provides insight into the current state of data rights practices in the consumer IoT (Section 4).
2. We use **data flow monitoring in conjunction**

---

[1] Statista (2022, December). *Statista Market Forecast: Smart Home.* https://www.statista.com/outlook/dmo/smart-home/united-kingdom#revenue

[2] Westcott, K., Loucks, J., Littmann, D., Wilson, P., Srivastava, S., & Ciampa, D. (2021). *Build it and they will embrace it: Consumers are preparing for 5G connectivity in the home or on the go.* The Deloitte Centre for Technology, Media & Telecommunications.

with data transparency rights to uncover certain data practices, and consider how vendors' rights responses align with and relate to the data flows that we observed (Section 5).

3. We consider the **potential and limitations for *"blocking"*** certain communications (preventing particular data flows) from IoT products (Section 6).

1.2 Scope

Our work aims to assist in building understandings of certain data-related practices in the consumer IoT, in order to support stakeholders in finding constructive ways forward.

It is important to note that our work is exploratory. Given that it is based on observations, which themselves are necessarily based on a range of assumptions, constraints, and uncertainties (see **§3.4**), this report should only be interpreted as indicative of some of the potential issues and trends. That is, the findings presented should not be considered absolute, complete, definitive, nor fully representative of the nature of the vast and complex consumer IoT ecosystem. Moreover, in recognising that technology is fast-moving, our report reflects a snapshot at a certain moment in time (2021-2022).

Further, our work is not intended as a consumer product testing study that involves direct comparisons and assessments of particular products. Our approach was designed to uncover general data-related practices in the consumer IoT, rather than supporting one-to-one comparisons between products or vendors in an environment of many variables. Further, our work cannot be considered or construed as making assessments of legal compliance or value judgements, as advising on which products to purchase or not purchase, nor as challenging particular vendors.

1.3 Report outline

- **Section 2** provides background on the IoT and its risks, data protection law, and data transparency rights.
- **Section 3** explains the methodology and methods used in this study.
- **Section 4** describes our findings and experiences regarding vendors' implementations of data transparency rights.
- **Section 5** presents our data flow analyses of the results from product traffic monitoring and transparency rights.
- **Section 6** discusses the potential regarding the blocking of data flows.
- **Section 7** concludes the report and discusses areas for attention.

2 Background

We begin by providing some background on the key concepts of the consumer IoT (**§2.1**), as well as presenting a brief overview of data protection law with a key focus on data rights (**§2.2**). We then highlight some particular data risks inherent in the consumer IoT and their relationship with data protection principles (**§2.3**).

2.1 An overview of the consumer Internet of Things

2.1.1 Key definitions in the IoT

This research focuses on the consumer *"Internet of Things",* referred to as the ***"IoT".*** The IoT consists of physical devices that contain sensors and/or actuators. *"Sensors"* can detect or measure properties from the physical environment. Examples of sensors common in consumer IoT products are those that detect motion, pressure, temperature, images (cameras), or sound (microphones). An *"actuator"* is a component that can take a certain physical action or make a change in the environment, such as to move something, turn something on or off (lighting, heating), and so on. With these embedded sensors and actuators, coupled with their network connectivity, IoT devices can interact with their surroundings as part of a broader ecosystem. In this way, the IoT can be seen as a physical extension of the Internet. The *"consumer IoT"* refers to IoT products that are marketed to consumers or individuals, aimed for domestic or personal use.

IoT devices are characterised by their ability to communicate, or exchange data, with others (directly or indirectly) over the Internet. Most consumer IoT products include a *"companion app":* an application that runs on a mobile device (i.e., a phone or tablet that is Internet-connected) that allows individuals to use the device by controlling it remotely or by receiving data or analyses relating to the device; or to configure the device and change settings. Companion apps may provide (potentially real-time) feedback and insights into the functioning and operation of the device or patterns in the lives of users; or data obtained by the device, such as images of a visitor via a video doorbell. A companion app is typically provided by a product's vendor, though in some cases the companion app is the product of another organisation.[3] Although a mobile app in isolation would not generally be considered as part of the IoT, in this study, the companion app is, when linked to the IoT device, considered an integrated part of an *IoT product*. This is illustrated in Figure 1.

We use the following terminology throughout this report:

- **IoT device *("device"):*** the *"IoT device"* is the physical IoT device, often equipped with some sensors (i.e., a security camera or wearable) and/or actuators.
- **Companion app *("app"):*** the *"companion app"* (or *"app"*) is a mobile application that can be used to control the IoT device and retrieve and visualise information from the device via a wireless connection such as Internet or Bluetooth.
- **IoT product *("product"):*** the *"IoT product"* is the combination of the IoT device and the companion app associated with the device (if present).

---

[3] Moreover, some devices can work with more general apps or other infrastructures that can facilitate interoperability across a range of devices.

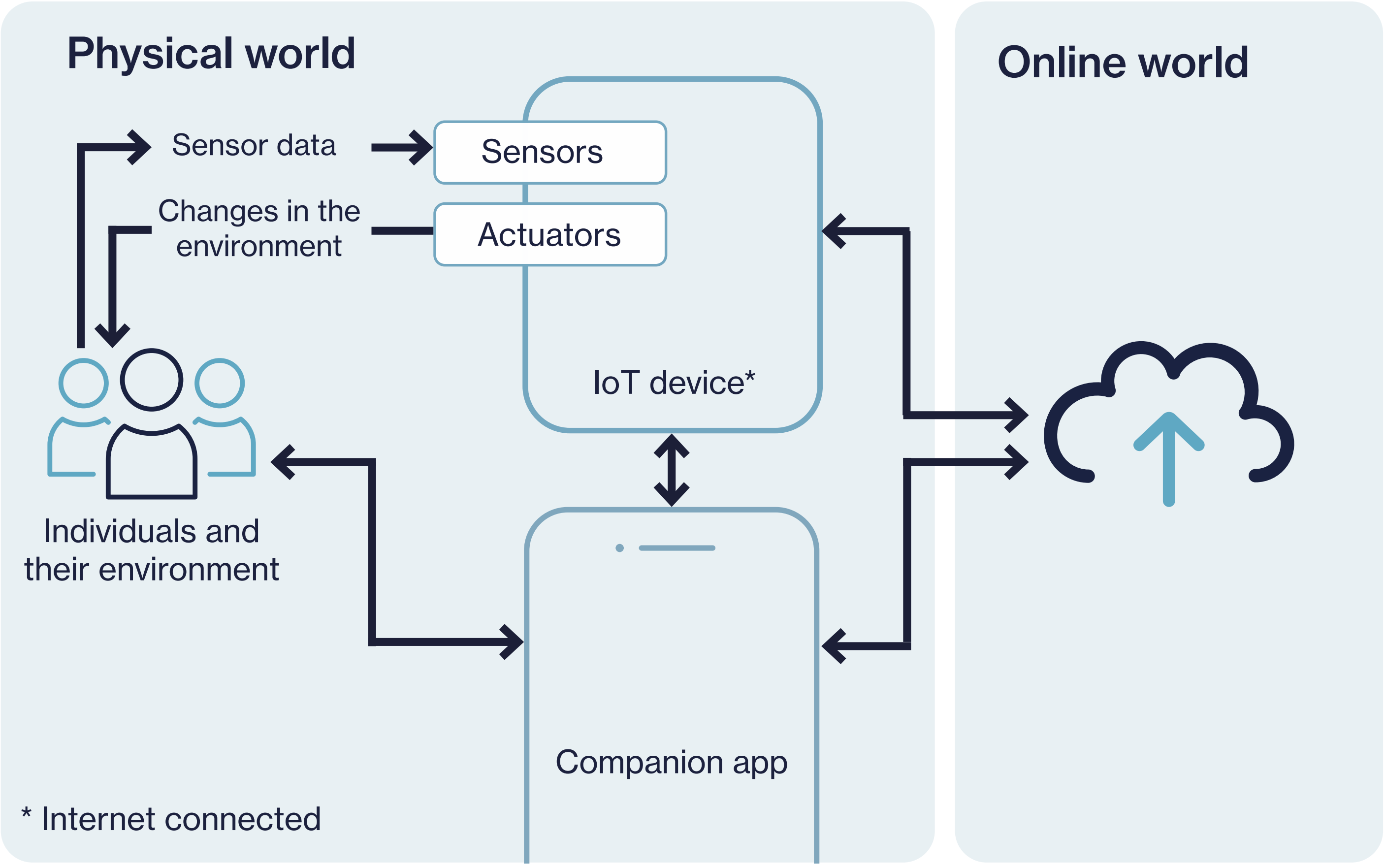


Figure 1: Illustration of the possible interactions between IoT devices, companion apps, and the Internet

2.1.2 Key stakeholders in the IoT

The IoT entails a multi-stakeholder ecosystem, with many different actors. This research focuses primarily on the following:

- **Individuals**: when we speak of *"individuals",* we refer to those who are formally called *"data subjects"* in the context of data protection law: th identified or identifiable living individual to whom personal data relates.[4] The data subject can be the *user* of the device, which is the individual that directly operates or interacts with the device. However, the data subject may also be an individual who is not the primary user of the product, but whose data is captured by the device; e.g., a delivery person captured by a security camera.
- **Vendors**: in this report we use *"vendor"* to refer to the organisation that puts the product on the market, and often is visibly associated with the product (e.g., via branding). Even though vendors are sometimes referred to as "manufacturers", we choose "vendor" throughout this report, given that products from the same manufacturer might be branded, sold, and marketed by different vendors.
- **Cloud providers**: a *"cloud provider"* is a company that provides computing resources as a service, for example regarding storage, computation, content distribution, analytics, etc. Cloud services play an important role in underpinning the vast majority of online services, the dominant players being Amazon (AWS), Microsoft (Azure), and Google (Cloud).[5]
- **Advertisers**: We use the term *"advertisers"* to refer to organisations that participate within a broader advertising, or "AdTech", ecosystem; particularly those that use personal data, or inferences drawn from the data, to target individuals with advertisements or support such processes.

Note that there are overlaps between the categories of stakeholder just described; for instance, some vendors also provide cloud services and sell analytics and advertising products.

---

[4] Art 4, UK GDPR

[5] Cobbe, J., Norval, C., & Singh, J. (2020). What lies beneath: Transparency in online service supply chains. *Journal of Cyber Policy, 5*(1), 65–93. https://doi.org/10.1080/23738871.2020.1745860

2.2 Data protection law and data rights in the IoT

Given that the consumer IoT typically entails the processing of personal data, data protection law is highly relevant. We now provide an overview of some aspects of UK data protection law, data transparency rights, and the applicability of data transparency rights in the consumer IoT. For more details and specifics of data protection in the UK, see the online *Guide to Data Protection* by the Information Commissioner's Office (ICO).[6]

2.2.1 A brief overview of data protection law

Here we provide a very brief overview of some key aspects of the UK General Data Protection Regulation (UK GDPR), the UK's Data Protection Act 2018 (DPA 2018), and the data protection principles and rights as relevant for this report. Additional guidance is provided by data protection authorities in individual EU member states or the European Data Protection Board (EDPB). The UK supervisory (data protection) authority is the Information Commissioner's Office (ICO). The functions and competences of the ICO are set out in the DPA 2018.[7]

The UK GDPR and the Data Protection Act 2018
In 2018, the EU General Data Protection Regulation (*"EU GDPR"*)[8] came into effect, laying down rules to protect the "fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data". The GDPR was retained in domestic law after the UK left the EU.[9] In January 2021 the GDPR applying in the UK (the *"UK GDPR"*) was amended by Regulations, including to extend the UK GDPR to cover activities which were outside of the scope of the EU GDPR (such as national security and defence).[10] The UK GDPR is largely based on the EU GDPR, containing the same key principles, rights and obligations, but allows the UK the independence to keep the framework under review.[11]
The UK GDPR is complemented by the DPA 2018.[12] This clarifies, modifies, qualifies, extends, and restricts various aspects of the UK GDPR[13]; incorporates the EU Law Enforcement Directive into UK law to cover law enforcement processing[14]; and establishes a data protection regime for personal data processing by the intelligence services.[15] When we refer to *"data protection law"* or *"UK GDPR",* we use this as shorthand to refer to the GDPR as retained and amended in UK law.

Material and territorial scope of data protection law
Data protection law regulates the processing of personal data. *"Personal data"* is "any information related to an identified or identifiable individual ("*data subject*").[16] *"Processing"* is defined broadly, not restricted to computation, but rather "any operation or set of operations performed on the personal data", such as collection, recording, sharing, storing, disclosure, combining and erasure.[17] Generally speaking, an entity having anything to do with personal data will be considered to be processing that data.

The UK GDPR has wide territorial scope. The law does not only apply to organisations based in the UK or EU, but much more broadly to the processing of data of UK citizens or residents, or to the processing of personal data by organisations that offer goods or services to UK citizens or residents.

Controllers and processors
Data protection law describes two key roles: the data controller and the data processor. The *"data controller"* is "the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data".[18] A data controller is an organisation or individual that makes the decisions about processing activities, such how the data will be collected, stored and shared.

A *"data processor*" is an actor which processes personal data on behalf of the controller.[19] In the context of this project, the IoT product vendor is typically the data controller, as they have the "control" over the purposes and means of data processing relating to the IoT product (in that they determine, through the ways they have designed their product and organisational practices, the associated data flows, computation and other processing). The vendor may assign certain data processing to other actors, such as cloud platforms that might provide some infrastructure services, where such providers are typically considered data processors.

---

[6] Information Commissioner's Office. *Guide to Data Protection.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/
[7] Information Commissioner's Office. *Guide to Data Protection: About the DPA 2018. I*CO. https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-dpa-2018/about-the-dpa-2018/
[8] *General Data Protection Regulation (GDPR):* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.
[9] S. 3, European Union (Withdrawal) Act 2018.
[10] The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019.
[11] Information Commissioner's Office. *The UK GDPR.* ICO. https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/data-protection-and-the-eu-in-detail/the-uk-gdpr/
[12] Information Commissioner's Office. *Guide to Data Protection: About the DPA 2018.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-dpa-2018/about-the-dpa-2018/
[13] Pt 2, Data Protection Act 2018
[14] Pt 3, Data Protection Act 2018
[15] Pt 4, Data Protection Act 2018
[16] Art 4(1), UK GDPR
[17] Art 4(2), UK GDPR
[18] Art 4(7), UK GDPR
[19] Art 4(8), UK GDPR

Lawful basis

For the processing of personal data, a lawful basis is required. Data protection law provides several lawful bases. These include bases where processing is *necessary* for certain *permitted purposes*, or those where it is not necessary for such a purpose but where *consent* to processing has instead been obtained from the data subject. Permitted purposes for processing include (a) to perform a contract to which the data subject is party, (b) to meet a legal obligation imposed on the controller, (c) to protect the vital interests of an individual, (d) to fulfil a public interest task, or (e) for the legitimate interests of the controller where they are not overridden by the data subject's rights and interests.

For the processing of special category data, i.e., that data considered particularly sensitive such as that concerning health, there are more restrictive bases for processing. Data Protection Impact Assessments (DPIA) should be conducted for data and processing that is likely to pose high risk to people's rights and freedoms.

Perhaps the most common and relevant lawful bases for the processing of personal data in the consumer IoT are consent and legitimate interest. *"Consent"* must be freely given, specific, informed and unambiguous, and for those under 13 years of age, requires parental permission. Though consent is common and encouraged, it is not always required; *"legitimate interest"* is a lawful basis which can, under certain circumstances, allow processing for the interests of controllers (here, vendors) or others without consent. However, this requires that controllers can demonstrate that the legitimate interest exists, that the processing is *"necessary"* to achieve this (i.e., there is no other, less intrusive way for achieving the same result), and that they balance the processing interests against the rights, interests and freedoms of the individual (data subject).[20]

Data protection principles

Data protection law presents a list of data protection principles that should be guiding the processing activities of a company.[21] These principles are set out in Table 1. This report will, where relevant, refer to some of these principles to indicate how our findings may relate to the data protection law.

---

[20] Information Commissioner's Office. *Guide to Data Protection: Right to data portability.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/

[21] Art 5, UK GDPR

Principle	Explanation
Lawfulness, fairness and transparency	Personal data must be processed lawfully, fairly, and in a transparent manner in relation to individuals.
Purpose limitation	Personal data must be collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes.[22]
Data minimisation	Personal data must be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.
Accuracy	Personal data must be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that is inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.
Storage limitation	Personal data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data is processed.[23]
Integrity and confidentiality	Personal data must be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organisational measures.
Accountability	The controller shall be responsible for, and be able to demonstrate compliance with, all of the principles above.

Table 1: Overview of data processing principles as defined in Article 5 of the UK GDPR[24]

Technical and organisational measures

Data protection law requires that data controllers and data processors implement appropriate technical and organisational measures to demonstrate that they comply with data protection law. In addition, they should maintain relevant documentation on processing activities. Measures to meet the requirements in the UK GDPR of data protection by design and default include data minimisation, anonymisation, pseudonymisation, transparency, allowing the user to monitor processing, and creating and improving security.

Data subject rights

All individuals within the scope of the data protection law enjoy certain data subject rights. These rights are briefly explained in Table 2.

---

[22] Art 5(b), UK GDPR further states that “further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes”.

[23] Art 5(e), UK GDPR further states that “personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of individuals”.

[24] Information Commissioner’s Office. *Guide to Data Protection: The Principles. ICO.* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/#the_principles

Data Subject Right (UK GDPR)	Explanation
Right to be informed (Arts 13-14)	Individuals have the right to be informed about the collection and use of their personal data. This involves the provision of a notice that explains the processing of personal data, including disclosure to third parties and retention periods. Such notice should provide contact details of the data controller as well as information on the data rights of individuals.
Right of access (Art 15)	Individuals have the right to obtain a copy of their personal data, as well as other supplementary information related to data processing. It helps individuals to understand how and why the organisation is using their data, and check if the organisation is operating lawfully.
Right to rectification (Art 16)	Individuals have the right for inaccurate personal data about them to be corrected, or to have incomplete personal data completed.
Right to erasure (Art 17)	Individuals have the right for data to be erased if there is no compelling reason for retention. This right is also referred to as "the right to be forgotten".
Right to restrict processing (Art 18)	Individuals have the right to the (temporary) restriction of processing under certain circumstances. These circumstances apply when the individual believes that the information is inaccurate, the processing is unlawful, or other specific reasons.
Right to data portability (Art 20)	Individuals have the right to obtain a copy of their personal data in a machine-readable format and to reuse this for other services.
Right to object (Art 21)	Under certain circumstances, individuals have the right to object to the processing of their personal data, unless the organisation has a compelling reason to continue processing.
Right in relation to automated decision-making including profiling (Art 22)	Individuals have the right to get information on automated decision-making, request human intervention, or challenge a decision. Organisations can only apply automated decision-making under certain circumstances and should regularly check whether systems are working as intended. Automated decision-making in this context refers to algorithmic or computational decision-making without any human involvement which has legal or similarly significant effects for the data subject.

Table 2: Overview of data rights as laid down in Chapter 3 of the UK GDPR

For this report we primarily focused on the right of access and the right to data portability, which we collectively refer to in this report as data transparency rights.

Our focus on the right of access and the right to data portability

The right of access and the right to data portability seek to support data subjects by providing more transparency over data processing. The idea is that they can help data subjects to be better informed, so that they are best placed to take appropriate actions in response. The right of access, in particular, can play an important role in "unlocking" other data rights; for example, insight about the data that a company holds about you may be needed to help to decide whether you want data to be deleted ("right to erasure") or whether the dataset needs corrections ("right to rectification") or whether you want a machine-readable copy of the data ("data portability").[25]

Note that the rights of access and data portability can be used as methods for enhancing visibility over data processing practices, by providing information that would otherwise be hard to obtain. Therefore, these rights assist researchers in understanding digital infrastructures and practices, as well as the possible impact of data processing.[26] We use these transparency rights as part of our methodology in this report (see **§3**).

Exercising the right of access and the right to data portability

Data controllers are responsible for responding to subject access requests. The *right to be informed* requires that vendors should provide the name and contact details of the organisation, and, if present, the Data Protection Officer (DPO).[27]

Data subjects (individuals) can exercise their *right of access* by issuing a controller with a "Subject Access Request" (SAR).[28] A SAR allows an individual to obtain a copy of their personal data, as well as details of how that data is being processed. This involves information about the purposes of the collection, retention and processing; categories of personal data collected, held, and processed; identities of those to whom data is disclosed; data retention periods; and other details around data processing practices. A SAR is not bound to a specific format, and organisations cannot require a specific format for a request. Responses to

---

[25] Ausloos, J., & Dewitte, P. (2018). Shattering One-Way Mirrors. Data Subject Access Rights in Practice. *International Data Privacy Law 8*(1), 4-28. https://papers.ssrn.com/abstract=3106632

[26] Ausloos, J., & Veale, M. (2020). Researching with Data Rights. *Amsterdam Law School Paper No. 2020-30, Institute for Information Law Research Paper,* 136-157. https://doi.org/10.2139/ssrn.3465680

[27] Information Commissioner's Office. (2022, October 17). *Guide to Data Protection: Right to be informed.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-be-informed/

[28] Art 12, 15, UK GDPR

access rights should be accessible, concise, and intelligible for individuals, in order to make them comprehensible.[29] This implies that responses to SARs should be tailored towards the specific request.

The *right to data portability* is closely related to the right of access and allows users to obtain a copy of all of their personal data, or request a transfer of their data from one organisation to another.[30] The right to data portability differs from the right of access in that it only applies to data that is *provided* by the data subject to the organisation, so vendors do not have to return details of its processing, and it applies when the data controller is carrying out processing by automated means (i.e., not "paper" records). Further, it applies when the legal basis for processing of the information is "consent" or the "performance of a contract", excluding the common "legitimate interests" legal basis,[31] and only applies to the extent that it does not interfere with the rights and freedoms of others. The personal data returned should be provided in a structured, commonly used, and machine-readable format. This allows users to transfer the data to another controller or analyse the data themselves.[32]

Organisations that receive a "*rights request*" are required to respond within one calendar month of receiving the request. The time frame starts on the day on which the organisation receives the request, irrespective of whether this is a working day.[33] However, when the request is complex or consists of multiple requests, organisations can extend the time to respond to three months. Organisations can ask the individual making the request for a proof of identity or other information useful for verifying the identity of the requester. They should request this promptly, as the timeframe for responding to a rights request commences after the company receives the identification material that was requested.[34]

2.2.3 Data subject rights in practice

Despite the importance of data subject rights, researchers have identified a number of challenges facing those wishing to exercise these rights in practice. For example, there is a range of barriers facing those wishing to request a copy of their personal data, including barriers to communication with the data controller, validating their identity, and avoiding obstacles and delays.[35] Further, it has been shown that the way organisations implement data-protection-by-design or privacy-by-design can "clash" with data subject rights,[36] which can make it more difficult, if not virtually impossible, for those wishing to exercise their rights to actually do so.[37] It follows that it can often appear to be an uphill battle for an individual to exercise their data rights.

There is also work highlighting that there are issues even when a data subject is successful in exercising their data transparency rights. For example, responses to portability requests have been found to be highly inconsistent, and can come in various different file formats, be they technical data structures, tabular files, graphical screenshots, or others.[38] This inconsistency not only means that technical files may make it challenging for non-technical individuals to navigate and interpret their data (a critique raised of information disclosures more broadly[39]), but also limits the ability of more general tooling and guidance to aid in their interpretation. This information can also introduce security implications for the organisation in question, such as instances where the disclosed data reveals something about the software or technical infrastructure that an organisation uses.[40] In all, it has been argued that the content of disclosures is often not fit for purpose in supporting wider aims of transparency and oversight.[41]

---

[29] Information Commissioner's Office. *Guide to Data Protection: Right of access.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/

[30] Art 20, UK GDPR

[31] Art 20(1)(a), UK GDPR

[32] Information Commissioner's Office. *Guide to Data Protection: Right to data portability.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/

[33] Information Commissioner's Office. *Time limits for responding to data protection rights requests.* ICO. https://ico.org.uk/for-the-public/time-limits-for-responding-to-data-protection-rights-requests/

[34] Information Commissioner's Office. *Guide to Data Protection: Right of access.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/

[35] Ausloos, J., & Dewitte, P. (2018). Shattering One-Way Mirrors. Data Subject Access Rights in Practice. *International Data Privacy Law 8(1),* 4-28. https://papers.ssrn.com/abstract=3106632

[36] Veale, M., Binns, R., & Ausloos, J. (2018). When data protection by design and data subject rights clash. *International Data Privacy Law, 8*(2), 105–123. https://doi.org/10.1093/idpl/ipy002

[37] Norval, C., Janssen, H., Cobbe, J., & Singh, J. (2018). Reclaiming data: Overcoming app identification barriers for exercising data protection rights. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 921–930. https://doi.org/10.1145/3267305.3274153

[38] Wong, J., & Henderson, T. (2018). How Portable is Portable? Exercising the GDPR's Right to Data Portability. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 911–920. https://doi.org/10.1145/3267305.3274152

[39] Norval, C., Cornelius, K., Cobbe, J., & Singh, J. (2022). Disclosure by Design: Designing information disclosures to support meaningful transparency and accountability. *2022 ACM Conference on Fairness, Accountability, and Transparency,* 679–690. https://doi.org/10.1145/3531146.3533133

[40] Singh, J., & Cobbe, J. (2019). The Security Implications of Data Subject Rights. *IEEE Security & Privacy, 17*(6), 21–30. https://doi.org/10.1109/MSEC.2019.2914614

[41] Norval, C., Cornelius, K., Cobbe, J., & Singh, J. (2022). Disclosure by Design: Designing information disclosures to support meaningful transparency and accountability. *2022 ACM Conference on Fairness, Accountability, and Transparency,* 679–690. https://doi.org/10.1145/3531146.3533133

2.3 Exacerbating data risks

There are many data protection risks already prevalent in the online world. However, certain properties of the IoT may contribute to *exacerbating these risks,* some of which are listed below.

2.3.1 Expanded scope for data processing

The IoT can be regarded as an extension of the online world into the physical world. Thereby, it expands the potential scope and scale of data collection. IoT devices use sensors to translate information from the physical environment into data, which is then potentially further processed and transmitted. Consumer IoT products typically operate in people's homes and on their bodies ("wearables"), and sometimes offer functionalities that are particularly sensitive, such as in areas related to our health and well-being, and to minors. Therefore, IoT products have the potential to reveal to others a wide range of highly personal and intimate information about ourselves and our lives. Moreover, as IoT products are often sensor-based and potentially "always-on", there is real scope for great increases in the volume and detail of personal data that is processed.

2.3.2 Passive data collection

In the IoT, there is much potential for passive data collection. Typically, those using the Internet will have some degree of awareness as to when they are online, for example when using their phone or computer, opening an app, and so on. This can also be the case in the IoT, for instance, when a user deliberately seeks to interact with a device: perhaps to control the lighting, or stream some content, or by addressing a voice assistant. However, interactions between users and certain IoT products can occur passively, in the background. For instance, a security camera or voice assistant can be a quiet observer from a corner of the living room, or a fitness tracker could be collecting and sharing real-time heartbeat, temperature, and location data. This possibility of passive data collection, without requiring the direct awareness or involvement of individuals, can represent a more surreptitious form of data collection outside of the control and oversight of the individual.

2.3.3 Complexity

The IoT embodies a web of different devices, networks, systems, mobile apps, and Internet applications, and more. While certain aspects may not raise concerns by themselves, issues can arise from the combination. For example, a security camera, a facial recognition application, and a set of smart locks together can collectively form a powerful security system that can automatically detect and provide access to household members, but which also adds complexity which can lead to challenges regarding accountability.[42]

2.3.4 Risks and data protection principles

These risks, among others inherent in the IoT, relate to data protection principles (see **§2.2.1**) in various ways – some of which we now elaborate.

Transparency, lawfulness and fairness (UK GDPR, Art 5(1)(a))
All of the abovementioned risks, in particular those around complexity, touch upon the principle of *"transparency"*. The complexity and interconnectedness of IoT systems can make it difficult to ascertain what happens where, when and why. This complicates transparency, and thereby oversight and strategies for governance.

Furthermore, there also appears a *"lawfulness"* dimension, given the range of data that may be obtained and the question of whether there are valid legal bases for the collection of such personal data. *"Fairness"* also plays a role in the IoT, given the potential harms that can flow from the sensitivity and diversity of the data involved, and because people may not always be conscious that data is being collected about them.

Data minimisation (UK GDPR, Art 5(1)(c)) and purpose limitation (UK GDPR, Art 5(1)(d))
The expanded scope for the collection and processing of data, as well as the capacity for passive data collection, both of which characterise the IoT, may be at odds with the *"data minimisation"* principle when more data is collected than is relevant and necessary for a specified purpose. Furthermore, when the purposes for data collection are not specified, explicit, and legitimate, or when data is further processed (collected, stored, analysed) in a way that is not compatible with the specified purposes, there may be frictions with the *"purpose limitation"* principle.

Integrity and confidentiality (UK GDPR, Art 5(1)(f))
Although not a focus of this report, it is important to acknowledge that there are security risks relating to the data (e.g., data leakage or unauthorised access to data by actors) as well as risks stemming from the interactions of various systems and components, and the presence of sensors and actuators (such as the smart lock example in 2.3.3).

Accountability (UK GDPR, Art 5(2))
The complexity and opacity of the IoT means that there can be a lack of certainty about the responsibility of actors. As the consumer IoT proliferates and IoT ecosystems expand, there is a risk that it will become increasingly unclear where the data is going, and which vendor will be accountable for what.[43]

---

[42] Millard, C., Hon, W. K., & Singh, J. (2017). Internet of Things Ecosystems: Unpacking Legal Relationships and Liabilities. *2017 IEEE International Conference on Cloud Engineering (IC2E)*, 286–291. https://doi.org/10.1109/IC2E.2017.46

[43] See Millard, C., Hon, W. K., & Singh, J. (2017). Internet of Things Ecosystems: Unpacking Legal Relationships and Liabilities. *2017 IEEE International Conference on Cloud Engineering (IC2E)*, 286–291. https://doi.org/10.1109/IC2E.2017.46

3 Methodology

This report concerns transparency in the consumer IoT. Specifically, we focus on the data flows associated with IoT products, as a means for uncovering and providing insight into the data processing undertakings of consumer IoT vendors. In this section, we detail the specifics of our methodology, which entails obtaining and experimenting with a series of IoT products and monitoring and blocking their data flows, in addition to submitting and analysing rights requests and other legal documents.

3.1 Selecting and obtaining IoT products

We sought a range of different IoT products to analyse in order to gain insight into the current consumer IoT ecosystem. Our aim was not to exhaustively test a large list of products, given that this would not be feasible from a technical, time, and financial perspective; and further, because such a survey would quickly become obsolete given the ever-increasing number of consumer products being launched and being made "smart". Rather, in this report we consider 43 products in total, across a range of product types, to provide an indication of certain data processing practices in the consumer IoT.

Our process of selecting and obtaining IoT products to use for the experiments first involved deciding on a range of 11 *IoT product types*. Then, for each product type, we chose three to five *IoT products* based on predefined criteria.

3.1.1 Selecting IoT product types

An *"IoT product type"* refers to a class or category of products that typically have the same or similar functionalities, targeting similar usage. Our criteria for selecting consumer IoT product types were as follows:

1. We only considered product types that are fairly commonplace, reflecting a degree of **popularity** and relevance of the product.
2. To be able to monitor a wide range of potential issues in the consumer IoT landscape, our aim was to capture a **diversity** of product types. As such, we sought product types that vary in degrees of complexity, and in the capabilities and functionalities the products offer. Moreover, we included products that involve the processing of inherently sensitive data, such as fitness-related data or data from children.
3. Products were selected that would be **amenable to testing**, given the practical constraints of our research project. This means that product types needed to be both affordable, and able to operate in the freestanding space of our offices; this excluded products such as smart refrigerators, washing machines, products built into cars or workspaces, and so on. Generally, we selected small, "standalone" products that can be bought and easily used "off the shelf".
4. We considered purely medical devices to be out of scope because these can be subject to other regulations.

The selection process resulted in a collection of IoT products shown in Figure 2. The selection includes the following product types:

1. **Baby Monitors**: Cameras with an app that are marketed as baby monitors.
2. **Child Smart Watches**: Wearables for children, sometimes with GPS (Global Positioning System) functionality.
3. **Fitness Trackers**: Wearables equipped with sensors and GPS.
4. **Motion sensors**:[44] Sensors detecting the motion of nearby objects. The motion sensors in this study are used in combination with "hubs": products that can connect a range of devices within a local network, often using a specific communication technology to manage the products in close proximity.[45]
5. **Smart Lighting**: Connected light bulbs.
6. **Smart Plugs**: Connected power plugs enabling remote switching and power consumption monitoring.
7. **TV Streaming Sticks**: TV or music streaming sticks that can be plugged into a screen.
8. **Smart Scales**: Electronic scales with a range of sensors and functionalities beyond weighing.
9. **Security Cameras**: Cameras used in home settings
10. **Video Doorbells**: Electronic doorbells with video camera functionality.
11. **Voice Assistants**: Smart speakers with voice assistants

---

[44] Two motion sensor products were paired with light bulbs from the same vendor, that were also connected to the hub, as this was required for the functioning of the products. However, only the motion sensing functionality of the products was tested.

[45] Such products tended to use Zigbee, which is a short range radio signal technology that certain products use to communicate.

Figure 2: Overview of product types that are included in this study

3.1.2 Selecting IoT products

For each product type[46], we obtained at least three, and up to five, IoT products.[47] The criteria guiding this selection process were: product functionality, network types, manufacturer, price, and popularity, as outlined in Table 3. The applicability of these criteria depended on the respective product and product type.

Theme	Criteria	How the criteria are applied
Product functionality	Select products that appeared similar in their offerings	Find products with similar sensor types and features
Network types	Include products covering different network types	A mix of devices with Internet connectivity and devices connected by Bluetooth or other radio connectivity
Vendor	Variation in business type, location, and business size to capture diversity	Screening websites and privacy policies
Product price	(Reasonable) variation in cost of products to capture diversity	Searching prices on online stores
Popularity	Where possible, select popular products, to reflect consumer behaviour	Consulting online blogs, Amazon rankings, and popularity statistics. Most of the products we obtained from Amazon were ranked with "best seller" or "Amazon's choice", and shown on the first page

Table 3: Criteria for selection of IoT products

---

[46] Note also that some products can serve several purposes, e.g., a baby monitor and security camera might have similar functionalities. As our focus is on consumer products, we considered a product's category based on the way it is being advertised.

[47] We aimed for at least three products per product type, depending on price, availability, shipping period, ease of procurement, functionality, relevance, and connectivity type. Upon unpacking, some products we purchased were not amenable for testing for various reasons, including that some were repackaged and rebranded duplicates of devices we already had. Nevertheless, we achieved our aim of examining at least three products per product type.

For each product type, we aimed to balance the criteria of the products selected. However, strict balances were not feasible, nor necessarily relevant for every product type. For example, regarding the voice assistants, we have only selected products from "big tech" companies, because the prominent voice assistants are provided by these vendors. In contrast, the child smart watches suitable for research were mostly sold by smaller vendors. For TV streaming sticks, there were essentially four popular brands, none of which were small nor located outside of the EU, UK, or US. In some cases, we bought different products from the same vendor.

Note that different IoT products have varying communication capabilities. Certain IoT devices (which we call *"Internet-enabled")* can connect to the Internet by themselves, which allows the device to directly communicate with external actors, as well as with the companion app. Some IoT devices were Bluetooth only. Because Bluetooth is a short-range radio technology, such devices tend only to interact directly with the companion app (on the mobile device), and depend on the companion app to communicate with others via the Internet.

Overall, the 43 products we examined entailed 33 Internet-enabled products, 30 of these through Wi-Fi and three with Ethernet (as a hub with local radio connectivity). Ten products were Bluetooth-only. Of the 43 products, 39 had a companion app. The 43 products in our experiment came from 38 unique vendors, because some product vendors belonged to the same corporate group.[48]

We also obtained three tablets, all running Android 10, which were used for installing and monitoring of the companion apps of each product, enabling concurrent analyses.

3.2 Interactions with the IoT products

After obtaining the products, we conducted a series of experiments involving interactions with the product. The experiments were designed so as to be reasonably consistent with expected user interactions with such products. During the experiments, we monitored data traffic to and from the IoT products (**§3.3.1**), and afterwards we sent data transparency rights requests to product vendors (**§3.3.2**).

3.2.1 Phases of experiments

The experiments involve different phases that are characterised by types of interactions with the device:

A. The account registration and set-up phase
B. Controlled interactions
C. Idle mode for short period of time
D. Controlled interactions (phase B repeated)
E. Idle mode overnight

The experiments were conducted at various times, with only the companion app running on the tablet. An overview is provided in Figure 3. Next, we describe the specifics related to the device interactions in more detail.

A. Account registration and setup
Most IoT products require registration, which entails the creation of a personal account though the product's companion app or the company's website before it can be used. The account registration and setup phase is included in the experiments, because registration is likely to entail the transmission of personal data. Moreover, this phase can provide insight regarding any data transmission involved with *"device bootstrapping",* i.e., on its initial connection to the Internet and/or companion app.

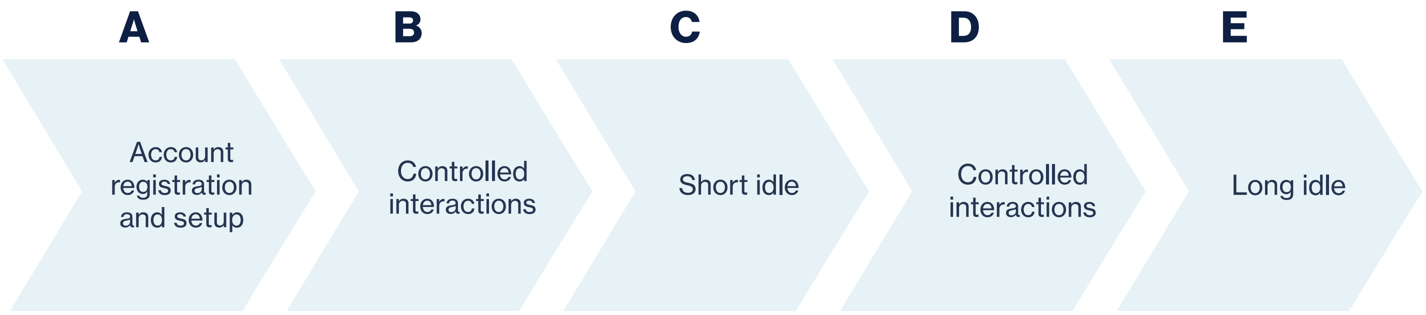


Figure 3: Phases of experiments with IoT devices

---

[48] Note that we observed substantial differences in process and form of responses among members of the same corporate group.

Account registration typically requires personal information such as name, email, and password; but sometimes also requests payment details, weight, or age. The setup phase sometimes, depending on the device, also requires permissions from the mobile device on which the app is installed, for example to use the camera for scanning a QR code, or for access to location or other services to allow the system to detect the IoT device that should be connected.

B. Controlled interactions (product usage)
We "used" each IoT product by undertaking a series of controlled interactions with the device. The interactions were designed before the experiments were conducted, in line with each product type, ensuring that we interacted with the different sensors or functionalities of the device. For example, interacting with the security cameras involved separate interactions with the motion sensor, microphone and camera, and we displayed our hand and face separately, in an attempt to reveal whether the camera recognises certain "objects" of potential interest. The purpose is monitoring for differences in data transmission or usage for different device sensors or functionalities. We defined a process to ensure that each experiment was conducted as similarly as possible among the different products of the same type.

C. Short idle mode
After the first round of controlled interactions, the products were also left running in idle mode with minimal to no interactions. The rationale behind this is that it is common for many IoT devices to *"passively"* collect or send data, that is, without active engagement with the device by the user. The short idle mode took between 70 and 120 minutes. This timeframe was meant to capture periodic data transmission with frequencies of an hour or less.

D. Controlled interactions (product usage) repeated
The controlled interactions of phase B (above) were repeated at a later stage, for three reasons. Firstly, repetition allows comparisons with the first round of controlled interactions, to see whether the monitoring patterns are similar. Secondly, there may be analytics or other processing whereby data is subsequently interpreted differently; a fitness tracker, for example, may only recognise particular physical activities after several uses. Thirdly, it may be that some threshold of activity or data collection is required before being sent to certain actors. Finally, further interactions with the product can help to account for any bootstrapping, such as updates, that might occur during the product's initial use.

E. Overnight idle mode
Finally, we repeated the idle phases of the experiments for a longer duration. This is because periodic data transmission may also occur in patterns longer than the hour of phase C.[49] Here, the idle experiments were conducted over a period of eight to 12 hours, where devices were left in an environment minimising sensory input.

3.2.2 Recording experimental specifics

The experiments provided the foundation for our analysis methods, which include the monitoring of data flows and the data rights requests. We systematically recorded all details and specifics during each stage of the experiments, including device identifiers, the actions that were performed during device usage, the times of these interactions, the personal data manually provided to the device and/or app, and any Android permissions that were granted. These experimental records allow us to:

1. **Interpret the outcomes of data flow monitoring and analysis**. For example, we could link the timestamps of particular uses and interactions with the products with patterns in the data flows we observed, and interpret data flows based on permissions or types of data manually provided.
2. **Guide the correspondence with vendors and the assessment of responses.** The information we recorded assists us in engaging vendors when exercising our data rights, to see if the information provided in vendors' responses aligned with what we had observed.

More specifically, we kept manual records of:

- Dates and times of experiments;
- Device model and app name;
- MAC addresses and IP addresses of the device and the companion app, and other possible identifiers;
- Any data required by the app for device registration, whether this was optional or mandatory; and the data we provided;
- Any permissions that were required (e.g., access to location, camera, contacts), whether these were optional or mandatory, and whether these were granted;
- Whether the app asks for permissions for personalisation or data collection for analytics, whether these were optional or mandatory, and whether these were granted;
- Whether the app asks the user to accept the privacy policy and terms of service, and whether these were asked for separately or simultaneously;
- Whether there were any other points worth noting.

---

[49] Clearly, other durations might be relevant for monitoring, e.g., daily, weekly, monthly, or more; though we did not explore these given the practicalities of conducting such experiments.

3.3 Research approach

To obtain insights into the data processing practices in the consumer IoT, we employed a mix of research methods: (1) data flow monitoring; (2) exercising data rights and analysing responses; and, where relevant, (3) privacy policy analyses.

3.3.1 Data flow monitoring and analysis

We observed where data was transmitted, and the patterns of data flows, to provide insight into IoT data flow practices. This data flow monitoring and analysis was used to obtain empirical insights on the data transmissions between IoT products and actors, be they vendors or third parties.

Some background on network traffic

We first introduce some key technical concepts:

- **IP addresses.** An IP *("Internet Protocol")* address serves as an identifier for a device on a network. One organisation can have multiple IP addresses. Amazon, for example, may host several servers that each have a unique IP address.
- **Network traffic**. Communication over a network from one point (IP address) to another, is what we collectively call "*network traffic*". We also refer to this as *"data flow"* or *"data transmission".*
- **Volume (bytes)**. Indicates the amount of data that flows over the network. *"Bytes"* are units of information used in computer storage and processing. The higher the number of bytes, kilobytes *("KB")* or megabytes *("MB"),* the higher the volume of data.
- **Encryption.** Often a message will be *"encrypted"* to prevent it being read by others while in transit. Unencrypted data flows are like a postcard, in that anyone that comes across one can read the message, whereas encrypted transmissions are comparable with a sealed envelope. When the communications are encrypted, this generally means that nobody but the intended recipient can read the message.

Data collection: setup and network traffic monitoring

The purpose of the network traffic monitoring in this study is to obtain insight into where data goes, and the patterns of data transmission associated with device usage. Monitoring network traffic allows us to observe where data is coming from and going to (through recording the IP addresses involved), as well as the volume of data that is sent, the frequency of data transmission, and other associated information.

The methodology and software used for traffic monitoring was designed by colleagues from Imperial College London in collaboration with others.[50] This involved taking a Raspberry Pi (a small, portable computer) that is configured to provide a *Monitoring Access Point* (MAP) – essentially a network access point that openly accepts connections – which intercepts and records all data traffic flowing to and from anything connected to it. The Raspberry Pi is connected to the Internet, such that the MAP operates as a conduit forwarding any data between whatever is connected to it, and the Internet, as appropriate. In this way, the infrastructure enables one to "spy" on the data transferred to and from the Internet. The software is built on "tcpdump", software that runs on a device to capture data that is transferred to and from the MAP over a Wi-Fi network or an Ethernet (wired) connection. The system does not capture Bluetooth data.

To gain insight into their data flows, we used Wi-Fi or Ethernet (as appropriate) to connect the IoT products (both the companion app and the device) with the MAP (Raspberry Pi), which in turn was connected to the Internet. In other words: data from the IoT products flows via the MAP to the Internet, and vice versa (Figure 4). All experiments were done from a UK-based Internet access point.

Because we are only able to capture data flows relating to the Raspberry Pi's MAP, the nature of data flow monitoring is different for devices that are Internet-enabled, and those that use Bluetooth:

- For **Internet-enabled products**, both the IoT device itself and the companion app can *directly* send information to, and receive information from, IP addresses via the Internet connection.[51]
- **Bluetooth** products, however, only exchange information with the companion app that is connected to the Internet.[52]

In other words, Bluetooth enabled products can only transmit information over the Internet by going through the companion app, whereas Internet-enabled products may communicate directly with the companion app, but can also communicate with others over the Internet directly (see Figure 4). Note that we refer to any data transmission to destinations other than the companion app or the IoT device, as an *"external"* transmission.

---

[50] Mandalari, A. M., Dubois, D. J., Kolcun, R., Paracha, M. T., Haddadi, H., & Choffnes, D. (2021). Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. *Privacy Enhancing Technologies Symposium (PETS) 2021.* ArXiv:2105.05162 [Cs]. http://arxiv.org/abs/2105.05162; Ren, J., Dubois, D. J., Choffnes, D., Mandalari, A. M., Kolcun, R., & Haddadi, H. (2019). Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. *Proceedings of the Internet Measurement Conference*, 267–279. https://doi.org/10.1145/3355369.3355577; See also: https://github.com/djdubois/moniotr-core

[51] Some devices have a hub that connects to the MAP with an Ethernet cable; and then uses a different radio communication type, like Zigbee or a proprietary radio, with which the devices interact. Given that these hubs generally connect directly to the Internet, and are similarly monitored by our infrastructure (albeit through Ethernet rather than Wi-Fi), for simplicity we treat these as "Internet devices" in this report.

[52] Note that Bluetooth is a short-range radio communication mechanism, very commonly used in the IoT.

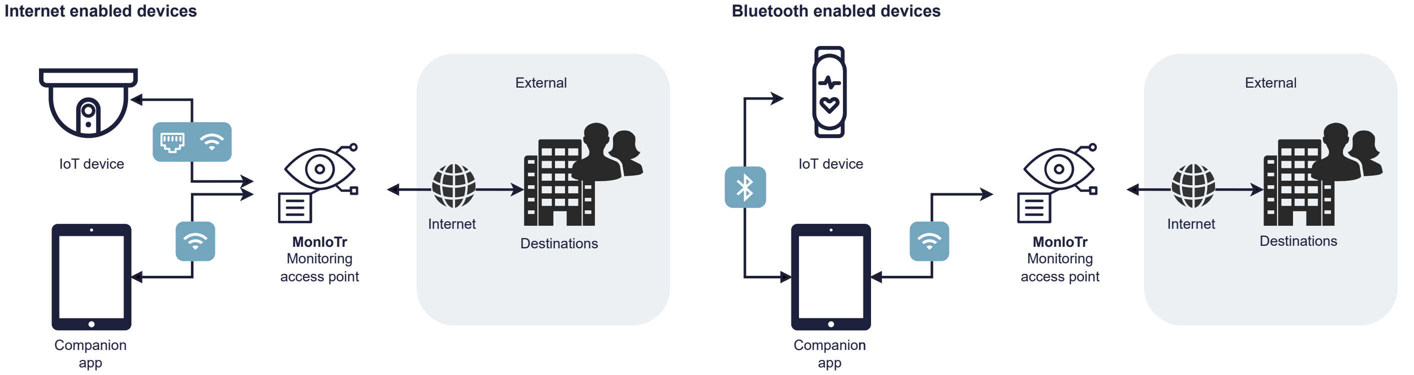


Figure 4: Overview of the experimental setup. The monitoring access point (MAP) was used to intercept and record data flows between devices with an Internet connection, the apps, and external actors

Data collection and preparation

The data monitoring infrastructure captures all traffic for each IoT component that is connected to the broader Internet via the MAP (the monitoring access point hosted by the Raspberry "Spy"). Therefore, monitoring an IoT product entailed capturing information on (1) the companion app, by connecting the tablet on which it operates to the MAP; and (2) for Internet-enabled devices, also connecting the IoT device to the MAP. The network traffic data is stored separately for each component that is connected to the MAP, by saving it in "pcap" files (a file format for storing computer network traffic). We captured the data in separate files for each phase of the experiment.

Note that the information about the companion apps, as captured from the tablet, raises particular uncertainties, in that there is other software which operates on the tablet – including the operating system, background applications and so forth – that can also involve network traffic. This means that during experimentation, traffic unrelated to the particular companion app being examined could potentially be captured.

We undertook several steps attempting to mitigate this problem of capturing traffic unrelated to the experiments. Though we only had the companion app for the product we were testing open on the tablet, and the other apps closed, we also undertook a range of different monitoring exercises to help us to identify, and thereby filter out, particular communications (IP addresses) that appear unrelated to our experimental analyses. In particular, we monitored the connections of the tablets when they were running idle after a factory reset, as well as those involved in downloading the app from the "app store" (but not the app's initial execution), and filtered these addresses out of the datasets we used in our analyses. We also monitored the tablets with all companion apps installed but with none of the applications open, to indicate which companion apps might have background transmissions; and undertook analyses to link these with particular products. This helped to inform which addresses to consider, and which to ignore, for the analysis of a particular product.

We note that, while this is a highly imperfect process and uncertainties remain, this issue concerns only the companion apps and not the devices, and many substantive traffic observations are clearly related to the product (e.g., the timing and volumes of traffic observed directly correspond to product use). Therefore, these observations still provide useful indications of trends and behaviours in this space.

Data analysis

The "pcap" files (network traffic logs) were analysed using network analysers called Wireshark, Tshark, and Pyshark (a python wrapper for Tshark). The data analysis consisted of destination analysis and data flow analysis.

Destination analysis

Using the pre-processed information from the data transmission logs ("pcap" files), we undertook analyses to extract the IP addresses involved in the data transmissions, and then derive the organisations and the countries associated with these addresses. For this, we first extracted the IP addresses occurring in the transmission logs for each experiment. For each IP address, we counted how many bytes were either sent to or retrieved from this address during each phase of the experiments (see **§3.2.1**). Then, we used third-party IP-address directory services (IP Stack[53] and IP Info[54]) to obtain information about the organisation and country associated with the IP address. We supplemented this information with manual analyses, involving searching for specific aspects of the identities related to certain IP addresses deemed to be of potential interest.

We gathered further information on destinations by extracting information from DNS (Domain Name System) data and from HTTP (Hypertext Transfer Protocol) data. DNS and HTTP are Internet protocols. The Domain Name System

---

[53] IP Stack. https://ipstack.com/
[54] IP Info. https://ipinfo.io/

(DNS) supports Internet users in navigating to a certain web page, such as www.wikipedia.org, by connecting the domain name to the IP address. As such, domain names can potentially tell us something about the identity related to the IP addresses involved in data transmission. The Hypertext Transfer Protocol (HTTP), which helps to drive the web, is essentially a set of rules used for the transfer of data. HTTP data contain a “request URI” (Uniform Resource Identifier) that sometimes reveals information about the type of destination involved. For example the following request URI shows that “aliyuncs”, which represents Alibaba Cloud, is involved in this data flow: [Request URI: http://britain.oss-eu-west-1.aliyuncs.com/EKDB_2D882E85-XXXX-5175-DE4A-F5EB6039846E/20220322/20220322140206.ts2?append&postition=0]

Data flow analysis

Because most data transmission is encrypted, we only know that data transmission occurred, and cannot see the actual content of the data exchanged. However, knowing when and where data was transmitted, among other aspects of the transmission itself, is still informative even without seeing the content. As such, we have analysed the frequencies and patterns in data transmission to identify whether and how particular interactions and uses of an IoT product, as recorded in our manual experimentation records, are reflected in the product’s data transmission. This can help to indicate that some data about the IoT product and its usage is being communicated.

To support our analyses, we visualised data transmission in bytes transferred for each companion app and Internet-enabled device during the experiments, separating the incoming from the outgoing data flows.

3.3.2 Data rights requests and analysis of responses

After conducting the experiments with the IoT products, we issued data access and data portability requests (“transparency rights requests”, “rights requests” or “requests”). The purpose was to assess the implementation of these data transparency rights by organisations in the IoT, and to obtain more insight into the data processing practices of vendors. There were three rounds of request: the initial request, a follow-up email emphasising some elements of the requests, and a second follow-up email for some vendors, presenting certain observations from our traffic analysis. There were also various reminders sent in between. We next describe the nature of the requests and the process of sending requests. We then analysed the replies we received from vendors about these requests (“responses”).

Designing and sending data rights requests

The initial data rights request consisted of two parts.[55] The first part was a request for a copy of the personal data provided by and observed from the user. It was explicitly stated in the request that this included personal data provided or otherwise obtained by account registration and setup of the product; data obtained by sensors from the product; or any other data obtained through or about the product or a device on which the companion app was installed. In addition, we asked about settings, preferences, and device permissions. The data was requested in a structured, machine-readable format with a description of all fields.

The second part of the request consisted of a set of questions concerning the *use of data* by the vendors. This included questions on the identities of joint controllers, any parties to whom the data had been disclosed, and the purposes and lawful bases for processing of the personal data, by category of personal data. Vendors were asked if any automated decision-making had been done or inferences made based on the data, and if so, to provide information on the logic behind this. Moreover, we asked about the periods for which data would be retained and the locations where the retained data would be held. In some cases, we included specific questions such as clarifications about sections of the privacy policies or about vendor’s business structures (e.g., whether data is shared with sister companies or subsidiaries).

The initial request was a detailed request sent to all vendors, after having executed the experiments. The email was lengthy (over 900 words), included legal wording, and contained references to the GDPR. To most vendors, the request was sent by email. However, some vendors did not refer to an email address in their privacy policies, but instead had an online portal with a text box with limited word count on their websites. In these cases, the same request was submitted with a link to an Internet-accessible PDF.

Given that the response rates to the initial request were low, and that none of the vendors had sent a full response addressing all elements of our requests, a follow-up request was sent. This follow-up differed from and complemented the initial one, in that it provided a summary of a few key elements of the initial request, including questions about data collection, purposes, data disclosure, storage, automation, and analytics to support the initial request. This follow-up request included a table that vendors could use to set out the purposes, lawful bases, and categories of personal data (Table 4). The purpose of sending this follow-up request was twofold. First, the aim was to trigger a response from vendors that had not yet responded to the first request, or vendors that had only, for example, returned a copy of a personal dataset but did not respond to our questions. Second, it was to invite a more specific, accurate or complete response from vendors that had already sent some (generally inadequate) response.

---

[55] The content of our initial request was informed by, and adapted from, a template for data access: Veale, M. (23 July 2019). *A better data access request template*. https://michae.lv/access-template/

Purpose	Lawful basis	Category or categories of personal data

Table 4: The empty table provided to vendors in the second round of requests, as a way to encourage responses that include information on the purposes of data processing

Vendors who responded to either of the requests, but whose responses were incomplete, received a further follow-up request to guide, challenge and encourage them towards providing a more detailed and meaningful response. To this end, the third round of requests included findings resulting from our traffic monitoring analyses, such as graphs of data flows, locations of the destinations involved, indications of the number of different IP addresses that had received data, permissions or data provided to the device, and other relevant information. Presenting evidence to vendors in this way was used as a strategy to invite a better quality of response.

The process of making this request consisted of three rounds with different formats. An overview of these formats is provided in Table 5.

Analysing the requests

For each product, we collated the communications with vendors and the copies of personal data and other information they provided in their responses. We then conducted a thematic analysis[56] to identify themes and patterns regarding the process of interacting with vendors, and the format and nature of their responses (see Section 4).

For each vendor that provided more meaningful responses, we extracted the data types they collected, the identities to whom data was disclosed and countries to which they stated data was shared, and any other information relevant for guiding the analysis of data processing practices (Section 5).

3.3.3 Analysis of legal documents

To guide the analysis of data rights requests, we analysed privacy policies and other legal documents of the vendors from their websites. This involved examining the documents for provisions regarding the:

- Data types collected;
- Purposes of collection;
- Identities of those to whom data has been disclosed;
- Categories of organisations involved in data processing;
- Countries to which data is transferred and/or processed;
- Details of storage and retention.

For the legal analysis related to the blocking of data flows (**§6**), we analysed the terms of service, terms of use, terms and conditions and other documents of a similar nature, to identify any statements, restrictions or prohibitions that could potentially be interpreted as relating to the blocking of data flows.

	What it entails	To whom	Result
Initial request (generic)	- Data access and portability request - Lengthy (900-1000 words) - Legal wording	All vendors with contact details (number of requests = 38)	16 responses
First follow-up email (generic)	- Data access and portability request - More concise (300-400 words) - Summary of initial request	All vendors with contact details (number of requests = 38)	26 responses (of which 14 from vendors that responded to our initial request)
Second follow-up email (specific)	- Presented outcomes and observations of data flow analysis (e.g., graphs and geographical locations) - Specific questions about previous responses, privacy policies and other details	Vendors that were responsive (number of requests = 22)	10 provided significantly more specific answers.

Table 5: Vendor responses for different rounds of data transparency requests

---

[56] See Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological*, 57–71. American Psychological Association. https://doi.org/10.1037/13620-004

3.3.4 Blocking data flows

Blocking data flows prevents certain data transmissions from occurring. It has been suggested that blocking mechanisms can empower IoT users by allowing them to take some control over data transmission (see **§6**).

We considered whether blocking appears feasible, desirable and sustainable. To this end, we conducted two case studies to ascertain whether blocking selected IP addresses, without stopping a device from functioning, appears possible from a technical perspective. Informed by the data flow analysis, we selected two products – one of which communicated with relatively few destinations, and the other which communicated with many destinations – and leveraged the functionality of our traffic monitoring software to block particular IPs. We also explored whether blocking already appears to be considered by vendors. This involved collecting and analysing the Terms of Service across the range of products, to investigate whether any terms appear to relate to data flow blocking.

3.4 Limitations

Our methodology has some limitations. First is that all experiments, responses and privacy policies *reflect a specific moment in time*. Different elements of the IoT landscape, involving product design and operation, privacy policies, and the behaviour of vendors, can of course change. Further, there are several specific limitations relating to certain aspects of the methodology, which we now describe.

3.4.1 Device selection and purchase

Although we have attempted to obtain a diverse batch of IoT products, *our device selection and purchase does not reflect the full consumer IoT market.* Various factors limited the selection and purchase of devices – naturally, available budget, time and space are key factors determining the breadth of our experiments. In terms of purchasing, with the exception of four devices from very well-known brands that were not available via the retail platforms we used, the selection was limited to what was available on prevailing online retail platforms in the UK, and purchases had to accord with the conditions of our institutional procurement policy, thereby limiting the retailers from which we were able to purchase products.

3.4.2 Data flow monitoring

Our observations about the flow of data are not necessarily indicative of the impact of data sharing, because *our knowledge of data transmission between the IoT product and the IP address is limited*. One issue is that because the data transmitted is most often encrypted, data flow monitoring does not allow for analysing what data is sent.[57] Moreover, products can communicate with an IP address without necessarily sharing any meaningful information. Therefore, when we mention that a product "talks to", "interacts with" or "contacts" an IP address, relating to a particular organisation or country, it just means that a communication channel was established and some transmissions occurred. For example, product A might transmit high volumes of fairly benign data to a large number of destinations, while product B could send small amounts of sensitive data to one significant destination. This means that it is important to be cautious in drawing conclusions about the actual implications of the communications that were observed.

Our data flow monitoring approach provides visibility over the data flows *directly between* the IoT product and an address, but *not what happens to the data after arrival at that address*. For instance, we may observe that data is sent to a server in California, but we cannot see what happens afterwards. Indeed, this highlights the importance of transparency rights, as a mechanism to provide visibility beyond what we can observe through technical monitoring.

3.4.3 Device interactions

Our *experiments represent simulated, rather than real-world device usage*, which might impact the data that vendors hold about us. Although the experiments involve different types of device interactions (**§3.2**), that were designed to accord with the nature of the product, the experiments were conducted in a systematic way over a relatively short period of time. As a result, the data generated by our experiments might differ from that collected from a typical user. For example, vendors may not have shared our data with advertisers, because they had not yet collected enough data on us to create a sufficiently detailed user profile; or perhaps certain fitness trackers may need more personal data for undertaking more advanced profiling and inferences.

3.4.4 Data rights requests

Our process of exercising data rights likely differs from that of a common IoT consumer. It seems unlikely that the average user would achieve similar responses, given the time, energy and persistence required in continuing to advocate for meaningful responses to such requests. This means that the response rates in this study may differ from the average response rates for consumers.

Vendors may only hold certain data for limited periods, hindering visibility. Vendors have one to a maximum of three months for responding to our requests. However, we saw, for example, that one vendor only holds video recordings for several weeks. This illustrates that data may have already been deleted before the company receives the rights request or finalises their response to the request. As a result, certain data (or certain types of data) may not have been included in the copy of personal data returned.

---

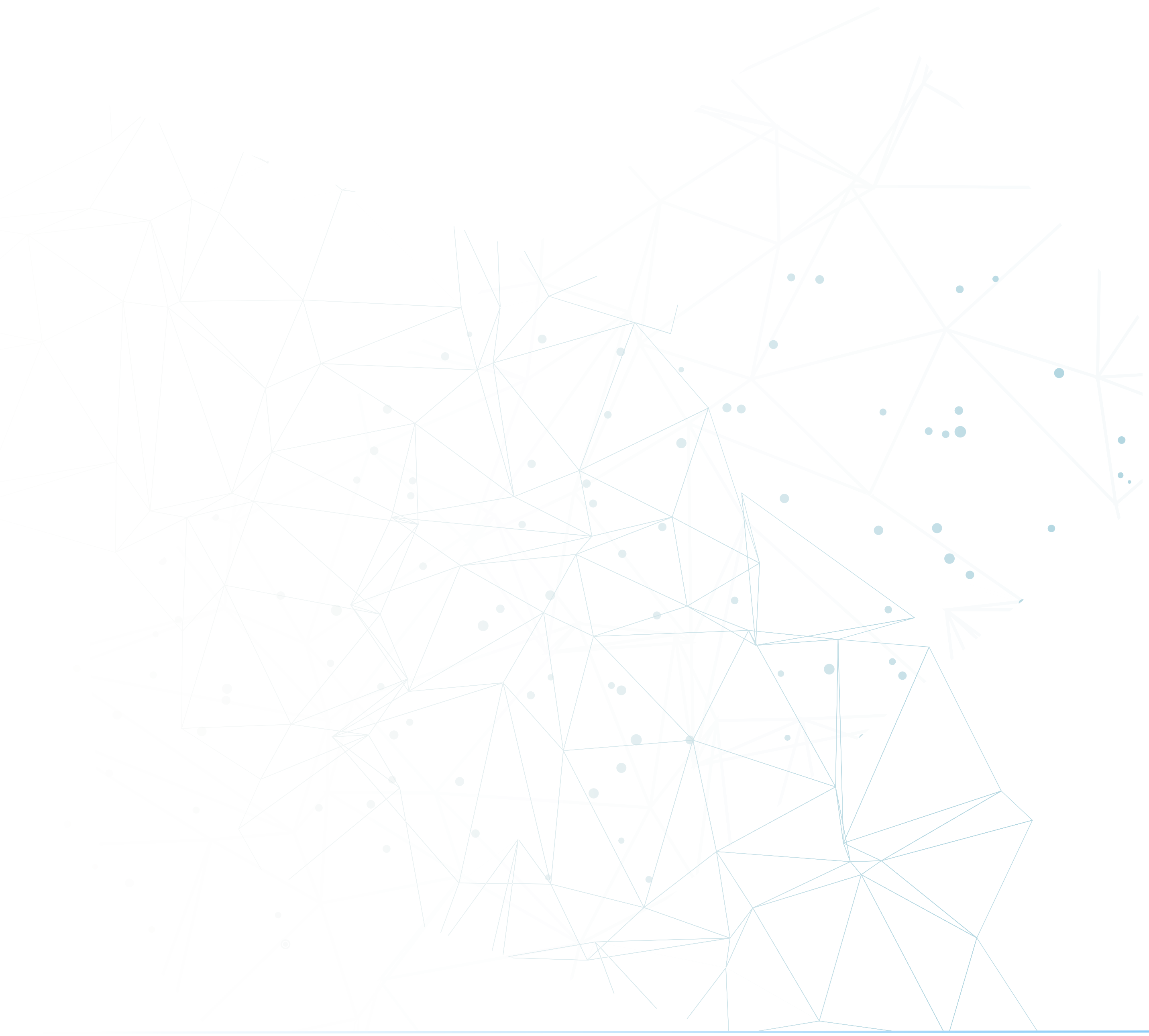
[57] There are approaches that can attempt to work around the opacity of the content of data transmissions, but these are limited in general, wide-scale efficacy. See Ren, J., Dubois, D. J., Choffnes, D., Mandalari, A. M., Kolcun, R., & Haddadi, H. (2019). Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. *Proceedings of the Internet Measurement Conference,* 267–279. https://dl.acm.org/doi/10.1145/3355369.3355577

Section 3: Summary

Our research approach entailed the following:

- We selected and obtained 43 IoT products across 11 different product types;
- With these products, we conducted a series of experiments involving the setup, operation and use of the products, while monitoring their data transmissions;
- We sent vendors subject access and data portability requests to obtain copies of the personal data they held, and information about what data they collect, who the recipients are, how long it is stored, and other aspects of data processing;
- We analysed the observed data flows, responses to requests, and any relevant legal documents;
- We conducted case studies from technical and legal perspectives to explore the considerations of blocking certain data flows.

The following sections discuss the outcomes and findings of these analyses.



4 Vendor implementation of data transparency rights

This section describes the ways vendors dealt with our data transparency rights requests. In particular, we explore the response rates of vendors (**§4.1**), the process of executing the requests and highlighting some of the challenges in dealing with vendors (**§4.2**), and the forms of vendor responses (**§4.3**). This provides an indication of the current state of vendors' data rights practices more broadly.

4.1 Overview of response rates to data access and portability requests

The response rate to our data transparency requests was generally poor. From the vendors of 13 products, we did not receive any information about the data held nor their data processing practices. Most of the remaining vendors needed to be encouraged with follow-up emails before they provided some form of response aligning to the rights requests. While perhaps predictable that vendors based outside of the UK or EU were particularly lax in responding to rights requests, those with a UK or EU presence did not always perform better. We now provide some details about vendor response rates.

	Number of products
Vendors we could not contact to issue a request (i.e., no contact details)	5
Vendors to whom we sent a request, but who did not respond at all	4
Vendors to whom we sent a request, but who did not send a relevant response	4
Vendors from whom we received a response that was at least partially relevant to the rights request	29
Vendors from whom we received a copy of personal data held	23

Table 6: Response rates to data transparency requests

4.1.1 Several vendors did not address any aspects of our data access or portability rights requests

Of the 43 products we tested, 13 vendors did not appear to comply at all with the rights of access and data portability. Based on their websites and public legal documents, most of these vendors were relatively small companies, and almost half of those appear to have a presence in the UK, be it an office, a UK-oriented website or a UK phone number. As Table 6 illustrates, some of these vendors we could not approach, others never responded, and some did not provide a meaningful response.

To elaborate, five product vendors did not provide contact details at all. This is concerning, given that the applicable data protection law states that all vendors processing data of EU or UK citizens or residents should provide their contact details to the data subjects.[58] The products without contact details were mostly low-budget, representing two vendors of small fitness trackers, two video doorbell vendors, and a child smart watch vendor[59]. It is unclear if these vendors have a UK or EU presence, as none of the products had a website.

In addition, four other vendors that were contacted did not respond at all, neither to our original request nor to our follow-ups. These include smart lighting vendors, one based in The Netherlands and the other in the US, and a child smart watch vendor with presence in Israel and Latvia.

A further four vendors did answer, but did not provide any meaningful responses to our questions about data processing practices or the data they collected. The interactions with these vendors are described in **§4.2.**

---

[58] Art 13(1), UK GDPR

[59] This company now has a website, but did not at the time of analysis.

4.1.2 Generic and specific follow-up emails were sometimes effective in eliciting responses

The vendors for the remaining 29 products provided some form of response that related to our rights request. However, many vendors were superficial in their responses, as far as the information they provided on data transmission and usage was concerned. In an attempt to obtain as much visibility as possible over data collection and usage, we sent vendors follow-up emails to provoke more meaningful answers. This entailed generic follow-up emails being sent for all products, and a second round of specific follow-up emails for 22 products for which we had already received some response to our questions about data processing but where such responses were incomplete. The follow-up emails aimed to encourage a response and to provide more guidance for vendors in dealing with the rights request. The process and nature of these follow-up emails is further described in **§3.3.2**.

As Table 7 illustrates, we only received 16 responses from vendors to our initial request. To our first follow-up email, we received 26 responses. Fourteen of these 26 came from vendors that did not respond to our initial request, suggesting the follow-up email helped to encourage a response. The remaining 12 responses came from vendors that had already responded to the initial request, but whose responses to the follow-up request were somewhat more complete; for example, by answering previously unaddressed questions. After sending a second round of follow-up emails, we received somewhat more detailed responses relating to 10 products.

Overall, we received responses regarding 29 (of 43) products that appeared to be related to our rights requests, though we note that the process was cumbersome, and the information contained in the responses varied significantly.

4.2 The process of interacting with vendors

In **§4.1** we described the response rates to our data transparency requests. However, we came across several issues in the process of exercising our data access and portability rights, which we now discuss.

4.2.1 Automated data rights processes can help or hinder

A number of vendors return a copy of the personal data held by means of an automated process, which limits the scope for executing data access and portability rights. Such vendors would offer portals on a website where users can log in and download a copy of their personal data. In some cases, users can select which types of data they seek and which not. The advantage of such systems is that retrieving a copy of personal data is relatively quick, individuals are not bothered with having to author and send an email, and copies of personal data were generally detailed.

However, for additional questions about data processing practices, related to the right of access, the same vendors would typically provide a portal with a text box, sometimes with limited character count, for sending privacy or data related questions.[60] Where the word count of our subject access and data portability request exceeded this limit (as occurred on several occasions), we would submit a link to a PDF detailing our request.

We observed that automated methods of providing access to personal data were most common among the more established technology vendors. However, this approach limited our freedom to specify form and content.

	What it entails	**To whom**
Initial request (generic)	- Data access and portability request - Lengthy (900-1000 words) - Legal wording	All vendors with contact details (n=38)
First follow-up email (generic)	- Data access and portability request - More concise (300-400 words) - Summary of initial request	All vendors with contact details (n=38)
Second follow-up email (specific)	- Presented outcomes and observations of data flow analysis (e.g., graphs and geographical locations) - Specific questions about previous responses, privacy policies and other details	Vendors that were responsive, i.e., sent some form of reply to our questions (n=22)

Table 7: Vendor responses for different rounds of data transparency requests

---

[60] For instance, one vendor provides a Data Access Request Form with space for a maximum of 1000 characters; another offered a portal for a Personal Data Management Request with a text box for a maximum of 4000 characters.

The automated processes typically, at least initially, resulted in more prefabricated responses to our data rights related requests, rather than responses tailored to our particular concerns (see **§4.3**).

4.2.2 Employees handling requests may lack relevant knowledge and skills

In several cases, it appeared that the employees handling the requests were unfamiliar with data subject rights. This is reflected by some vendors apparently treating our request as a concern, warranty claim, or complaint. For example, a smart scale vendor asked for the order number. When we indicated that we had ordered the product via Amazon and not the vendor's online shop, the vendor responded: "As this was purchased from Amazon, I'm afraid you'd have to go through them". Another vendor initially responded with "we are investigating the problem mentioned in your email" and "your feedback is well received" despite the fact that our requests did not mention concerns or feedback.

There are further indications that vendors' staff lack the necessary technical understanding to respond to the requests. A smart scale vendor indicated that the scale's data is stored through a cloud system. When we mentioned that we observed data flowing to multiple destinations and asked about the identity of their cloud provider, the vendor replied: "sorry that we also don't know about the information about the identity of the cloud provider", followed by the comment: "As a representative of [vendor] support team, we are glad to offer some common consults and technical support." When we asked if they had consulted colleagues that are concerned with data management whether they can return a copy of personal data or know the identity of the cloud provider, the vendor ignored this and replied with instructions to access our data via the companion app.

We have many other examples suggesting that frontline, consumer-facing staff lack relevant knowledge of data rights. Some staff indicated that they needed to consult a technical or legal expert. A smart scale vendor employed external legal counsel for responding to data access requests, which resulted in them revealing a list of third parties to whom our data is disclosed. Another vendor told us they needed to consult their technical team, but they ultimately never responded (see "Narratives of company interactions"). When confronted with our observations concerning numbers and locations of recipients of data flows from the device, a video doorbell vendor responded: "Regarding the last inquiry [question about recipients], we already send [sic] the relevant department and will provide you with the proper answer ASAP", and afterward they provided an overview of processors, details of whether the processing is done internally or externally, and the associated countries and purposes.

4.2.3 Vendors sometimes requested more information

Some vendors requested additional information to fulfil the request. This varied from providing an order number or the device info, such as serial number and account ID, to more private information such as an identity document.

Again, it appears that certain vendors seemed to confuse our request with warranty claims. For instance, three vendors asked for order numbers to process the request, possibly showing a lack of awareness of the difference between the exercise of data rights and product-specific (e.g., warranty) concerns.

Two vendors asked us for documents as proof of identity. For instance, a child smart watch vendor and a smart scale vendor required a proof of identity (passport or driving license). The watch vendor also required proof of residence. Vendors are entitled to request identity documents to verify the identity of the requester;[61] however, it is questionable whether it is proportionate in these circumstances. Proof of identity was not required for registering any one of the IoT products we tested; indeed, users could just as well have provided random names and details for account registration so that their identity information would not be directly and obviously linked to them as an individual. Furthermore, there could be other ways to identify the user that are potentially less privacy invasive. For example, some vendors provided mechanisms to do this through their software, the supplied phone number, etc.

Moreover, the identification process, be it an order number or an identity document, can add delay to the process of executing data rights, as it is after proving identity (if requested) that the timeframe to respond to a rights request begins.[62] We observed this in practice, where a smart scale vendor only asked the proof of identity months after the initial request was sent, thereby adding additional delay to an already lengthy process.

4.2.4 Other issues in the process of communicating with vendors

There were several other obstacles that complicated the process of executing our data rights. For example, we came across issues with the email addresses of vendors. Two vendors, after weeks or months of waiting for a response, referred us to another email address and required us to

---

[61] Information Commissioner's Office. *Guide to Data Protection: What should we consider when responding to a request?* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/right-of-access/what-should-we-consider-when-responding-to-a-request/

[62] Information Commissioner's Office. *Guide to Data Protection: Right of access*. ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/

submit or forward the request again, thereby significantly delaying the process. One vendor, over the course of our communications, deactivated their email address without updating us about it ("Oh no! The email address you contacted is no longer active" was the response we received when following-up).

We also observed that internal issues in the vendor's business could affect the responses. A vendor mentioned that they were in a receivership process, and that they are, therefore, "obligated to convey all GDPR requests through their secondary approval process, as they now act as interim controllers with obligations under the GDPR act" and as a result there would be "a slight delay in providing you with your data request". This vendor eventually sent us a reasonable response.

Some vendors told us to "trust" and "rest assured" that they care about our privacy or that our data is in good hands, without coming back with a substantive response. These responses suggest a general lack of appreciation among some vendors of the importance of transparency and data rights.

Narratives of company interactions

Building on **§4.2**, we present four narratives highlighting the cumbersome experience of our interactions with certain vendors when exercising our transparency rights. Note that none of the vendors discussed here ultimately provided an adequate response.

A smart plug vendor from China that has a presence in the UK began by asking for our order number. After providing that, they indicated that they did not understand the request. When we provided a short explanation of the GDPR and data rights, they responded with:

"We believe that you are a professional and warm customer. We have forwarded your relevant questions to the R&D team. Once I receive their reply I will reply to you as soon as possible".

Subsequently, after some reminders, and having shared an extract from their privacy policy which indicates that the user can exercise data rights including the right to access and data portability, along with the contact email to which we initially sent our query, the vendor told us to contact another company. This other company never responded, even after we sent them a reminder.

A baby monitor vendor from the US with a UK webshop responded to the initial request, but only after we sent a reminder:

"We will get back to you with the detailed reply shortly."

Over the months after this message was sent, we sent two follow-up emails, including an overview of certain outcomes of the data analysis, as well as a reminder:

"Is there any additional information you need to be able to fulfil my request?"

but we did not receive any further response.

A child smartwatch vendor with a UK customer service phone number started by asking for our order number, as well as the model of the mobile phone linked to the watch, because they said that their privacy policies for Android and iOS are different. When we indicated we had used an Android tablet for the app, the vendor (incorrectly) said their watches are not compatible with tablets and can only be connected to mobile phones. Eventually, after some further interactions, including providing links to the ICO website, the vendor said:

"We still don't understand what you mean. However, you can rest assured that downloading the application on your phone does not violate your privacy."

After having forwarded them our initial request and referred them again to the website of the ICO, we have not heard anything further from this vendor.

A security camera vendor headquartered in China but with offices in the UK and the EU had to consult their technical team for our request, and eventually argued that they cannot access the data for security reasons. To our initial request, they responded:

"We understand how important this is for you. However, as much as we would love to help you further, we do not have any information, the one that you are looking for. The reason why is because this is for security purposes of our customers and users. We know that this is not the answer that you are looking for but we are hoping for your kind understanding".

We asked them if they could clarify if they do not have this information or cannot provide this information, and for what reasons. Their response was:

"we do not have any information, the one that you are looking for. The reason why is because this is for security purposes of our customers and users."

In response, we sent them follow-up emails with clear and specific questions about the destinations of the data flows we observed, data retention and sections of the privacy policies. We also presented them with examples of data we provided to them, such as account registration information. The vendor said they forwarded this to their technical teams. After having sent a reminder and after months of waiting, the vendor told us:

"Regarding your concern, this has already been forwarded to our advanced technical team however, due to your own security and privacy purposes, we are not able to access any of your information".

We responded to this with:

"If your advanced technical team is not able to access the data, could you please inform me what technical and organisational measures they have used to limit access to the data, and if and whether any third parties will be able to access the data obtained from the [vendor] camera or the companion app? Furthermore, I observed data going to AWS and other servers. Could you please inform me what data are flowing to which third party, and for what purposes?"

Instead of responding, the vendor told us:

"Rest assured, this has been raised to our relevant team. We'll get back to you as soon as we get a reply from them."

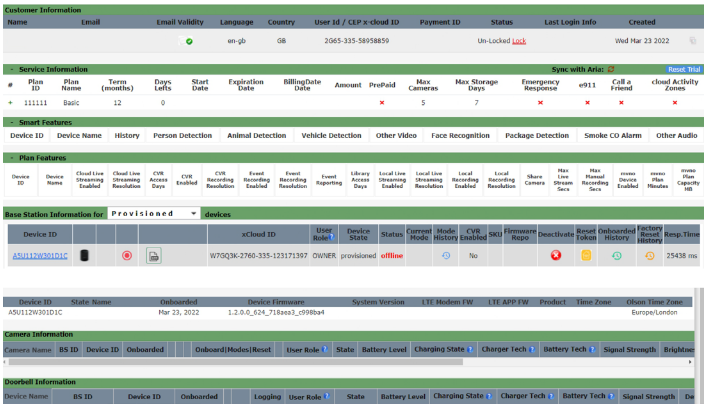
They also told us they had a new email address. We forwarded our request to the new email, but heard nothing further.

4.3 Form and content of the responses to data transparency rights requests

Our request explicitly mentioned the right to data portability as well as the right to access. We now describe the form and content of the responses we received for these two types of rights request.

4.3.1 Data access: responses were generic and sometimes prefabricated

In our requests, in addition to asking for a copy of the personal data held, we asked specific questions about the sources, purposes, sharing, storage duration, and other aspects of the processing of our personal data. However, the responses we received from vendors were often generic and not adapted to specific questions, especially in the initial response. Vendors would often, for example, only selectively respond to questions (e.g., mention types of data collected but not disclose categories or identities of recipients), simply return parts of their privacy policies, or send instructions for how to manage data in their app or on their websites. Several vendors would only send a copy of our personal data and a link to their privacy policy or to privacy settings on their website or app. Even when vendors responded to our questions, their responses rarely addressed all the questions, nor did they address the entirety of the questions they did answer. For example, we asked vendors to specify to whom they have disclosed data, and how this relates to the particular categories of personal data and purposes they describe, yet only four vendors returned information in this way.



Customer Information

Name	Email	Email Validity	Language	Country	User Id / CEP x-cloud ID	Payment ID	Status	Last Login Info	Created
		✓	en-gb	GB	2G65-335-58958859		Un-Locked Lock		Wed Mar 23 2022

- Service Information Sync with Aria: Reset Trial

#	Plan ID	Plan Name	Term (months)	Days Lefts	Start Date	Expiration Date	BillingDate Date	Amount	PrePaid	Max Cameras	Max Storage Days	Emergency Response	e911	Call a Friend	cloud Activity Zones
+	111111	Basic	12	0					✗	5	7	✗	✗	✗	✗

- Smart Features

Device ID	Device Name	History	Person Detection	Animal Detection	Vehicle Detection	Other Video	Face Recognition	Package Detection	Smoke CO Alarm	Other Audio

- Plan Features

Device ID	Device Name	Cloud Live Streaming Enabled	Cloud Live Streaming Resolution	CVR Access Days	CVR Enabled	CVR Recording Resolution	Event Recording Enabled	Event Recording Resolution	Event Reporting	Library Access Days	Local Live Streaming Enabled	Local Live Streaming Resolution	Local Recording Enabled	Local Recording Resolution	Share Camera	Max Live Stream Secs	Max Manual Recording Secs	mvno Device Enabled	mvno Plan Minutes	mvno Plan Capacity MB

Base Station Information for Provisioned **devices**

Device ID					xCloud ID	User Role	Device State	Status	Current Mode	Mode History	CVR Enabled	SKU	Firmware Repo	Deactivate	Reset Token	Onboarded History	Factory Reset History	Resp.Time
A5U112W301D1C			●		W7GQ3K-2760-335-123171397	OWNER	provisioned	offline			No			✗				25438 ms

Device ID	State Name	Onboarded	Device Firmware	System Version	LTE Modem FW	LTE APP FW	Product	Time Zone	Olson Time Zone
A5U112W301D1C		Mar 23, 2022	1.2.0.0_624_718aea3_c998ba4						Europe/London

Camera Information

Camera Name	BS ID	Device ID	Onboarded			Onboard|Modes|Reset	User Role	State	Battery Level	Charging State	Charger Tech	Battery Tech	Signal Strength	Brightnes

Doorbell Information

Device Name	BS ID	Device ID	Onboarded		Logging	User Role	State	Battery Level	Charging State	Charger Tech	Battery Tech	Signal Strength	De

Figure 5: Overview of copy of personal data retrieved from a video doorbell vendor in PDF format

While such non-specific responses were common, these were typical for vendors that automated the processes of exercising access requests. From these vendors, we generally received prefabricated written responses, with links to relevant documents and references to privacy policies or portals for downloading personal data. Their responses were generic, standardised, and ignored certain elements of our questions. While these sorts of responses were typical of the larger vendors, not all were like this. One large tech firm responded promptly and addressed each section of our request, and another prominent TV streaming stick vendor sent a specific response, albeit after many follow-up emails and months of waiting.

4.3.2 Data portability: copies of personal data were generally provided in mixed formats

As discussed, we eventually received copies of personal data from 23 vendors. Most data was returned in a machine-readable tabular form, though in various different file formats. Nine vendors, mostly established tech firms and TV streaming stick vendors, returned multiple directories with files in various formats: seven vendors returned one spreadsheet or table of information, six vendors returned one or more spreadsheets, one vendor sent a short JSON file and two returned data in PDF format. Figure 5 shows a screenshot from a copy of personal data from a video doorbell vendor in PDF format. PDF files (particularly those embedding images) can be problematic, as the format is not sufficiently structured to meet the requirements around data portability[63] nor should it be considered machine readable (see Figure 5).[64]

---

[63] Article 29 Working Party, Guidelines on the right to data portability (WP 242) (n 9) 14. (2017). https://ec.europa.eu/newsroom/article29/items/611233/en

[64] Ausloos, J., Mahieu, R., & Veale, M. (2019). Getting Data Subject Rights Right. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law, 10*(3), 283–309.

Section 4: Summary

This chapter considered the current state of implementation of data transparency rights in the consumer IoT, by highlighting our experiences in interacting with IoT product vendors.

§4.1 described how the process of executing our transparency rights requests was slow and laborious, required much chasing, and eventually resulted in only about two thirds of the vendors' sending some form of response relating to the substance of the requests.

We came across several factors hindering the issuing of the requests (**§4.2**), such as automated processes for making requests which limited their specificity and customisation, employees lacking relevant knowledge about data rights, delays due to vendors requesting additional information, as well as other obstacles.

§4.3 showed that even when companies did respond, the responses were often generic and incomplete (as we will further discuss), and the form in which we received responses and copies of personal data varied greatly.

In all, our experiences point to an inadequate implementation of data rights by consumer IoT vendors. In Section 5, we discuss insights we gained from the responses in relation to specific data processing practices, and compare these, where relevant, with observations from data flow monitoring.

5 Data processing observations

This section considers the data flows involved in the consumer Internet of Things. For this, we undertook traffic monitoring to observe and provide insight into the transmissions associated with product use. We then used data transparency rights to request details from vendors on the data they held and the processing practices associated with the product. This section analyses these two aspects, outlining the observations relating to the types of data that IoT vendors collect (**§5.1**), the destinations to which data flows from the product **(§5.2**), the patterns in data flows and their relation to product use (**§5.3**), and data retention considerations (**§5.4**).

5.1 Data types

We first categorise the types of data involved in the consumer IoT, informed by the privacy policies and responses received from vendors. We did this by conducting a thematic analysis, by first enumerating all types of data we saw occurring in privacy policies and rights responses, resulting in 63 data types, and then ordering and grouping these into associated categories.

This process resulted in five categories: account registration data, device details, sensor data, usage logs, and inferences (Table 8). Note that these categories naturally overlap and are not intended to be mutually exclusive, but broadly indicate different types of data involved, which may be used by vendors for different aims, and can entail different governance and sensitivity considerations.

5.1.1 Account registration data

Most IoT products required that a registration process be undertaken before the product can be used. This involved the creation of a personal account, where users typically need to provide their name and email address and set a password. In some cases, this also entailed specifying phone numbers, birth dates, gender, home address and payment details. We saw that privacy policies, where present, are generally upfront about registration data being collected. This was a common form of data returned in rights responses, with the vast majority of companies responding providing this information in their response (18 companies in total).

It is questionable, however, whether account registration is actually *necessary* for the use of many of these devices.

Resulting categorisation	Example data	Numbers of products
Account registration data	Name, email address, password, home address, weight, birth date, user ID, account creation date	18
Device details	Device ID, IP address, configurations, settings, operating system, automatic updates. Applies to the IoT product as well as the mobile device on which the app is installed	21
Sensor data	Data and values from sensors, e.g., audio, video, heartbeat, temperature, weight	4
Usage logs	Records about how the product was used, e.g., startup or shutdown, actions performed, etc.	10
Inferences	Conclusions drawn from data, user profiles and classifications	3

Table 8: Categorisation of the types of data that are returned by vendors. In total, 23 vendors responded with copies of personal data in their rights responses. The numbers indicate the number of vendors that returned data of each type

Several devices, both from established vendors, as well as from relatively unknown vendors, did not require account registration. This can reduce certain privacy risks, in the sense that if less data is collected by the company, there is less that can be (mis)used and less that can leak. Indeed, names, email addresses or other forms of directly identifiable information do facilitate profiling and data linkage. However, not collecting such identity-related information also means that other mechanisms are required to facilitate rights requests. This is because there is no account registration (i.e., typically identity-related) data that the individual can directly point to when exercising their request.[65] For two vendors that did not require account registration data, we saw that they required individuals to provide a product identifier from the app when exercising rights requests. While this reflects a privacy-by-design approach, it could become problematic if a user wants to exercise their rights when they no longer have access to the app and/or the device.[66]

5.1.2 Device details

Device details refer to aspects of the product itself, which might relate to the IoT device and/or the mobile device on which the companion app was installed. Such information includes device identifiers, firmware and software versions, as well as settings and preferences related to the device, such as time zone settings or notification preferences. It may also include other actions related to the management of the device, such as automated firmware updates and diagnostics. Insight into the type of device, settings and software can be helpful for companies for troubleshooting or security updates and can also support analytics and other processing.

Device-related data was returned by 21 companies. This commonly included device identifiers (IDs), and some of these also involved settings or preferences derived from the companion app. Only two of the copies contained any data pertaining to the mobile device, such as its model and operating system or permissions to access certain functionality (e.g., location, camera and microphone, contacts, notifications, etc). Tables 9 and 10 show examples of device specification data that was returned.

5.1.3 Information directly perceived by sensors (sensor data)

We use the term “sensor data” to refer to data that aligns to the device’s sensors that perceive aspects of the physical environment; such as audio from the microphone, video from a camera, heartbeat, temperature, weight from a scale, and so on.

Device unique number	MAC address	Device SN number	Local IP address	Internet IP address	Zone	Battery power left
7ed3b47c08 c7c0e990 f1b6edefafda1e	18:af:a1:a8:d4:b4	1	1527032000	2130706433	0	0
WIFI connection signal	**SD card status**	**SD card memory usage size**	**Detection settings**	**Timing settings**	**Detection strength setting**	**Name**
-33	4	0	{"motion_sensitivity":0,	{"motion_plan":	0	Living Room

Table 9: Device data from a baby monitor, including various settings such as “crying detection status”

alias	bindAt	deviceId	fwVer	hwVer	Mac	model	name
kitchen	2022-03-26T12:05:03.853000+00:00	8006****95B0C8	1.0.3 Build 191111 Rel.144915	4.1	1C3B****7EDC	HS100(UK)	Smart Wi-Fi Plug

Table 10: A smart plug device specification

---

[65] Norval, C., Janssen, H., Cobbe, J., & Singh, J. (2018). Reclaiming Data: Overcoming App Identification Barriers for Exercising Data Protection Rights. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 921–930. https://doi.org/10.1145/3267305.3274153

[66] Veale, M., Binns, R., & Ausloos, J. (2018). When data protection by design and data subject rights clash. *International Data Privacy Law, 8*(2), 105–123. https://doi.org/10.1093/idpl/ipy002

Only four companies returned recordings or values that appear closely related to the data obtained by sensors. This included audio recordings from a voice assistant, values on weight and body composition from a smart scale, and information obtained by a fitness tracker such as heartbeat and temperature data. It was not unexpected that fewer companies overall would return sensor data because some products have fewer sensors or are less reliant on sensors, such as light bulbs. However, given that many IoT products we surveyed are "sensor driven", we expected greater levels of response. We elaborate this discussion in **§5.3.3.**

5.1.4 Usage logs

Usage logs can be understood as a representation of events, typically as a log marked by timestamps, relating to the interactions with the user and the product, and any related occurrences. Such events include those describing any activities and usage of the product, such as opening the app or setting a timer for the lighting, device start-up and shutdown, and so on. Usage logs might also encapsulate timestamped data from sensors, such as a trigger representing movement detected from a motion sensor, or the timing when a particular inference was made. These are important as they present useful representations of what vendors are recording about product usage.

Usage logs were returned for 10 products, varying widely in the level of detail provided. In some instances, event logs were linked with locations, activities, IDs, and settings.

5.1.5 Inferences

Inferences involve attributes, conclusions, or characteristics inferred from various data sources. Inferences can be made based on sensor data or device usage information, and potentially other characteristics and analyses. Examples of inferences include classifications, data-derived profiles, and conclusions about an individual and their behaviour. Vendors can, for instance, combine account registration data with usage data to construct profiles that can be used for predicting user behaviour and preferences. During account registration, 12 apps explicitly asked for permission for collecting data for analysis, for personalisation, or both, to which we gave consent. Out of all products we explored, only three returned information directly describing inferences.

5.2 Data transmissions

IoT products involve data flows with various entities, including the vendor and other parties. In this subsection we look at the outgoing transmissions from IoT product to IP addresses, and the countries and organisations associated with these addresses.

As described in **§3.3.1** and **§5.3**, the data traffic monitoring infrastructure is set up between the device or companion app and the Internet. This means we can only observe the data flows that occur directly between the IoT product and an external IP address. After receiving data from the product, the entity behind the IP address can pass the data, along with the results of any computations or processing, to another entity – a secondary data flow. Though our monitoring infrastructure cannot observe these secondary flows, the rights to data access and portability could potentially give insight into the existence and nature of these secondary (and possibly subsequent) flows.

The following subsections present findings on the IP addresses with which products interacted, yielded by the methodology described in **§3.3.1**. Specifically, we consider:

1. The **involvement** of products with various IP addresses (**§5.2.1**);
2. The **countries** associated with these IP addresses (**§5.2.2**);
3. The **organisations** associated with these IP addresses (**§5.2.3**).

5.2.1 IP addresses

Our analyses show that IoT products communicate with a number of IP addresses. Across all of the IoT products that we tested, the mean number of IP addresses contacted per product was 88.3, with a median of 57. The total number of IP addresses contacted was 3796, of which 2966 were unique.

The smart TV streaming sticks, baby monitors, and video doorbells were the product types that tended to communicate with the highest number of IP addresses. For TV streaming sticks, this is perhaps unsurprising given that they enable content delivery from a range of sources, and therefore need to connect to various IP addresses to enable such functionality. The smart scales, fitness trackers and child smart watches involved the lowest number of IP addresses. Most of these were Bluetooth enabled devices, which interacted via the companion app. An overview of total numbers of IP addresses contacted by each interconnected device type is provided in Figure 6.

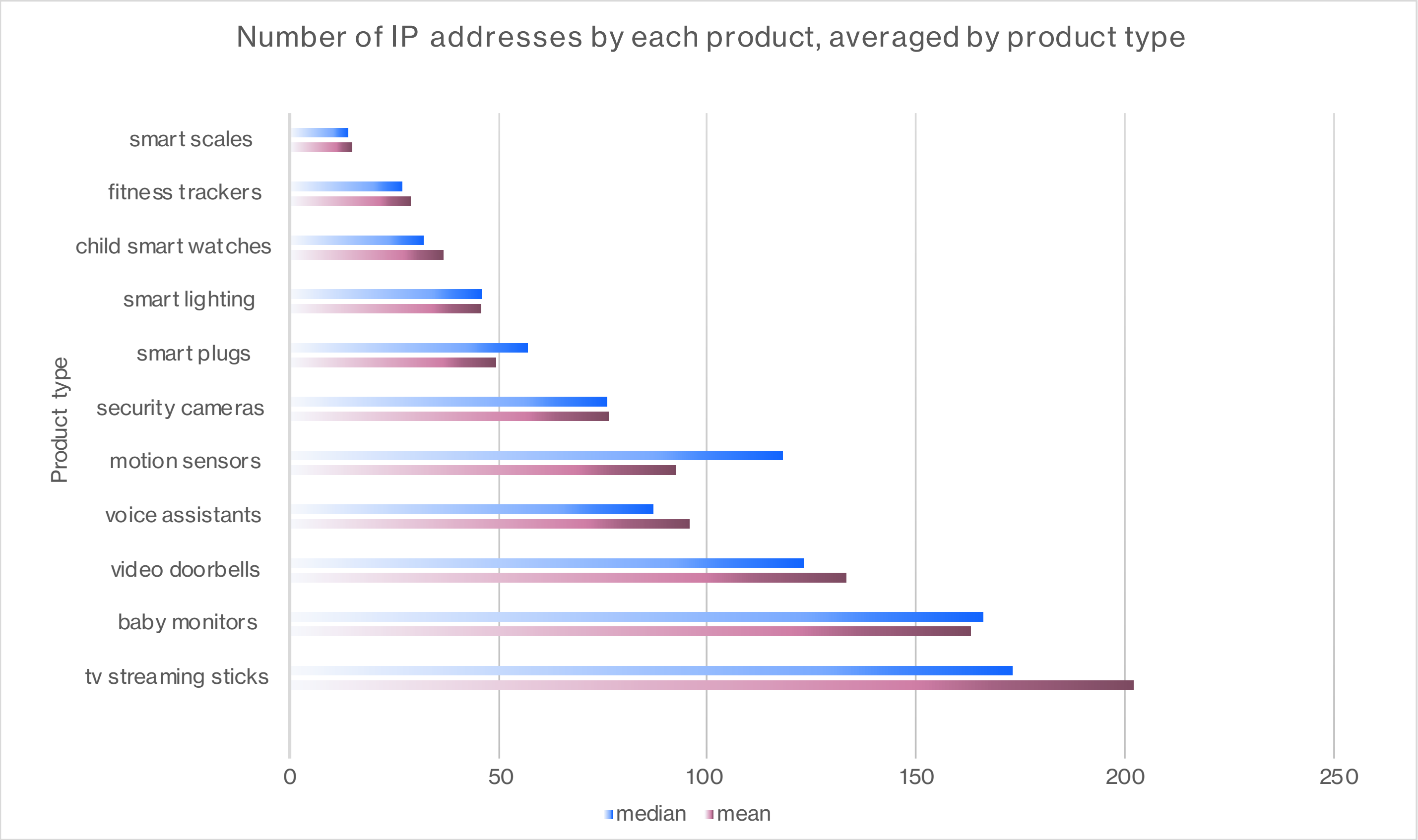


Figure 6: Total numbers of IP addresses contacted by each product, averaged by product type

In addition to the number of IP addresses with which a product interacts, we also considered the volume of data associated with (bytes transmitted to) those IP addresses, which is one factor that indicates the potential levels of involvement of an IP address with a product. Recipients of greater volumes of data will often be particularly interesting from a data protection perspective, given they likely receive more personal data from the IoT product. However, that does not imply that addresses that receive smaller amounts of data are of lower importance – there can be a low volume of highly sensitive transmissions. For example, a fitness tracker or a child smart watch may send GPS data, transmitting a relatively low volume of data which is nonetheless sensitive because it reveals the user's location.

Table 11 presents a detailed breakdown of our IP address analysis, as observed from the traffic flows generated by our product experiments (**§3.3.1**). It shows our observations about the total number of IPs with which a product interacts, the total bytes transmitted across all IP addresses over the duration of the experiments, and the total number of IPs involved with the different volumes of transmission over that period.

Table 11 shows that there is much variance in the number and volumes of data flows for each product. While there are trends – for example, products with audio or visual functionality tend to have IPs associated with higher volumes of transfers – overall, the analysis shows there are significant differences in the design and operation of these products, and therefore significant differences in volumes of data transferred and the spread of destinations.

Product	Total (count)		IP count by volume							Total (count)	
	IP addr	Volume (KB)	under 100 bytes	0.1 - 1 KB	1 - 10 KB	10 - 100 KB	100 KB - 1 MB	1 - 10 MB	over 10 MB	# Orgs	# Countries
baby monitor 1	166	136458	22	10	64	48	15	3	4	33	10
baby monitor 2	175	20216	1	28	60	60	16	10	0	31	10
baby monitor 3	229	5173	0	15	128	75	11	0	0	20	12
baby monitor 4	152	3072	9	4	101	35	3	0	0	35	14
baby monitor 5	93	14515	8	2	36	33	7	7	0	18	10
baby monitors mean	*163*	*35886.8*	*8*	*11.8*	*77.8*	*50.2*	*10.4*	*4*	*0.8*	*27.4*	*11.2*
child smart watch 1	38	748	2	2	25	7	2	0	0	14	7
child smart watch 2	25	201	1	1	18	5	0	0	0	8	9
child smart watch 3	58	1399	1	2	35	16	4	0	0	15	9
child smart watch 4	26	1316	2	5	11	4	4	0	0	6	7
child smart watches mean	*36.8*	*916*	*1.5*	*2.5*	*22.3*	*8*	*2.5*	*0*	*0*	*10.8*	*8*
fitness tracker 1	48	1625	0	2	28	15	3	0	0	9	9
fitness tracker 2	15	79	0	1	12	2	0	0	0	6	7
fitness tracker 3	14	82	1	1	9	3	0	0	0	8	8
fitness tracker 4	41	2348	0	4	28	7	1	1	0	9	6
fitness tracker 5	27	2935	0	1	15	6	5	0	0	8	6
fitness trackers mean	*29*	*1413.8*	*0.2*	*1.8*	*18.4*	*6.6*	*1.8*	*0.2*	*0*	*8*	*7.2*
motion sensor 1	120	12305	0	28	47	27	12	6	0	10	6
motion sensor 2	118	1861	1	2	92	20	3	0	0	11	9
motion sensor 3	39	560	2	5	21	11	0	0	0	6	5
motion sensors mean	*92.3*	*4908.7*	*1*	*11.7*	*53.3*	*19.3*	*5*	*2*	*0*	*9*	*6.7*
security camera 1	111	9582	3	4	52	34	18	0	0	13	8
security camera 2	42	468215	2	6	24	4	4	1	1	9	9
security camera 3	76	4087	3	7	37	20	9	0	0	17	10
security cameras mean	*76.3*	*160628*	*2.7*	*5.7*	*37.7*	*19.3*	*10.3*	*0.3*	*0.3*	*13*	*9*
smart lighting 1	24	1904	3	0	6	13	1	1	0	4	5
smart lighting 2	52	1861	0	4	28	16	4	0	0	7	6
smart lighting 3	46	746	1	3	28	13	1	0	0	8	5
smart lighting 4	61	1404	2	1	39	16	3	0	0	13	7
smart lighting 5	46	1875	0	4	25	12	5	0	0	13	7
smart lighting mean	*45.8*	*1558*	*1.2*	*2.4*	*25.2*	*14*	*2.8*	*0.2*	*0*	*9*	*6*
smart plug 1	61	639	25	6	18	11	1	0	0	21	7
smart plug 2	30	1309	1	4	14	8	3	0	0	4	8
smart plug 3	57	757	3	5	33	16	0	0	0	9	9
smart plugs mean	*49.3*	*901.7*	*9.7*	*5*	*21.7*	*11.7*	*1.3*	*0*	*0*	*11.3*	*8*
smart scale 1	10	200	1	2	3	3	1	0	0	4	5
smart scale 2	14	519	0	1	9	3	1	0	0	6	5
smart scale 3	21	202	0	3	12	6	0	0	0	8	5

smart scales mean	*15*	*307*	*0.3*	*2*	*8*	*4*	*0.7*	*0*	*0*	*6*	*5*
tv streaming stick 1	57	4764	1	2	34	11	8	1	0	9	4
tv streaming stick 2	164	5845	0	7	107	42	6	2	0	9	5
tv streaming stick 3	182	6086	0	8	102	53	19	0	0	7	6
tv streaming stick 4	405	17521	2	17	257	100	25	4	0	23	11
tv streaming sticks mean	*202*	*8554*	*0.8*	*8.5*	*125*	*51.5*	*14.5*	*1.8*	*0*	*12*	*6.5*
video doorbell 1	208	55967	1	9	106	71	9	10	2	16	9
video doorbell 2	36	102210	0	3	13	10	5	3	2	5	7
video doorbell 3	123	69170	4	6	81	24	5	2	1	19	8
video doorbell 4	69	4400	2	2	38	12	15	0	0	7	6
video doorbell 5	230	174073	8	6	100	63	19	31	3	14	10
video doorbells mean	*133.2*	*81164*	*3*	*5.2*	*67.6*	*36*	*10.6*	*9.2*	*1.6*	*12.2*	*8*
voice assistant 1	159	21474	3	9	96	21	25	5	0	7	6
voice assistant 2	87	5786	1	18	33	29	5	1	0	7	6
voice assistant 3	41	2580	0	2	25	13	0	1	0	7	6
voice assistants mean	*95.7*	*9946.7*	*1.3*	*9.7*	*51.3*	*21*	*10*	*2.3*	*0*	*7*	*6*
mean	***88.3***	***27164.4***	***2.7***	***5.9***	***47.7***	***23.2***	***6.5***	***2.1***	***0.3***	***11.9***	***7.5***
median	***57***	***2580***	***1***	***4***	***33***	***15***	***4***	***0***	***0***	***9***	***7***

Table 11: Traffic monitoring observations for the IP addresses, volumes in bytes, organisations and countries contacted by different products. In total, we observed 3796 IP addresses (of which 2966 were unique) associated with 136 organisations and 32 countries. Note this table refers only to the outgoing data from a product.

5.2.2 Countries

IP addresses are linked to physical devices, including IoT products but also computers, data centres, servers, mobile phones or any other device that is connected to a network. Data stored or otherwise processed at an IP address (including “in the cloud”), will ultimately reside at some physical, geographical location. When data is transmitted to addresses in countries outside of the EU or UK, it is possible that those countries maintain different data protection standards. This can have consequences for the safeguards and rights over the personal data held in those countries, and raise issues of legal standing and practical concerns for individuals seeking to challenge companies, states or other entities that are based in distant locales. This is a key reason for being interested in where the personal data physically goes.

From the outgoing traffic data analysis, we observed that the IoT products interact with servers in 7.5 different countries on average (with a median of seven), with a minimum of four and a maximum of 14 countries. Figure 7 provides an overview of the average number of countries contacted by each product, by product type.

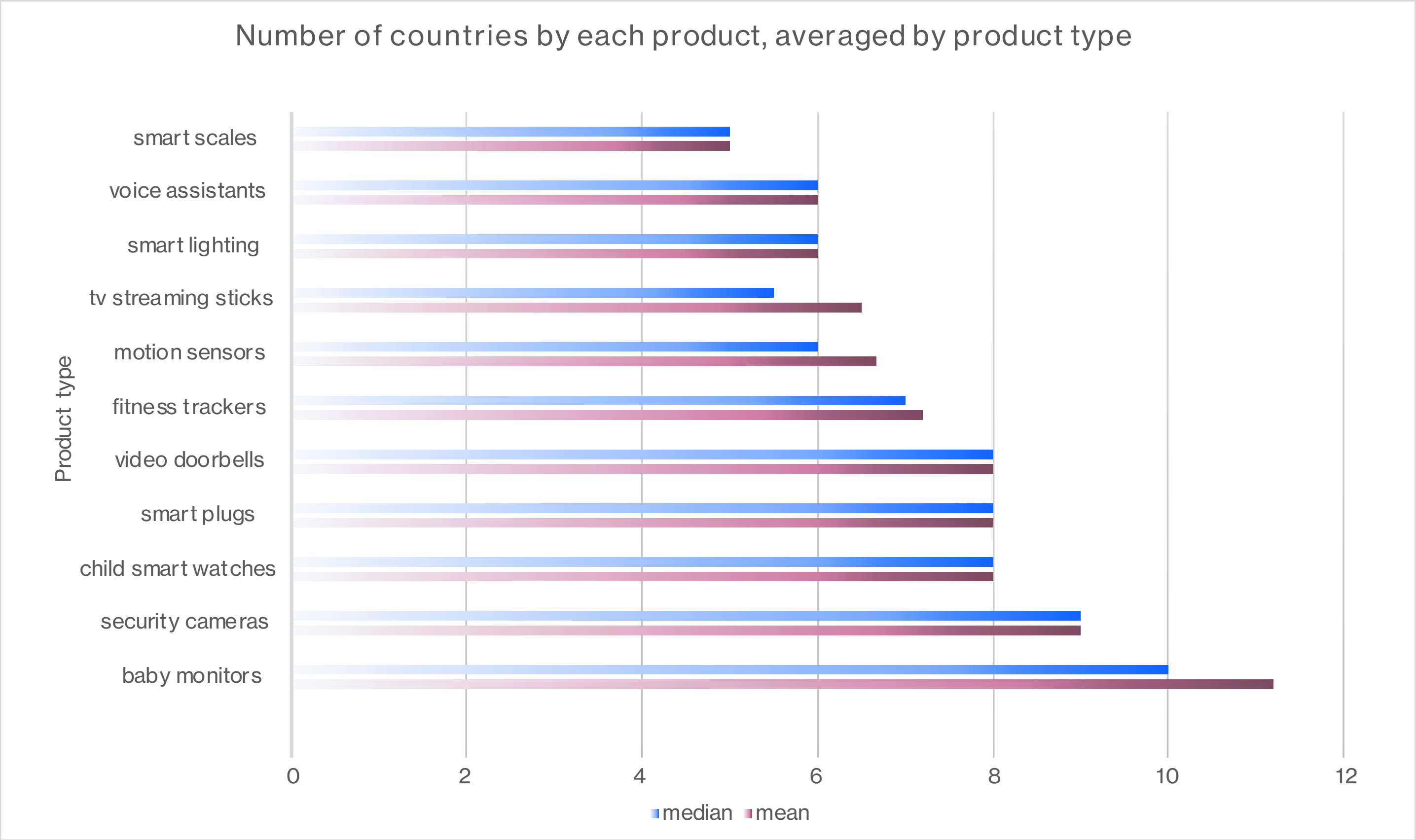


Figure 7: The numbers of countries contacted by each product, averaged by product type

Some devices involved interactions with addresses in a particularly remarkable list of countries. As examples of devices marketed for use cases involving children, a device sold as a baby monitor interacts with addresses in ten different countries, including Belgium, Canada, China, Germany, Hong Kong, Ireland, the Netherlands, Singapore, the UK, and the US. The vendor provided the identities of the actors but did not disclose their countries of operation. Another baby monitor communicated with addresses in seven different countries including Brazil, Macedonia, and Singapore. Furthermore, a child smart watch from a vendor with an EU presence established connections with China, Ireland, Japan, Russia, and the UK. Even the devices that appear to have relatively unsophisticated features can still involve data flows to a range of countries. For example, two smart plugs contact servers across six countries and three continents, with one establishing connections with Australia and New Zealand.

Considering all the products we surveyed, we observed that there were transmissions to at least 32 countries overall. The most prominent destination is the US, recipient of almost 48% of the total volume of data to 1105 IP addresses. This is followed by Ireland, the UK, and Germany, with between 10-20% of the total volume, transmitted to 300-700 addresses. In addition, The Netherlands, Japan, Hong Kong, Singapore, Finland, France, and China were recipients of 0.15-2.47% of the data at 31–280 IP addresses. Israel, Russia, Belgium, and Canada are each recipients of 0.01-0.05% of data linked with five to seven IP addresses. The remaining countries – Australia, Brazil, India, Italy, New Zealand, Sweden, and Switzerland - are all recipients of less than 0.01% of the data, linked with two to three IP addresses. Countries linked with only one IP address are South Korea, Taiwan, Turkey, and seven different EU member states.

5.2.3 Organisations

Given that IoT devices typically communicate with a number of IP addresses, we attempted to identify the organisations associated with those addresses.

Numbers and identities of organisations

We found that the IoT products contact 11.9 organisations on average (median nine; Table 11). Figure 8 shows the average number of organisations associated with each product, by product type. From different IoT device types, the highest number of organisations on average are contacted by baby monitors, security cameras, video doorbells, and TV streaming sticks.

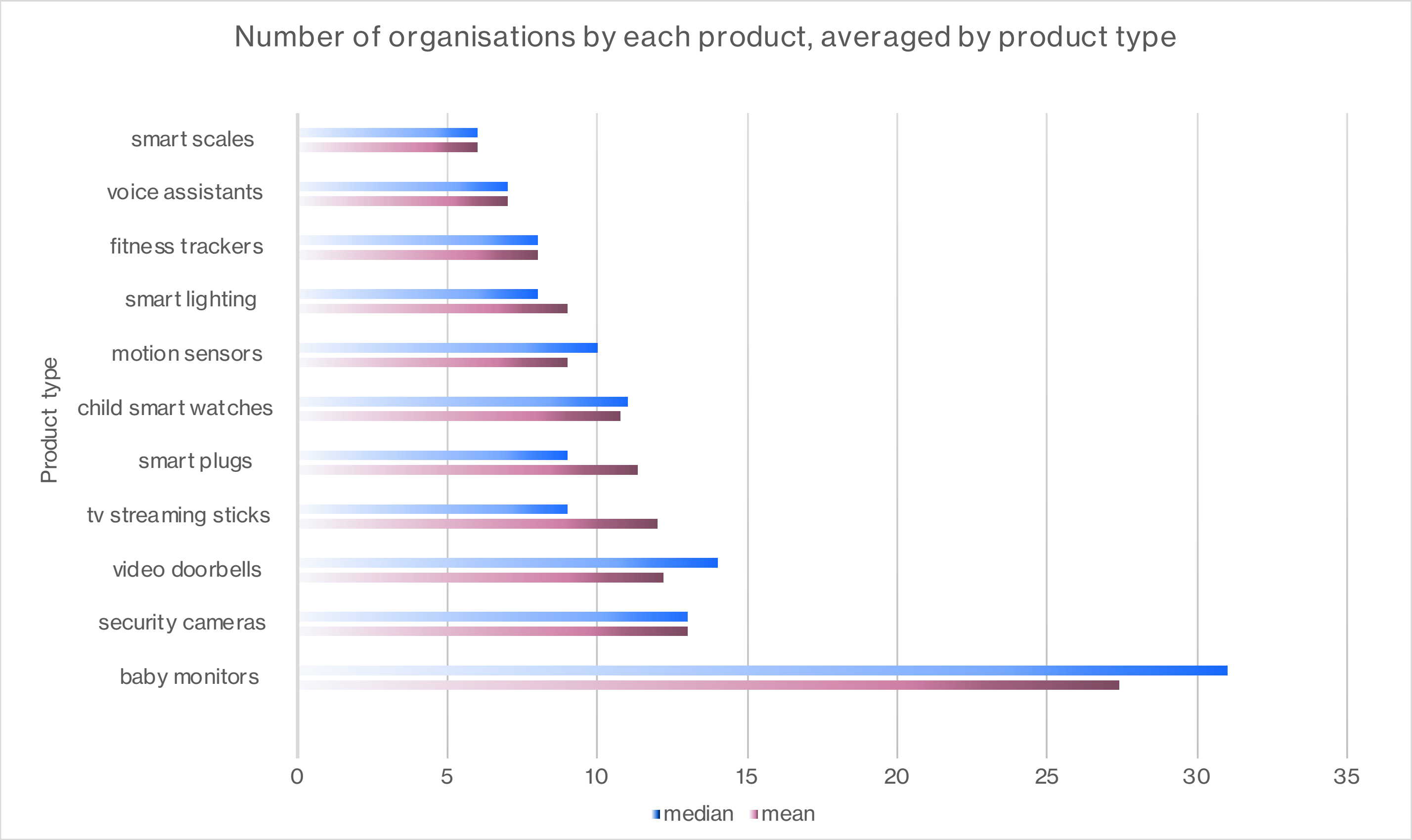


Figure 8: Average number of organisations contacted by each product, averaged by product type

Combining the IP addresses of all products that we tested, we identified a total of 136 different organisations linked to these. Recall that one organisation can be associated with multiple IP addresses. In our dataset, there are 1868 different IP addresses linked to Amazon, 217 to Google, 178 to Akamai, 58 to Cloudflare, 53 to Apple, 51 to Alibaba US, 35 to Alibaba Hangzhou Advertising, 30 to Microsoft, and 27 to Facebook. While we saw that almost half (65) of the organisations have unique IP addresses, we found that on average, each organisation can be linked with 3.3 IP addresses.

The most noticeable recipients of data by volume were Google (41.1% of the volume), Amazon (31%), Alibaba US (14.8%), Tencent (8.7%), Blackblaze (1%), Akamai (0.7%), and Cloudflare (0.4%). This is perhaps unsurprising given that most of those firms represent cloud platform and infrastructure organisations which are regularly used by a wide range of organisations for data storage, processing, etc., and because the platform we used was Android, which is owned by Google. The remaining organisations combined are the recipients of 2.3% of all the data.

Categories and roles of organisations

We next consider the potential *role* an organisation plays, based on the category of products and services that the organisation appears to deliver. This is a speculative exercise based on reconstructing information from various sources: privacy policies, responses to requests, IP address analysis, and the data flow analysis. Note that our rights requests explicitly asked about the identities of those to whom data has been disclosed, as well as what types of data they received, for what purposes, and where they are located.

Overall, 20 vendors provided *some information* about the actors involved in processing data in their rights responses, though the level of detail varied (Table 12). Of these, seven product vendors provided information in their response as to the *categories* of organisations they work with. This included, for example, a list, such as “cloud provider”, “analytics provider”, or “payment service provider”. Sometimes, categories were described in vague ways, such as “business partners”, “third parties”, or “service providers”: terms that would not leave the data subject much the wiser. Only four of the 12 vendors provided a list including the names or identities of those with whom data was, or can be, shared (see the baby monitor example at the end of **§5.2**).

	Number of products
We received some information on data sharing	20
We received information on locations	7
We received a list with the categories of organisations	7
We received a list with the names of organisations	4

Table 12: Breakdown of information received from vendors about the actors involved in data processing.

Note that even when an IP address can be linked to an organisation, there are circumstances that can introduce further uncertainties regarding its role. First, some organisations can fall under different "categories", particularly where they offer a range of services. For example, an organisation might provide cloud storage, an advertising platform, and a geographical mapping application. From the names of the organisations alone, as identified from our IP analysis, we often cannot derive the products and services that an organisation delivers in relation to the product and its data flows. Further, different parties might use similar infrastructure services, e.g., an advertising organisation and the vendor might use the same cloud provider for their service, which might make their respective roles difficult to disambiguate. Moreover, the IP analysis only provides information on the entities directly associated with the data flows, whereas flows to secondary and subsequent entities are out of sight.

Cloud providers

We observed that cloud providers are a common IoT traffic destination. Cloud services include storage or related computing services and infrastructure. As above, cloud providers that were prominent in the data retrieved in this study were Amazon, Google, Akamai Technologies, Microsoft, and Alibaba. We found that different data types may be stored by the same provider at different locations. For example, a baby monitor vendor contacted Amazon Web Services (AWS) servers in different countries. In their response, they indicated that they sent user info to AWS in Hong Kong, and live video, activity alerts, and debug logs to AWS in Germany, though they did not specify the reasons for this. This example shows servers from the same organisation serving different data types at different locations.

Advertisers

In the analysis of data traffic and privacy policies, we found some indications that advertisers, marketing organisations, analytics companies, or actors in the online advertising ecosystem are among the destinations of IoT data traffic. Privacy policies are often unclear about the specifics around data sharing with advertisers. Sixteen product vendors mention in their privacy policy that they might share data with advertisers. Half of them indicate that they may obtain data *from* advertisers. However, the nature of the relationships between the vendors and advertisers remains unclear from the privacy policies.

In the analysis of organisations that received data from the devices, we found some indications that data was sent to advertisers and analytics companies. Hangzhou Alibaba Advertising Co Ltd. represents one company that was contacted by four different IoT devices, and whose name suggests it has a role in advertising. Furthermore, several products contacted Facebook, even though, during the experiments, we never used a Facebook account to log in or otherwise connected to Facebook's social media services.-

Overall, the responses to the requests revealed little about the direct involvement of specific advertising-related actors. One exception was the TV streaming stick vendor, who confirmed in their response that they share data with advertisers. They provided us a "non-exhaustive" list of 90 advertisers whose interest-based ads may have been shown to us. This list includes large corporations such as Vodaphone, Philips, Nestle, BMW, and IKEA. Furthermore, the vendor indicated that they use pseudonymisation to provide "interest-based ads", which the vendor described as separating the names and email addresses of customers from information relating to that customer.[67] The other was a smart scale vendor, who mentioned in their response that they share data with Umeng, a mobile app analytics company, for a service they refer to as "U-FITDAYS/U-Web".

Payment service providers

Four vendors indicated in their responses that they share data with payment service providers. This is supported by privacy policies, which regularly mention payment providers in the list of organisations with whom data was shared (if these policies included third party information at all). We have seen in the traffic data that some apps sent data to Alipay, an online payment platform, even though we did not enter any payment details. PayPal was also mentioned in the response of one of the vendors. It is not clear if data was sent to PayPal during our experimentation, as we did not observe this through our traffic analysis. That said, none of our experiments involved conducting or attempting to conduct purchases after obtaining the product.

Company affiliates

Nine vendors mention in their response that company affiliates might receive data. One fitness tracker vendor, for example, indicates that the vendor can receive device details, order information, and potentially financial information in China. It also indicates that it has a business intelligence department in Finland and a business operation branch in the UK.

---

[67] The vendor's response with regards to the TV streaming stick: "We rely on pseudonymisation to serve interest-based ads while protecting our customers' identifying information. This one-way data transformation separates identifying information, such as a customer's name or email address, from information, such as audience segments associated with that customer, that we use to serve interest-based ads".

We also saw that company affiliates may have a different business focus from that of the vendors of a given product. For example, we found out that both a security camera vendor and a TV streaming stick business are owned by a large media company, which is the entity that handled our rights requests. Upon further inquiries about this relationship, the security camera vendor replied: “Limited information may be shared with [media company] where it is necessary to meet our legal and operational obligations (such as responding to your SAR/ portability request)”.

Research and education

We found some research and educational institutions among the organisations linked to the IP addresses contacted. A smart plug interacted with the University of Colorado Boulder, and another smart plug contacted three different research institutes, including Massey University in New Zealand, The University of Melbourne in Australia, and Verein zur Foerderung eines Deutschen Forschungsnetzes e.V., a German research and education institute. Furthermore, a baby monitor sent data to Gustavus Adolphus College, a liberal arts college in Minnesota; and another baby monitor contacted the Goce Delčev University of Štip in North Macedonia. Finally, a video doorbell interacted with the National Physical Laboratory, the national measurement standards institute in the UK. We did not receive any responses that shed light on these relationships.

Other organisations

Other categories of organisations that we observed in the responses include customer service providers, business analytics, and mapping or navigation platforms. Four vendors indicated that they share data with customer service providers. It is unclear whether these were company affiliates or external companies.

Moreover, from the various privacy policies, other third parties with whom data might be shared included delivery services, financial services, regulatory and legal services, credit reference agencies, warehouse and storage departments, retail outlets, telephone IT service providers, internal auditors, and system maintenance organisations.

Example: a baby monitor

One product marketed as a baby monitor contacted a particularly high number of destinations. The vendor was helpful in offering more information on the destinations. The table they provided is presented below.

The table shows that the baby monitor’s data can flow to one or more cloud platforms, payment service providers, or social media platforms; an email marketing provider; and three customer service providers. We see that data can flow to a number of larger technology companies, including Tencent, Amazon, PayPal, and Google.

Such an overview is helpful in identifying the types of actors and the purposes for which actors are contacted. However, while the vendor provided information about 10 “third parties”, our traffic monitoring suggests there may be 29 different organisations involved in the transmissions from just the device alone, before even considering the companion app.

Purpose	Third party	Related privacy policy
To provide the verification service for account registration	Tencent Cloud Computing (Beijing) Co., Ltd.	https://intl.cloud.tencent.com/document/product/301/17345
To provide the [product’s] Cloudplay Service	Amazon Web Services, Inc.	https://aws.amazon.com/privacy/?nc1=f_pr
To complete the online payment for subscribing to [the product’s] Cloudplay Service per the choice of [the product] User	PayPal Pte. Ltd.	https://www.paypal.com/c2/webapps/mpp/ua/privacy-full?locale.x=en_C2#PayPal
To complete the online payment for subscribing to [the product’s] Cloudplay Service per the choice of [the product] User	Stripe, Inc.	https://stripe.com/zh-cn-us/unsupported-browser?location=%2Fzh-cn-us%2Fprivacy
To provide message push service for applicable Huawei mobile phone only	Huawei Software Technology Co., Ltd.	https://developer.huawei.com/consumer/cn/doc/development/HMSCore-Guides/sdk-data-security-0000001050042177
To support the connection of the [product] account and your social media account per your choice	Google LLC	https://policies.google.cn/privacy#enforcement
To send the commercial emails to users with prior consent	Benchmark Email	https://www.benchmarkemail.com/privacy-policy/
To provide online customer service to users per the choice of customer	Guangzhou Yunqu-info technology Co., Ltd.	http://www.yunqu-info.com/about.html
	XTRASOURCE EUROPE B.V.	https://webhelp.com/wp-content/uploads/2020/12/EN_GPPrivPol-01-Privacy-Policy-20201125.pdf
	BPOnovo Inc.	https://callnovo.com/privacy-policy/

5.3 Patterns in data flows

Our experiments involved us interacting with, or “using”, the IoT products in line with the methodology described in **§3.3.1**. In the following subsections, we describe our observations about the transmissions by IoT products; outline the types of data flows and the relationships of data flows to device usage; and relate our observations to the transparency rights responses that we received from vendors.

5.3.1 Types of monitored flows

To reiterate (see **§2.1.1**), the distinction between IoT products, IoT devices, and companion apps is important for our analyses. An IoT product most often consists of a IoT device and a companion app. The device is the physical object, such as a light bulb or a fitness tracker. The app runs on a tablet or smartphone and is used to communicate with the IoT device.

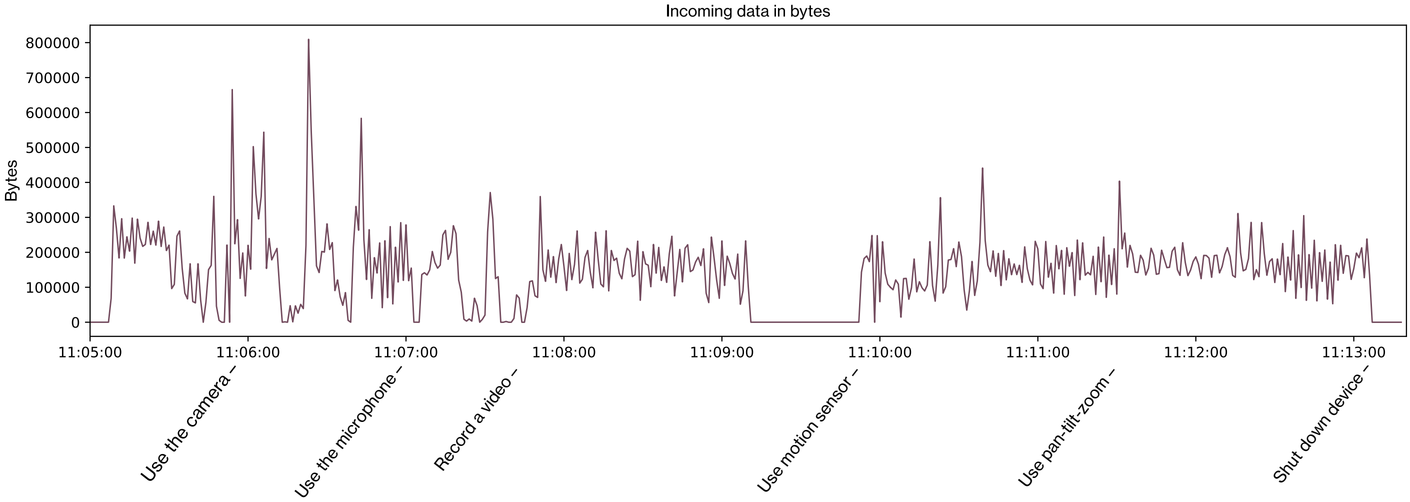
Data flows may occur directly between the app and the device. They can also occur from a device directly with an IP address on the Internet, or from the companion app to the Internet. We describe communication as being *external* if it is with an IP address other than those related to the IoT device or the companion app. This means that “external” addresses relate to the vendor *or* some other party.

Our monitoring focuses on the direct flows relating to an IoT device, its companion app, and external IP addresses. Specifically, we explore three core types of data flows, elaborated below.

Note that all of these flows entail direct transmissions (through a direct communication channel) and are of the type we are able to observe with our monitoring infrastructure (see **§3.3.1**). Again, we are unable to directly monitor Bluetooth exchanges, nor can we capture flows that occur beyond the realm of the device or app, i.e., those communications between external actors which do not directly involve the device or app.

Flow 1: Direct communication between the IoT device and companion app
Devices typically involve some direct communication with the companion app. In practice, such flows can occur for both Internet-enabled and purely Bluetooth-enabled product types. These flows are the norm for Bluetooth devices because they do not have direct Internet connectivity.[68]

Figure 9 presents an example showing the clear presence of direct interactions between a baby monitor and its companion app. The graph in purple visualises the outgoing data transmission from the (Internet-enabled) IoT device to a companion app, and the blue graph visualises the incoming data flow received by the app from the device, where we see that the timelines and sequences of events are matched. For this particular device, a baby monitor, around 97.5% of the data transmission happens directly between the app and device. Of this data transmission, 97% flows from the IoT device to the app, and 3% flows from the app to the device. This suggests that live observations obtained by sensors from the baby monitor were sent directly to the companion app.



---

[68] Note we do not directly monitor transmissions via Bluetooth. For Bluetooth devices, we do monitor the transmissions from the device’s associated companion app, which is the only means by which the device can leverage the Internet and transmit data externally (see Flow 3, below).

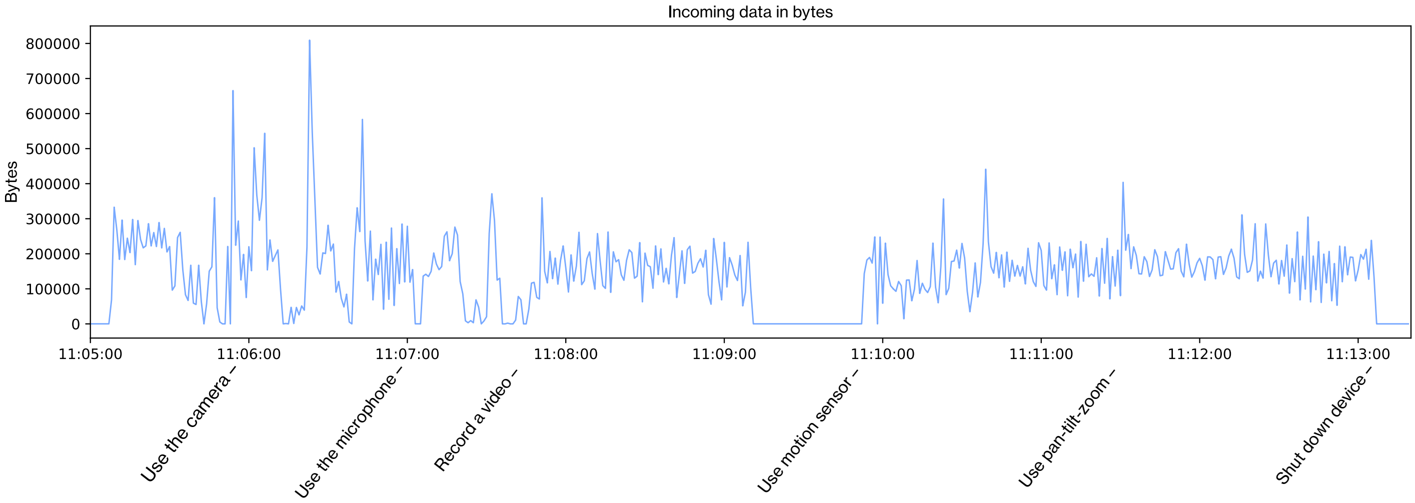


Figure 9: The graph on the bottom of p. 45 in purple shows the outgoing data from the baby monitor directly to the companion app. The graph in blue shows the incoming data from the baby monitor directly to the companion app. In combination, the graphs illustrate transmissions directly between the app and device in which timelines and events appear to align

Flow 2: Direct communication between the IoT device and external actors

Data can be transmitted directly between the IoT device and external actors via the Internet. This was the case for all of those devices that are Internet-enabled, which comprised 33 of the IoT products in our study.[69] Figure 10 presents an example illustrating the data transmission between an IoT device and external IP addresses, showing the external transmissions of a TV streaming stick in bytes per second.

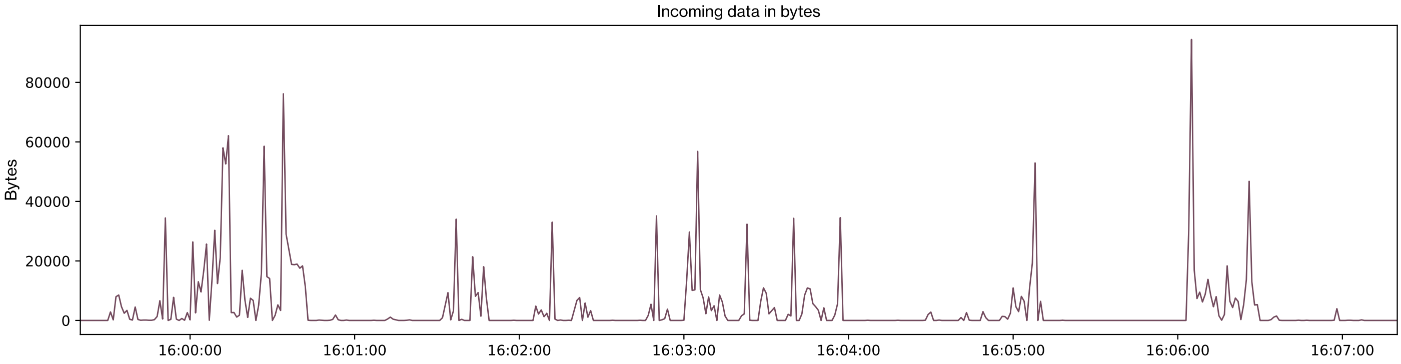


Figure 10: The data transmissions in bytes from a TV streaming stick to external actors

---

[69] Note that those that do not directly communicate with external actors were mostly Bluetooth devices, which cannot directly access the Internet, but instead transmit data to the companion app (see below).

Flow 3: Direct communication between the companion app and external actors
The companion apps, which operate on mobile devices (e.g., a phone or tablet), can also directly transmit data externally. As apps are a common feature of IoT products, generally this means transmissions from the app can relate to data about either a Bluetooth or Internet-enabled device. We observed that for all products, data flows from apps to external actors (**§5.3.2**),

The three flows we highlight are those which we can directly observe from our experimental infrastructure. Importantly, these form the basis for broader interactions, and in practice, IoT products can employ a combination of these types of flow. Note that when we observe a combination of data flows, for example when data flows first from the device to the companion app (Flow 2), and then from the app to external actors (Flow 3), the nature or content of the data flows are not necessarily the same. On the app, data may have been aggregated, summaries may have been made, or transmissions may relate to or incorporate other data. As a result, any given sequence of flows does not necessarily mean that subsequent flows are the same in volume, or occur over a similar time-frame. Such uncertainties further emphasise the need for vendor transparency about the involvement of other actors in these data flows.

5.3.2 Transmissions and device usage

By analysing the data flows that we captured, we were able to observe certain patterns of transmission indicative of the interactions between the user and the device. This is useful as it can suggest where and how certain events were communicated.

We found that, in many cases, we could observe direct mappings between the timings of the interactions, as recorded in our experimentation logs, with particular patterns in data transmission. We illustrate this with the following examples that show external data transmissions (over the Internet) that appear to reflect device usage.

Figure 11 shows the data transmission in bytes per second during interactions with a video doorbell. "Video storage" and "people detection" were enabled by default on this device. We see that data transmission is low when there are no user interactions with the device, and increases when the motion sensor, camera, microphone or app are used. This means that data about the usage of the device appears to be reflected in external transmissions. Note that this vendor did not, in their response, return any sensor (i.e., video, sound, etc) data or usage logs, but only general device information.

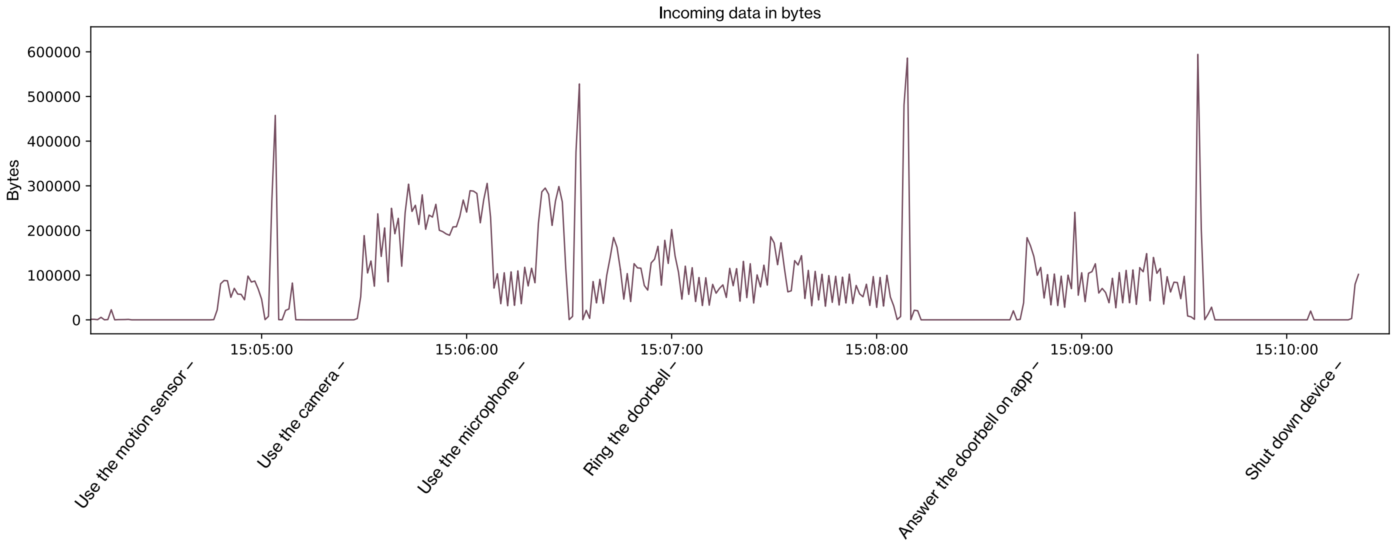


Figure 11: The data transmissions to external actors during interactions with a video doorbell

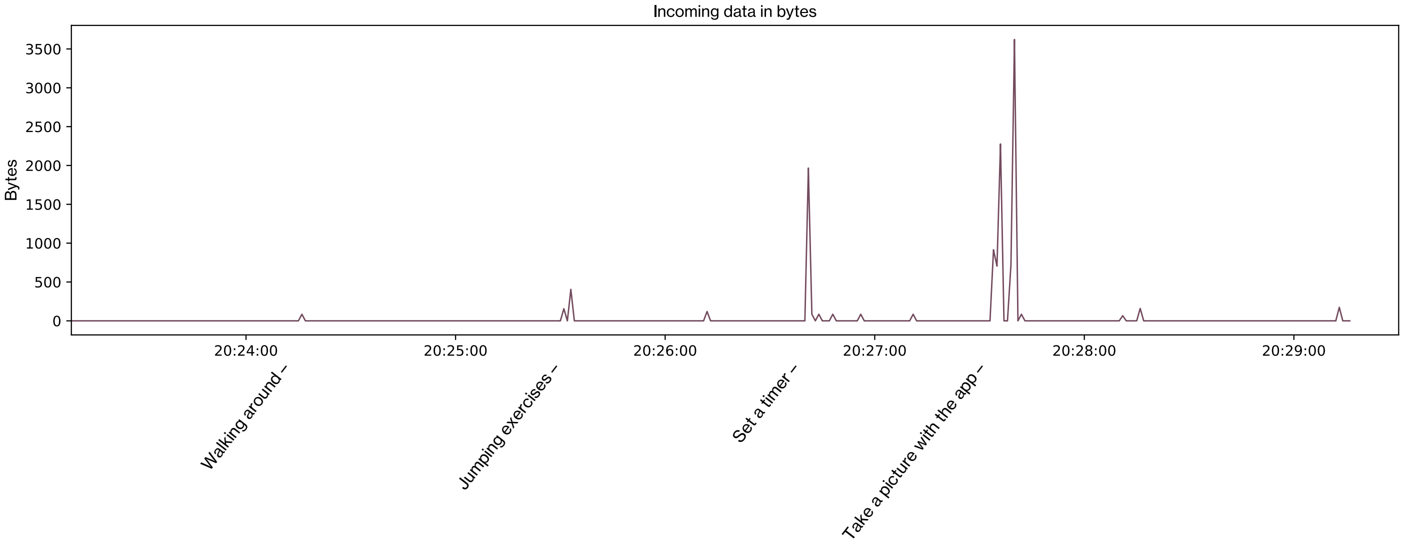


Figure 12: External data transmissions from the companion app of a fitness tracker

Figure 12 shows data transmission from the companion app connected to its fitness tracker via Bluetooth, meaning the fitness tracker's only connection to the Internet was via the app. The figure shows that data transmission from the companion app increases when the individual walks around, performs jumping exercises, or controls the device with the app. This illustrates that the fitness tracker sends information to the app via Bluetooth when the device is used, and suggests that the app informs actors about the nature of the user interactions and their activities.

Importantly, the monitoring of data flows can potentially indicate some instances where certain processing occurs. As an example, in monitoring the data from security cameras, we observed that data transmission rates were higher when we showed our face compared to when we showed a hand (see Figure 13), which indicates that facial recognition may have been activated.[70]

5.3.3 Responses and data flows

The previous s subsections show that IoT products transmit data externally. We now further explore the nature of the data transmitted. We combine and compare the observations from our data flow monitoring with the rights responses we received from the vendors to help to build a better understanding of the data transmitted, and how data rights responses can help to support this understanding. Given that the nature of transmissions will differ depending on the product and its functionality, we consider each product type in turn.

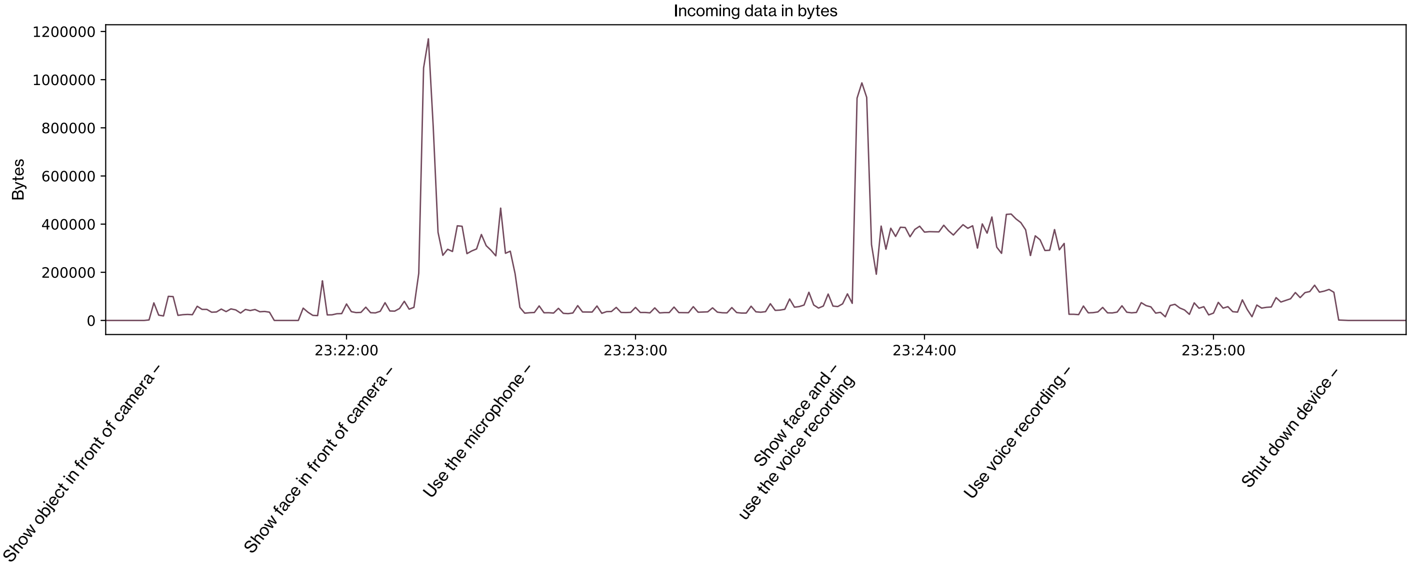


Figure 13: Data transmissions from a security camera to external actors during use

---

[70] Facial recognition is a feature this product supports and actively markets.

TV streaming sticks

We considered four TV streaming sticks, of which only one came with a companion app. The TV streaming sticks allow users to consume content, such as watching shows or listening to music. The data flow analyses of these devices show peaks in data transmission from the TV stick to external actors while the device was being used (selecting content to consume).

While all TV streaming stick vendors responded to our rights requests, their responses varied in level of detail. Two vendors returned streaming records in their rights responses. These typically included the start time, end time, and duration of content consumption; whether streaming was successful; and, in some cases, the name of the media being streamed. An overview of these data flows from one TV streaming stick is shown in Figure 14. Table 13 presents an overview of viewing data as returned from that vendor, demonstrating (unsurprisingly) that the use of this TV stick led to external actors' obtaining details about user consumption of content.

A third TV streaming stick vendor returned empty files with "viewing_history" and "voice_history", suggesting that they have systems in place to collect such data. Given that the data flow analysis shows similar flows from the device to other vendors mentioned above (Figure 15), finding the files to be empty was not something we would have expected. The vendor of a fourth TV streaming stick did not return any usage data, despite our directly observing the device externally transmitting data during usage.

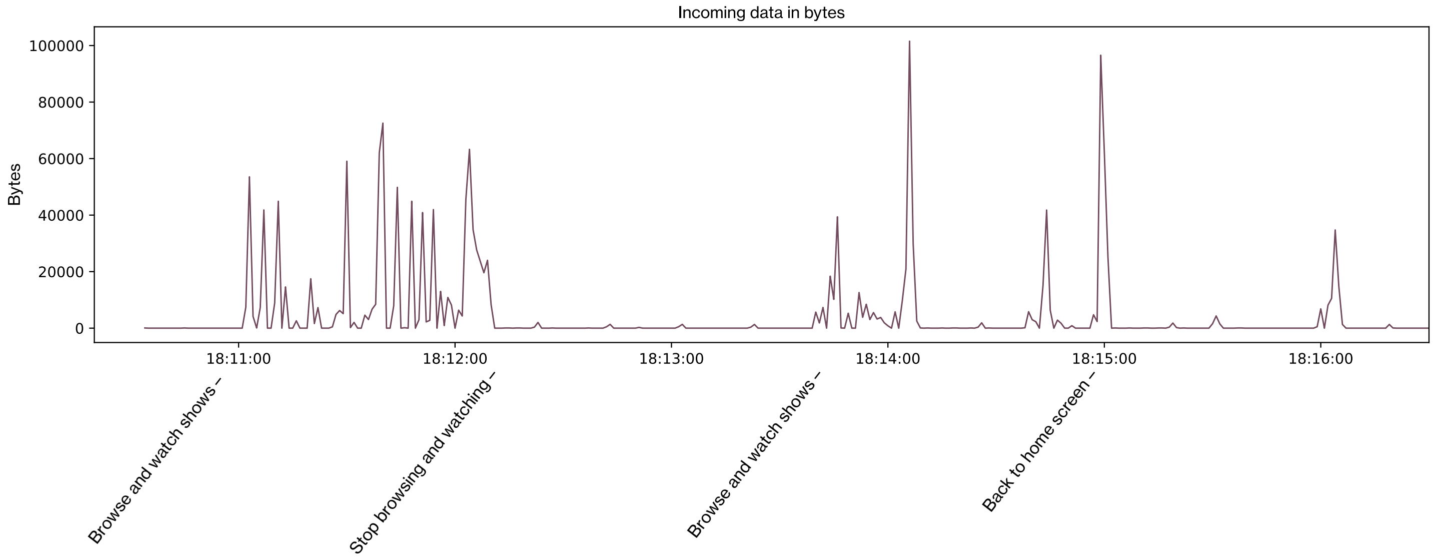


Figure 14: Data transmissions from the TV streaming stick to external actors during interactions with the device

Start TimeShow actions	End TimeShow actions	Asset NameShow actions
26 March 2022 – 17:11:35	26 March 2022 – 17:11:35	Live-1409-Sky Cinema Premiere
26 March 2022 – 17:12:03	26 March 2022 – 17:13:39	F1 Start of Season 2022
26 March 2022 – 17:14:07	26 March 2022 – 17:14:39	Movie Box Sets trailer
26 March 2022 – 17:15:00	26 March 2022 – 17:15:31	DC Heroes trailer

Table 13: An excerpt from the event log of a TV streaming stick, as returned by the vendor in response to our data rights request

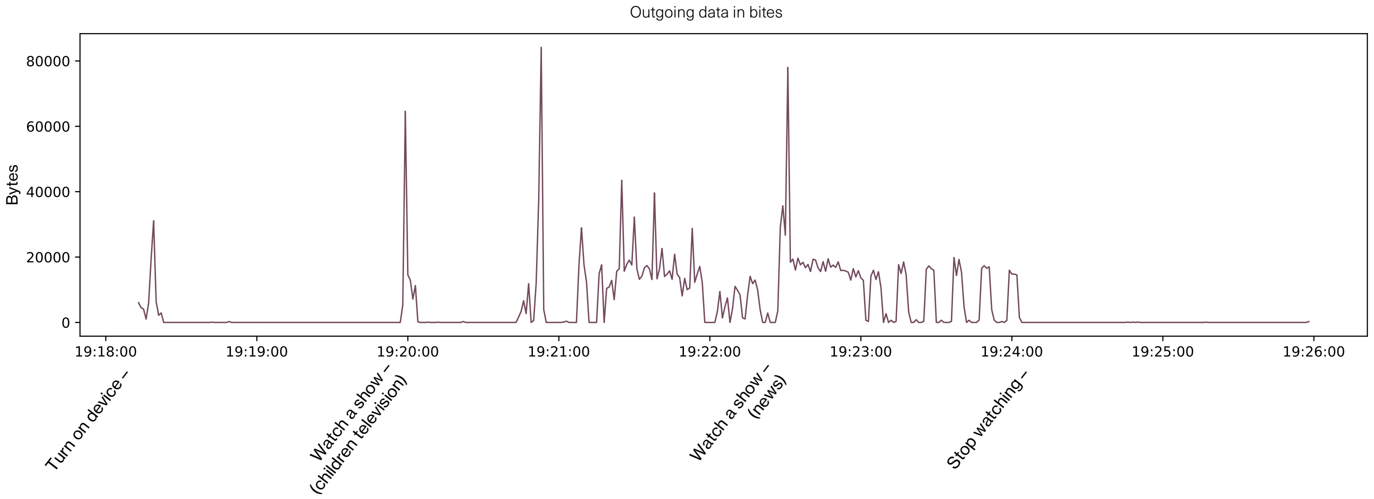


Figure 15: Data transmission from a TV streaming stick to external actors

Voice assistants

This study involved voice assistants from three different vendors. We see that the collection of audio from the microphone is reflected in the data flow analysis (Figure 16), where noticeable external data transmissions occur while the user is giving voice commands, and smaller volumes are transmitted from the device to the app. The vendor returned audio files and command transcriptions provided to the voice assistant using the relevant activating word (Table 14).

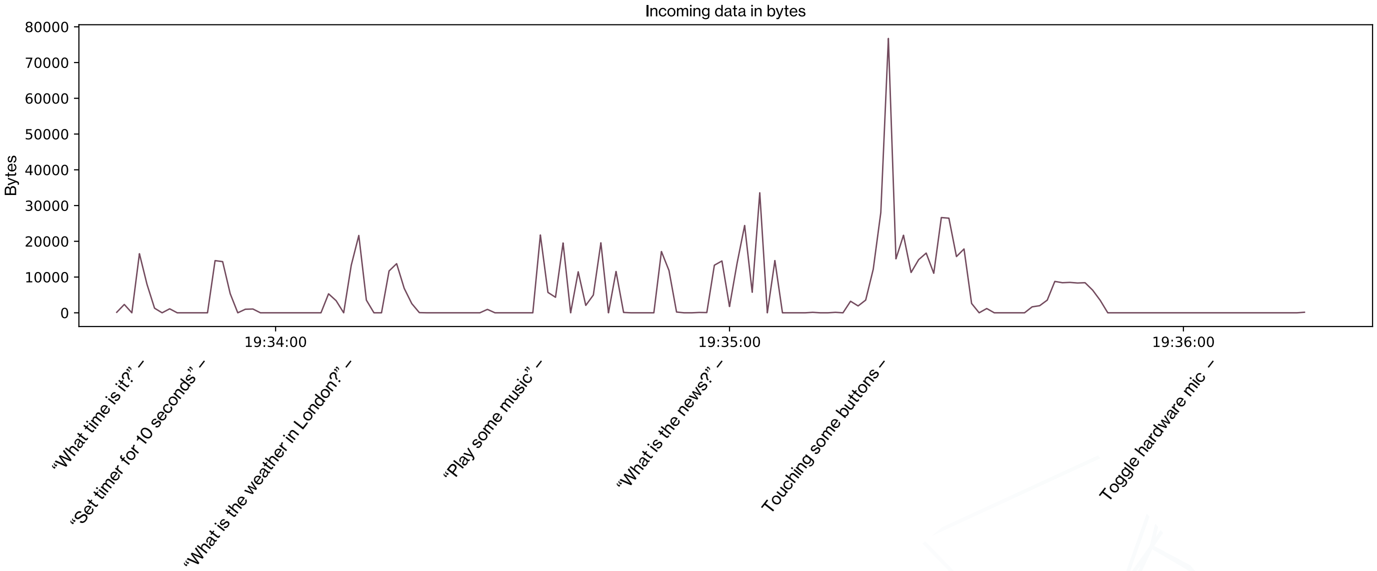


Figure 16: External data transmissions from a voice assistant

Timestamp (UTC)	Transcription	File	Response
2022-03-04T18:33:43.793Z	what time is it	4fd3ef2e42a3e87235c89 1be6140a640258542d2.wav	The time is 6:33 pm.
2022-03-04T18:33:54.442Z	set timer for ten seconds	f2f35a357c21c55da2 754a0e0514bbcec41b525b.wav	10 seconds, starting now
2022-03-04T18:34:12.123Z	stop	48dfa3e8fe5ff882e0e1ab6e a20374c039f17dc4.wav	Not Applicable
2022-03-04T18:34:17.823Z	what's the weather in London	9ca7a70031165eb09523 a9b7a29134538be26ff8.wav	In London United Kingdom, it's 45 degrees Fahrenheit with showers. Tonight, you can expect rainy weather, with a low of 40 degrees.
2022-03-04T18:34:36.514Z	play some music	40270063206aeace46f04 7ec58502408bdd19985.wav	Playing your personalised station, My Soundtrack, from [vendor].
2022-03-04T18:34:59.992Z	what's the news	6ffbef73427ecca36412 ec9e8510b55d2adb2d9c.wav	Not Applicable
2022-03-04T18:34:52.700Z	stop	e72309101fd6c577c88e65 31e0199855770a4e6a.wav	Not Applicable
2022-03-04T18:35:48.385Z	Data Not Available	159a6459d67e4730d07 7d3061e4b7a58eea3fbba.wav	Not Applicable

Table 14: Selection from responses that were returned by the voice assistant. The file was accompanied by audio fragments of the voice of the user giving the commands

Furthermore, a second voice assistant vendor also returned transcriptions and audio files of what the user said to the speaker (Figure 18). We did observe, as per Figure 17, that around 16:31:41 there is a peak in transmission that did not align with our records of experimental interactions. In the usage logs, the vendor did indicate that the assistant was used around this time, but without providing further details (see bottom of Figure 18). All entries in the usage log mentioned that "this activity was saved to your [product] Account because the following settings were on: Web & App Activity". According to the website of the vendor, Web & App Activity saves activity on the vendor's apps, including location data, to provide better recommendations and more personalised experiences. Users have the option to opt out of saving their activity (though we did not). The usage logs suggest that the speaker communicates with (at least) the vendor, even when the user does not use the assistant's "activation word", though such transmissions may not necessarily involve audio data. In short, the usage logs demonstrate that data was sent around 16:31:41, but do not indicate what the event was that happened at that time.

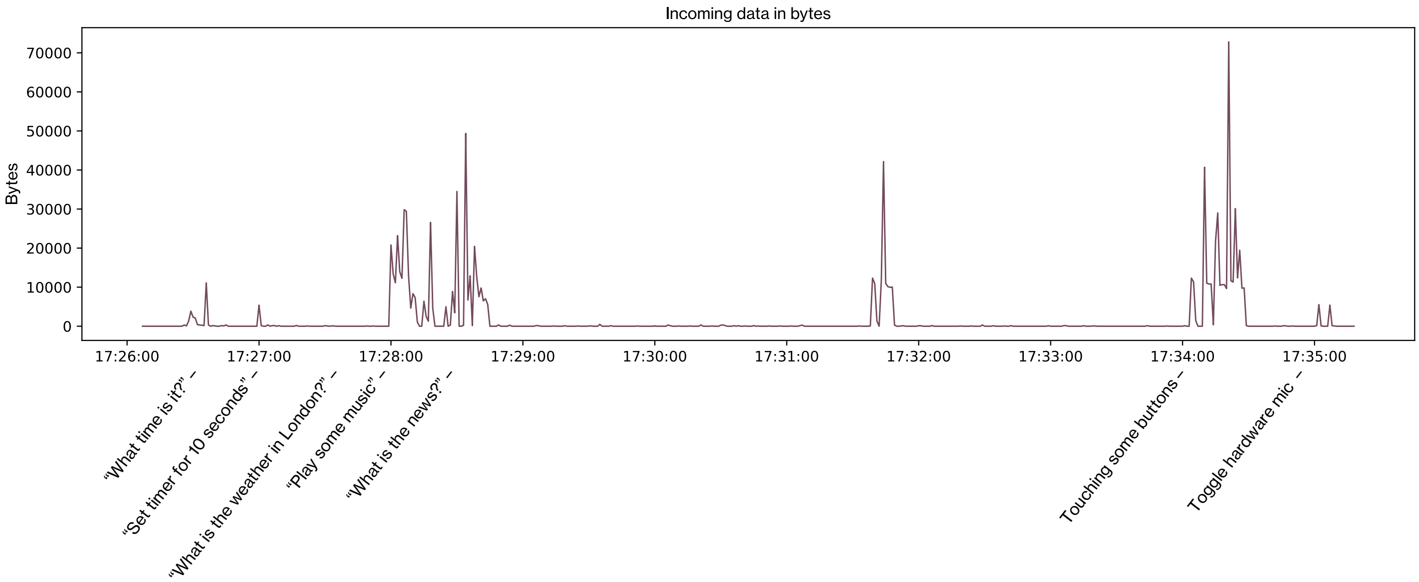
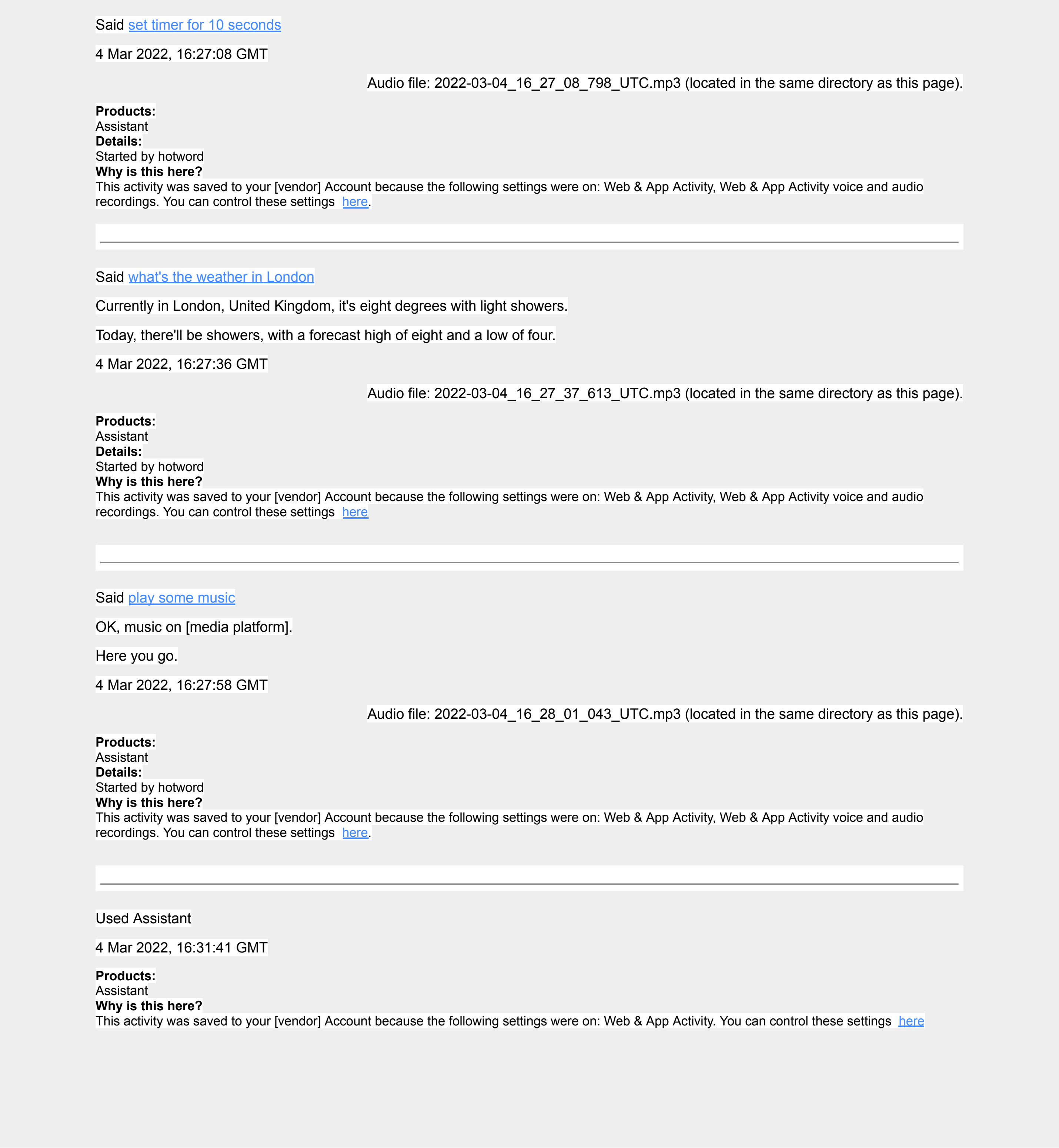


Figure 17: Data transmissions to external actors during interactions with a voice assistant



Said set timer for 10 seconds

4 Mar 2022, 16:27:08 GMT

Audio file: 2022-03-04_16_27_08_798_UTC.mp3 (located in the same directory as this page).

Products:
Assistant
Details:
Started by hotword
Why is this here?
This activity was saved to your [vendor] Account because the following settings were on: Web & App Activity, Web & App Activity voice and audio recordings. You can control these settings here.

Said what's the weather in London

Currently in London, United Kingdom, it's eight degrees with light showers.

Today, there'll be showers, with a forecast high of eight and a low of four.

4 Mar 2022, 16:27:36 GMT

Audio file: 2022-03-04_16_27_37_613_UTC.mp3 (located in the same directory as this page).

Products:
Assistant
Details:
Started by hotword
Why is this here?
This activity was saved to your [vendor] Account because the following settings were on: Web & App Activity, Web & App Activity voice and audio recordings. You can control these settings here

Said play some music

OK, music on [media platform].

Here you go.

4 Mar 2022, 16:27:58 GMT

Audio file: 2022-03-04_16_28_01_043_UTC.mp3 (located in the same directory as this page).

Products:
Assistant
Details:
Started by hotword
Why is this here?
This activity was saved to your [vendor] Account because the following settings were on: Web & App Activity, Web & App Activity voice and audio recordings. You can control these settings here.

Used Assistant

4 Mar 2022, 16:31:41 GMT

Products:
Assistant
Why is this here?
This activity was saved to your [vendor] Account because the following settings were on: Web & App Activity. You can control these settings here

Figure 18: A selection of transcriptions, activity data and audio files (where relevant) including the user's voice, returned by a voice assistant vendor

The other voice assistant vendor did not return any audio transcriptions or usage data, even though the data flow analysis shows similar peaks in transmission during usage, rather in line with that which we saw in Figures 16 and 17.

Fitness trackers

We experimented with five (Bluetooth enabled) fitness trackers, all of which showed some peaks in transmission during interactions with the device or with the app. From one response, we see data returned that includes overviews of stress levels, activity levels, menstrual health, and sleep score based on the sensory input. Figure 19 and Table 15 below show an overview of our actions as recorded from our experiments (Figure 19), as well as what the vendor returned (Table 15). This shows that the data obtained or inferred by the fitness trackers is known by at least the vendor.

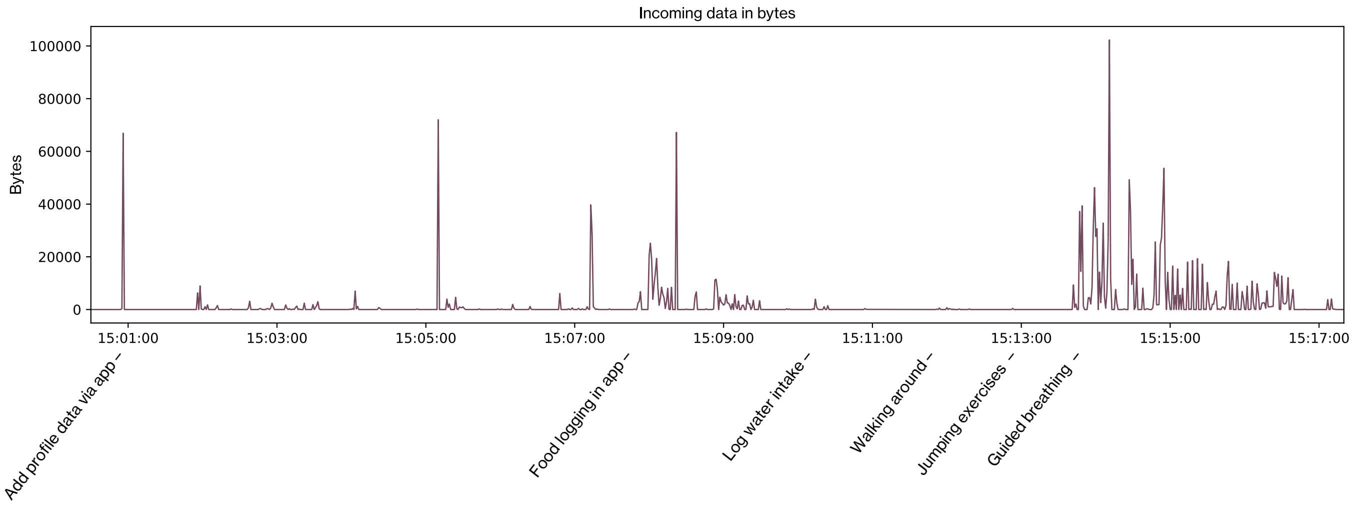


Figure 19: External data transmissions from the companion app of a fitness tracker

<table>
<tr><th colspan="3">What the vendor returned</th></tr>
<tr><th colspan="2">Profile data</th><th>Name, height, weight, birth date, profile picture</th></tr>
<tr><td colspan="2">food_logs</td><td><pre> "logDate" : "2022-03-20",
  "favorite" : false
},{
  "logId" : 28764798612,
  "loggedFood" : {
    "foodId" : 0,
    "name" : "Quick Calories",</pre></td></tr>
<tr><td colspan="2">Calories</td><td><pre>"dateTime" : "03/20/22 14:07:00",
"value" : "1.03"</pre></td></tr>
<tr><td colspan="2">water_logs</td><td><pre>"id" : 8944589841,
"date" : "03/20/22",
"waterAmount" : 500.0,
"measurementUnit" : "MILLILITER"</pre></td></tr>
<tr><td colspan="2">steps</td><td><pre>"dateTime" : "03/20/22 14:11:00",
"value" : "76"</pre></td></tr>
<tr><td colspan="2">steps</td><td><pre>"dateTime" : "03/20/22 14:12:00",
"value" : "49"</pre></td></tr>
<tr><td rowspan="16">Guided breathing</td><td>session_id</td><td>18d0faa0-a858-11ec-8080-808080808080</td></tr>
<tr><td>activity_name</td><td>2-Minute Stress Reliever</td></tr>
<tr><td>average_heart_rate</td><td>null</td></tr>
<tr><td>start_heart_rate</td><td>null</td></tr>
<tr><td>end_heart_rate</td><td>null</td></tr>
<tr><td>duration</td><td>156995</td></tr>
<tr><td>start_date_time</td><td>2022-03-20T14:14:46</td></tr>
<tr><td>end_date_time</td><td>2022-03-20T14:17:29</td></tr>
<tr><td>sleep_log_entry_id</td><td>3.6181E+10</td></tr>
<tr><td>timestamp</td><td>2022-03-21T07:45:30Z</td></tr>
<tr><td>overall_score</td><td>76</td></tr>
<tr><td>composition_score</td><td>17</td></tr>
<tr><td>revitalization_score</td><td>20</td></tr>
<tr><td>duration_score</td><td>39</td></tr>
<tr><td>deep_sleep_in_minutes</td><td>55</td></tr>
<tr><td>resting_heart_rate</td><td>60</td></tr>
<tr><td></td><td>restlessness</td><td>0.0810081</td></tr>
</table>

Table 15: The data returned by a fitness tracker vendor, that relate to the data flows shown in Figure 19 above

Other fitness trackers show similar transmission patterns in response to activities, which suggests similar communications of user-related activities or inferences are being made to external actors. Two of the fitness tracker companies confirmed in their responses that they do collect health data but did not return details of the data that they might have gathered, even though we explicitly asked for this. The remaining two fitness tracker companies did not respond at all.

Smart scales

Of the three smart scales we considered, one vendor returned weight and body composition linked to time of measurement, which is shown in Table 16. The values on date 3/3/22 appear in line with our experiment done on the same day (Figure 20), as derived from a pressure sensor for measuring weight, as well as sensors perceiving an electrical current that helps to calculate body fat and muscle mass. Though the vendor promotes and markets these features, their response shows that these measurements are known by at least the vendor.[71]

The two other smart scales demonstrate similar sorts of data flow during the same experiments. One of those confirmed that they share "body related data such as birthday, height and gender submitted" and use this for "the calculation of body fat rate / muscle mass / basal metabolic rate". However, they did not return this data in their response. The other smart scale vendor indicated that the personal data is stored in the app and that they do not have access to it without our permission. Although we explicitly asked for data access in the request, they responded with instructions for us to access the data via the app.

Time of Measurement	**Weight (kg)**	**BMI**	**Body Fat (%)**	**Fat-free Body Weight (kg)**	**Subcutaneous Fat (%)**	**Visceral Fat**
3/3/22 23:26	69.3	19.8	15.2	58.8	13.9	3
Body Water (%)	**Skeletal Muscle (%)**	**Muscle Mass (kg)**	**Bone Mass (kg)**	**Protein (%)**	**BMR (kcal)**	**Metabolic Age**
61.3	54.8	55.8	2.94	19.3	/	30

Table 16: Elements from the copy of personal data returned by a smart scale vendor

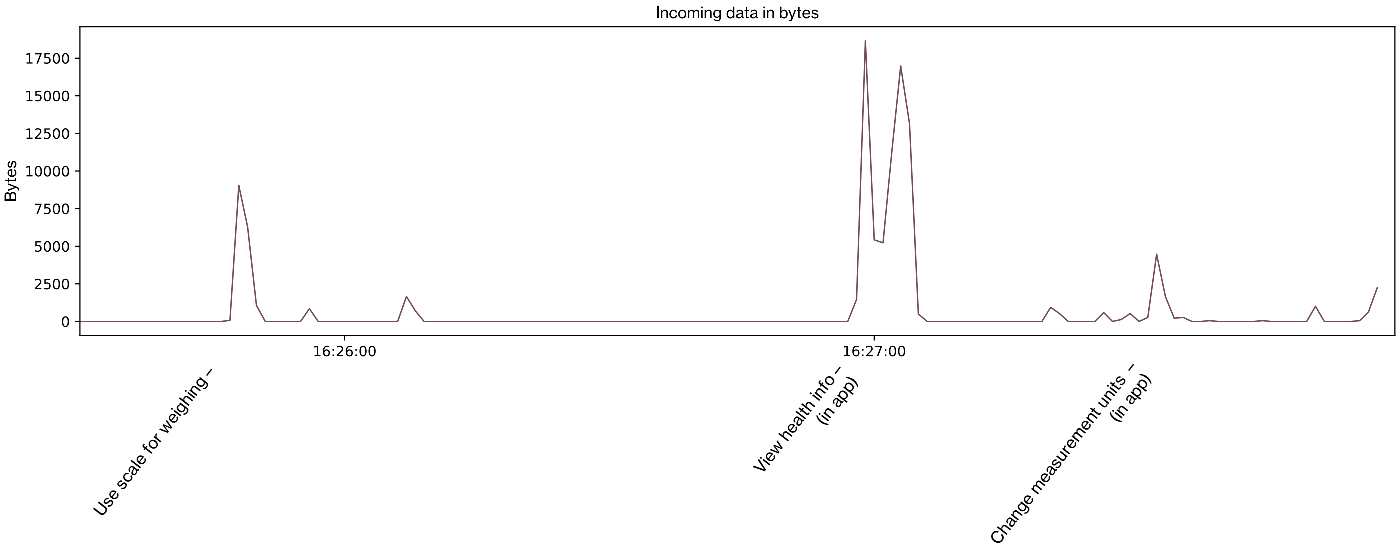


Figure 20: Data transmissions from the app of the smart scale to external actors

---

[71] As opposed to other possible implementations where such inferences might be made and retained locally, i.e. on-device or in-app only.

Cameras (security cameras, video doorbells, baby monitors)

This study involved different products with camera functionality, including security cameras, video doorbells, and baby monitors, which here we consider collectively. Vendors of these product types generally provided little information about what kind of data was transmitted, even though these products appear to communicate with a number of external actors (**§5.2**). None of the vendors returned sensor data (e.g., audio and video), and only two vendors provided usage logs. One usage log included times of events linked with locations, event IDs, and device data in JSON format, without much additional information for interpretation.

A security camera vendor returned usage logs that involve motion detection and events where the device was on and off, which are presented in Table 17. The logs show that the camera detected motion between 20:29:04 and 20:48:08. Our data flow analyses show that in the same time frame, both the device and the app sent data externally as we were conducting our usage experiments. This shows that the external transmissions include data about product usage, such as information on motion detection, whether the camera was on, and whether the camera stream has succeeded (Table 17).

Category	Text	Metadata
camera_stream_failed	[user name] opened a stream on Kitchen SmartCam	utc_device_time: 2022-03-24T20:29:04Z, os: Android, thing_id: 906625, error_code: -42, step: 0
smartcam_offline	smartcam_offline	thing_id: 906625, application: iot-adapter
smartcam_online	smartcam_online	thing_id: 906625, application: iot-adapter
motion_detected	motion_detected	thing_id: 906625, smartcam_clip_id: A83FA182DD0911648154007, application: iot-adapter
motion_detected	motion_detected	thing_id: 906625, smartcam_clip_id: A83FA182DD0911648154310, application: iot-adapter
motion_detected	motion_detected	thing_id: 906625, smartcam_clip_id: A83FA182DD0911648154626, application: iot-adapter
camera_stream_succeeded	[user name] opened a stream on Kitchen SmartCam	utc_device_time: 2022-03-24T20:48:08Z, os: Android, thing_id: 906625

Table 17: An extract from the usage logs returned by the security camera vendor

In addition, one video doorbell vendor indicated that they cannot provide the requested data because all video and audio is stored on the companion app or device, and that the information will not be transmitted to the vendor. This appears consistent with our observations, in that while Figure 21 shows that while some data was sent to external actors when the user was interacting with the device, the nature of the transmissions does not suggest that video data was sent externally; and further, there were negligible data flows from the companion app to external actors.

That said, there was some alignment between our actions and certain peaks in data flow, about which the vendor did not return anything to us (Figure 21).[72] Another video doorbell vendor indicated that they only store video data if the user selects a subscription plan, and one baby monitor vendor provided a "sample log" with time stamps and device data, with the remark that this was only stored for three months.[73] The remaining video vendors did not provide any information clarifying what data our transmission observations involve.

---

[72] The only data the vendor provided in their response related to device specifications.

[73] Notably, this vendor responded 5 months and 10 days after the initial request was made, and a follow up email and reminders had been sent; so the timeline for the response does not well-align with the vendor's stated retention period.

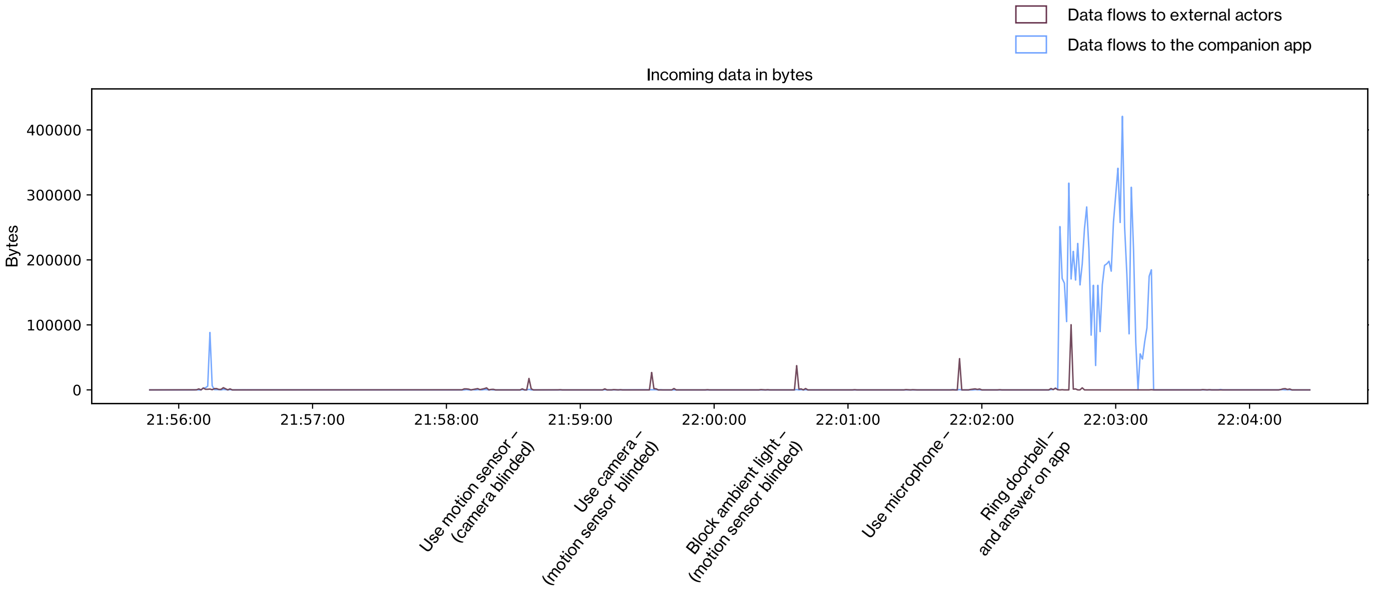


Figure 21: Transmissions from a video doorbell to the companion app and to external actors. The graph indicates flows from the doorbell to the app, and from the doorbell to external actors.

Remaining vendors: smart lighting, child smart watches, and smart plugs

We did not receive any meaningful usage-related data in vendor rights responses for these products. This was despite observing data being transmitted externally from apps and devices during use; for example, Figure 22 shows that data from a smart plug involved transmissions when the device was turned on and off.

The need for meaningful rights responses

Overall, the above shows that adequate responses to rights requests are important for interpreting the nature of data transmissions, and for shedding light on what data vendors might have about individuals, and what vendors might be doing with that data. Where vendors responded to such requests with details of the data they held, it helped to clarify what aspects of product usage (and therefore, what information about the user) are known by (at least) the vendor, as well as what the particular data flows we observed might represent. However, though we observed that every device transmitted at least some data to external actors during usage, very few vendors returned data that actually described usage specifics or that perceived by sensors – despite our explicitly requesting this information. The result is that there are significant uncertainties as to what the data flows we observed from products to external actors mean and represent.[74] Again, this emphasises the need and importance for IoT vendors to employ better data rights handling practices.

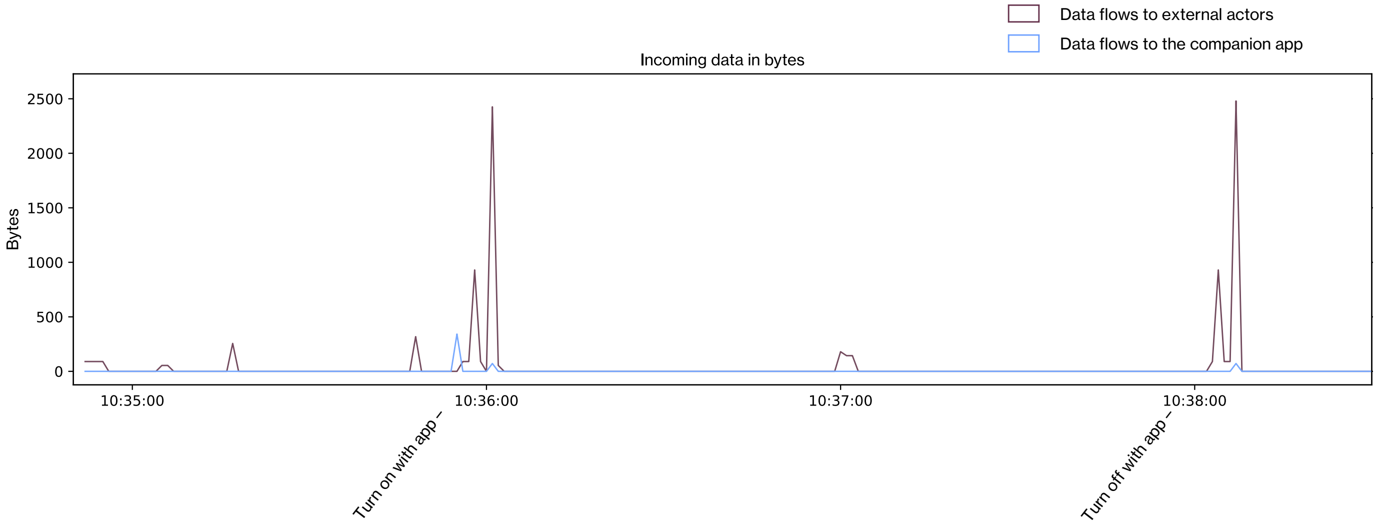


Figure 22: Data transmissions from a smart plug to the companion app and from the smart plug to external actors. The highest peaks align with times when the device was turned on and off.

---

[74] Though the information returned by vendors of similar product types can provide some indications of what we might expect from others.

5.4 Data storage practices

In this subsection, we unpack issues around data storage and retention. Storing and retaining data can be important in supporting particular product functions; for example, a smart scale might store data for a long period to allow customers to monitor shifts in weight and body composition over time. However, when data is stored, there are increased risks of unauthorised access, sharing, alternative uses, and data leakage. In contrast, if the data just passes temporarily through servers and does not persist, or indeed, only remains on the device or app, there are fewer opportunities for such issues to occur. This does not mean, however, that problematic data practices cannot and do not occur when data is just temporarily processed or held.

As we have made clear, our data flow monitoring only reflects a product's direct communications. It does not provide visibility over what happens with the data after it reaches a recipient. Therefore, to gain information on how the recipients of data from IoT products store the data, we rely on privacy policies and vendor responses to our rights requests.

5.4.1 Privacy policies

We analysed vendor privacy policies to gain insight into data storage practices, including retention periods and storage locations. We found that privacy policies typically indicate that, unless the user takes action (e.g., exercising their right to erasure), the retention periods depend on the purposes of collection. In these privacy policies, data is often said to be stored until no longer needed for the purposes outlined. However, these ***purposes are generally described in a generic way***. For instance, the most common purposes we identified in the policies concerned providing or improving products or services. Other common purposes broadly described using data for communication and marketing ends, fulfilling contracts, personalisation and analytics, or for advertising. While the broad definitions of purposes make it easy for vendors to argue that retaining data is useful and justified, for individuals, it generally remains unclear whether and when their personal data is stored, and whether and when it would be deleted.
Secondly, we observed that some vendors state that they hold the right to ***retain data even after the retention period expires***. For instance, a security camera vendor mentions in their privacy policies that data can, depending on the jurisdiction, be retained for longer periods, based on broad grounds such as "security", "abuse prevention", "public interest", "historical research" or "statistical purposes". This vendor further indicates in their privacy policy that it has data centres in China, Germany, India, Russia, Singapore, and the US. While such provisions may in some cases be lawful, this may be of particular interest, given that this vendor is headquartered in a foreign country with different views about citizens' rights, security, and public interest matters; and uses data centres in countries that may raise similar concerns.

Moreover, one vendor mentions in their privacy policy that they will retain "some data for longer periods of time when necessary for legitimate business or legal purposes, such as security, fraud and abuse prevention, or financial record-keeping". While it was already unclear when data would automatically be deleted, certain provisions in privacy policies provide additional reasons for vendors to store data for longer periods, which further adds to uncertainty for individuals.

Further, some vendors state that they might ***anonymise the data instead of deleting it***. In theory, anonymisation makes it impossible to identify the individual, and anonymous data is outside the scope of data protection law. As such, it is considered by some data controllers as an alternative to erasure. A security camera and voice assistant vendor, for example, writes in its privacy policy: "we anonymize advertising data in server logs by removing part of the IP address after nine months." This means that even if a user requests erasure or terminates their account, data relating to them might continue to be stored in some form. Anonymisation as means for erasure can be problematic, because "true" anonymisation is a difficult if not impossible task;[75] because anonymisation is not formally put forward by data protection law as an equivalent measure to erasure; and because it could disempower the individual.[76]

5.4.2 Vendor responses

In our transparency rights requests, we asked every vendor where and for how long the data is stored. Only 13 vendors provided ***some*** information about retention periods, storage locations, or storage providers for one or more types of personal data. Four of these vendors mentioned the country where the cloud provider stores the data or where the data is otherwise stored. We discuss three key issues from the analysis of responses.

First is that ***retention periods are seldom specified for types of personal data***. Only seven out of the thirteen vendors that mentioned retention periods in their responses indicated the periods for ***specific types*** of personal data (see Table 18). Six of those were vendors of products with some camera functionality (e.g., baby monitors, security cameras or video doorbells). These vendors sometimes indicated retention periods specifically for sensor perceived data such as audio and video recordings. The other vendor was a smart lighting vendor, that specified the retention of usage data. The retention periods provided vary, ranging between seven days and 13 months. In most cases, vendors indicate that certain data types, particularly video recordings, will be retained for a longer period if users choose a certain subscription plan. In one case, inactivity was listed as grounds for deletion. One security camera vendor told us that video images will

---

[75] Pseudonymised data is still subject to data protection law.

[76] Ausloos, J., Mahieu, R., & Veale, M. (2019). Getting Data Subject Rights Right. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, *10*(3), 283–309

be removed after the camera has been offline for 90 days.

We also observed in the responses that the *details regarding retention periods can be vague*. Some vendors indicate that they will store certain data for certain periods of time but remain unclear about the duration of this period, and again, the types of personal data they refer to. For example, a large vendor of TV streaming sticks indicated that “some data you can delete whenever you like, such as the content you create or upload”, and “other data is deleted or anonymized automatically after a set period of time, such as advertising data in server logs”. They also mention that they “keep some data until you delete your [account], such as information about how often you use our services”. Responses concerning a security camera and a voice assistant from that same vendor indicate that they will delete “video footage within a few days and most other account data within a few months”. “However”, they add, “some data may persist for longer periods of time”. These vague descriptions do not contribute much to user understanding of whether, and for how long, their data will be stored.

Thirdly, vendors provide *little visibility over who stores the data and where*. Table 18 provides an overview summarising the vendors from whom we retrieved some information about retention period, the data types retained, storage providers, or storage locations in their rights responses. Vendors whose replies did not mention storage or retention are not included. In their responses, only five of the 13 vendors specified the location of data storage. Four of these vendors mention the name of the company responsible for storage of datasets.

In all, our overarching finding is that *most vendors are unclear about the periods for which particular data is retained*, both in their privacy policies and in their responses to our rights requests.

Product type	Data type	Specified retention period	Storage provider	Storage location
Baby monitor 3	Activity detection and event logs	7 days	AWS	Germany, Hong Kong
Baby monitor 3	Audio and video data	Not stored		
Baby monitor 4	Account log data and device log data	3 months		
Baby monitor 5	Camera data	7 days, if there is no subscription plan		
Fitness tracker 1	Personal data for analytics, improvement, and sales and marketing	12 months		
Fitness tracker 1	Backups and app logs	6 months		
Security camera 1	Event data and usage logs	30 days		Ireland
Security camera 2	Some device and sensor data	13 months		
Security camera 2	Other data	may persist for longer periods of time		
Smart lighting 4				Frankfurt, Germany
Smart lighting 5	Device usage data	7 days	Hanzhou PlusMinus Smart Technology	Frankfurt, Germany
Smart plug 1			AWS	
Smart scale 1			AWS	US, Germany
Video doorbell 1		Depending on chosen plan		
Video doorbell 4	Audio and video		On the device only	
Video doorbell 4	Product device logs	5-7 days		
Video doorbell 4	Interface logs	20 days		

Table 18: Overview of details on data storage and retention in vendor responses. Empty cells denote that we did not receive information on that point

Section 5: Summary

This section focused on the data flows associated with IoT products, aiming to illuminate the data processing practices in the IoT. We undertook data traffic monitoring to provide insight into the data flows from IoT products, using that as a basis for exploring – by way of data rights – what information vendors (and others) processed, and how their responses aligned with the data transmissions we observed.

§5.1 showed that the consumer IoT entails vendors collecting various kinds of personal (and related) data about individuals. We saw that vendors were generally more straightforward with returning account information and device specifications, than data about usage such as sensor obtained data, inferences, or detailed usage logs.

§5.2 showed that IoT products transfer data to a range of different IP addresses, in various countries, belonging to various organisations. Perhaps unsurprisingly, infrastructure providers, such as cloud platforms, appear to play a prominent role in IoT ecosystems, but many other organisation types also seem to be involved. However, very few vendors provided information about these other parties.

§5.3 showed that product usage is reflected in data flows to external actors. We saw evidence of this in the few responses that provided sufficient detail, in which real-time data traffic appeared to correspond both with our interactions with the product as well as with information vendors held. However, despite seeing that many devices sent data externally during product use, very few vendors returned responses about transmissions and therefore what was communicated.

§5.4 considered data retention. Given we saw that every product transmitted data, we expect some data will be retained. We found that, generally, vendor privacy policies were vague in their retention criteria. In their responses, vendors seldom returned data, nor gave specifics as to the period for which particular types of data are retained, and provided little transparency into who stores the data and where. This is of concern, because if data is stored, vendors are obliged to return it (along with information about it) upon request.

In all, we observed that IoT products entail a range of external data flows, regarding several categories of data, and where details about product usage appear to be communicated to various actors. However, vendor responses to our rights requests rarely provided enough detail to support and clarify those observations, nor whether, where, how long, and by whom data is shared and stored. The few responses we did receive were rather helpful for interpreting what occurred, e.g., what particular data transmissions might represent. Overall, our findings emphasise the need for better data rights practices.

6 Blocking IoT data flows

In the previous sections, we showed that IoT products can involve transmissions to various destinations. Moreover, we concluded that the current approaches taken by vendors towards data transparency rights are generally insufficient for individuals to gain meaningful transparency over a vendor's data processing practices. In various contexts, some have proposed blocking (or preventing) certain data flows as a mechanism to provide the user with a degree of control. We consider the potential for blocking in the consumer IoT and, in particular, whether blocking is feasible,

In the previous sections, we showed that IoT products can involve transmissions to various destinations. Moreover, we concluded that the current approaches taken by vendors towards data transparency rights are generally insufficient for individuals to gain meaningful transparency over a vendor's data processing practices. In various contexts, some have proposed blocking (or preventing) certain data flows as a mechanism to provide the user with a degree of control. We consider the potential for blocking in the consumer IoT and, in particular, whether blocking is feasible, desirable and sustainable.

6.1 Background on blocking

Blocking works to prevent data flows to certain destinations. Blocking has been proposed in various domains as a means for empowering users with greater control over their data and privacy, one prominent example being the blocking of tracking cookies and advertisers in web browsers ("Ad Blocking").[77] Blocking has also been proposed in the consumer IoT context, where it could operate to prevent unwanted data flows from the product to particular IP addresses.[78] Because one still wants to use the device, the idea is to block the interactions with recipients that are non-essential for the functioning of the product.

Colleagues from Imperial College have developed and tested a systematic methodology for blocking traffic in the IoT to non-essential destinations.[79] They found that 16 out of the 31 IoT devices that they tested had at least one address that was not required for the functioning of the device. The work indicates the possibility for blocking data flows that are not essential for a product's core functionality.

To explore the potential and limitations of blocking, we conducted two technical case studies and outline the legal, social, and technical considerations of blocking in the consumer IoT.

6.2 Blocking case studies

We conducted case studies with two of the devices we surveyed: a light bulb that interacts with the fewest IP addresses and a baby monitor that interacts with a large number of IP addresses, meaning these devices represent products whose data transmission practices are at both ends of the spectrum. Because we blocked data transmission through the Monitoring Access Point (MAP), the IP addresses that were blocked affected the communications for both the IoT devices and companion apps.

6.2.1 Case study 1: the light bulb

The first case study involved a light bulb we selected because the number of destinations for its transmissions is particularly low: two IP addresses associated with AWS in Germany. We blocked one IP address at a time, while

---

[77] Achara, J. P., Parra-Arnau, J., & Castelluccia, C. (2016). MyTrackingChoices: Pacifying the Ad-Block War by Enforcing User Privacy Preferences. Workshop on the Economics of Information Security (WEIS 2016). arXiv:1604.04495. https://doi.org/10.48550/arXiv.1604.04495; Nithyanand, R., Khattak, S., Javed, M., Vallina-Rodriguez, N., Falahrastegar, M., Powles, J. E., De Cristofaro, E., Haddadi, H., & Murdoch, S. J. (2016). Ad-Blocking and Counter Blocking: A Slice of the Arms Race. 6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 2016). arXiv:1605.05077. https://doi.org/10.48550/arXiv.1605.05077

[78] For example, see: Haar, C., & Buchmann, E. (2019). Fane: A Firewall Appliance For The Smart Home. Federated Conference on Computer Science and Information Systems, 449-458. https://doi.org/10.15439/2019F177; Gupta, N., Naik, V., & Sengupta, S. (2017). A firewall for Internet of Things. 2017 9th International Conference on Communication Systems and Networks (COMSNETS), 411–412. https://doi.org/10.1109/COMSNETS.2017.7945418; Mandalari, A. M., Kolcun, R., Haddadi, H., Dubois, D. J., & Choffnes, D. (2020). Towards Automatic Identification and Blocking of Non-Critical IoT Traffic Destinations. ArXiv:2003.07133 [Cs]. http://arxiv.org/abs/2003.07133; Mandalari, A. M., Dubois, D. J., Kolcun, R., Paracha, M. T., Haddadi, H., & Choffnes, D. (2021). Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. Privacy Enhancing Technologies Symposium (PETS) 2021. ArXiv:2105.05162 [Cs]. http://arxiv.org/abs/2105.05162; Tagliaro, C., Hahn, F., Sepe, R., Aceti, A., & Lindorfer, M. (2014). I Still Know What You Watched Last Sunday: Privacy of the HbbTV Protocol in the European Smart TV Landscape; I just wanted to track my steps! Blocking unwanted traffic of Fitbit devices. Proceedings of the 12th International Conference on the Internet of Things, 96–103. https://doi.org/10.1145/3567445.3567457.

[79] Mandalari, A. M., Kolcun, R., Haddadi, H., Dubois, D. J., & Choffnes, D. (2020). Towards Automatic Identification and Blocking of Non-Critical IoT Traffic Destinations. ArXiv:2003.07133 [Cs]. http://arxiv.org/abs/2003.07133; Mandalari, A. M., Dubois, D. J., Kolcun, R., Paracha, M. T., Haddadi, H., & Choffnes, D. (2021). Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. Privacy Enhancing Technologies Symposium (PETS) 2021. ArXiv:2105.05162 [Cs]. http://arxiv.org/abs/2105.05162

interacting with the device in the same way as we did in the previous experiments, by turning the light on and off, and changing colours. We found that with either destination blocked, it was not possible to control the light bulb with the companion app. In this case, blocking data flows made the device unusable.

6.2.2 Case study 2: the baby monitor

The second experiment focuses on a baby monitor, which we found interacted with 178 unique IP addresses, relating to 34 organisations across 10 different countries. From the IP analysis described in **§5.2**, we selected two IP addresses that appeared interesting for our blocking experiments: one that appears linked with an advertiser, and one involving unencrypted data flows.

The first IP address was linked with "Gourdmobi", a company that appears to provide mobile advertising and "intelligent" marketing services. We repeated the usage experiments with the product, where we recorded for about 30 seconds, then used the microphone, then the pan-tilt-zoom (PTZ) function,[80] then taking two pictures using the "360 picture" and "snapshot" functionalities. When we blocked the IP addresses associated with Gourdmobi, the recording and microphone still functioned as usual and we were able to live watch the video on the companion app. However, when we used the PTZ, 360 picture, or tried to take a snapshot, a pop-up showed up in the app telling us the "operation failed". This suggests that the product was designed to require interactions with this IP to use these particular functionalities.

The other IP address was associated with unencrypted images that were received by the companion app. The IP address seems to be linked with Amazon servers in the US. Blocking this IP address did not appear to have any effect during the experiments.

This case study shows that it may be difficult to determine which IP addresses can be blocked and which cannot. Moreover, depending on how the product is designed, it shows that even if one wants to block traffic to an IP address – such as an organisation that could be an advertiser – this could affect the functioning of the device.

6.3 Socio-technical considerations

Overall, we confirm that blocking is potentially feasible at a technical level,[81] but that it comes with challenges. We next explore some limitations to blocking on a larger scale.

6.3.1 Technical considerations

Importantly, how and when data is transmitted is a design choice. Observing data flows to fewer destinations does not necessarily mean the situation is any less problematic, as the few recipients can always forward that data on to others. In line with this, if vendors perceive wholesale blocking as a risk to their business, they could modify their data flow architecture such that data flows are first directed to their servers – perhaps then as the ***only*** external destination – where the vendor themselves forwards that data to any third parties, as they see fit. This would not only further hinder the visibility of data flows in the IoT, but also render the blocking mechanisms largely ineffective (though there could be some scope for filtering the transmission of *particular data*, i.e. content, rather than connections). In all, we argue that an important way forward is to encourage better overall privacy by design practices in the IoT industry.

6.3.2 Legal considerations

We analysed Terms of Service, Terms of Use, Terms and Conditions, and legal documents of a similar nature (which we collectively refer to as "Terms"), to see whether vendors currently appear to consider blocking directly, or have conditions that might relate to or restrict the blocking or filtering of data flows. Overall, we found that, although blocking data flows currently does not seem to be *explicitly* considered by vendors, some Terms include provisions that might be applicable to blocking practices. We now discuss different types of provisions in the Terms of vendors included in this study.

No restrictions

More than half of the vendors did not mention anything that can be clearly linked to attempts to restrict blocking or other interference with the software, hardware, or systems.

Restrictions meant to protect the integrity of systems

We came across documents relating to 18 products (out of 43) that involve certain provisions that appear to be aimed at protecting the integrity of systems and services. Of those, 16 mention restrictions on reverse engineering the product.[82] Though it was not made explicit, blocking could potentially be interpreted as a means for reverse-engineering part of a product or modifying the function of a device – not least as an aim is to work out which data flows to block (or not), which then stops the device from operating in the manner in which it was originally designed.

A number of Terms contain provisions that prohibit the violation of "the security of any computer networks" or the incorporation of "any disabling code designed to permit improper use, access, deletion or modification of software or hardware programs or systems".

---

[80] PTZ (or "pan-tilt-zoom") refers to the capacity of a camera to move horizontally and vertically, and adjust the focus of the lens.

[81] See also Mandalari, A. M., Kolcun, R., Haddadi, H., Dubois, D. J., & Choffnes, D. (2020). Towards Automatic Identification and Blocking of Non-Critical IoT Traffic Destinations. ArXiv:2003.07133 [Cs]. http://arxiv.org/abs/2003.07133; Mandalari, A. M., Dubois, D. J., Kolcun, R., Paracha, M. T., Haddadi, H., & Choffnes, D. (2021). Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. Privacy Enhancing Technologies Symposium (PETS) 2021. ArXiv:2105.05162 [Cs]. http://arxiv.org/abs/2105.05162; https://moniotrlab.ccis.neu.edu/pets21/.

[82] Including some household name vendors.

Such prohibitions do not seem to explicitly prevent individual users from blocking traffic on their own home networks but rather appear to be about protecting the integrity and security of the vendor's systems, networks, and processes. It is debatable whether this might relate to blocking activities of the type we discuss here.

Restrictions that hint in the direction of blocking restrictions

One company explicitly states that the user "agrees **not to interfere with, disrupt** or attempt to gain unauthorised access to **the services or networks connected to the services** or **violate the regulations**, policies or procedures of such networks" (our emphasis added).[83] Although even this wording does not explicitly mention blocking itself, it could be argued that this would cover blocking.

Moving forward we could see vendors deciding to employ specific terms in an attempt to prevent users from blocking, should they see (large-scale) blocking as a threat to their business aims.

6.3.3 Social considerations

Blocking has been argued by some as a mechanism to empower users to take control over their data. However, such an approach, depending on how it is implemented, could disadvantage relatively less "tech savvy" individuals and households; potentially leaving particular demographics at greater risk of being subject to higher levels of surveillance.

Moreover, blocking is ultimately a form of "privacy self-management".[84] There are questions regarding the extent to which individuals should bear the burden of managing their interests in an inherently complex and opaque technical ecosystem, not least given the potential for questionable and possibly non-compliant data processing practices of vendors. In other words, is it desirable that individuals are burdened with the responsibility to block unnecessary data flows, when the existence and concerns regarding these data flows may be a direct result of vendors' not adhering to key data protection principles in the first place?

Section 6: Summary

We found that although blocking may technically be possible, it raises a number of considerations.

First are the practical concerns: knowing what to block is difficult; vendors could (re)design products such that blocking becomes impractical or irrelevant; or vendors may attempt to employ legal mechanisms to try and prohibit blocking being employed by users.

More broadly, mechanisms such as blocking raise questions as to whether it is appropriate that users should be burdened with managing the problems that ultimately result from vendor data processing practices in IoT ecosystems.

---

[83] "(e) Certain Restrictions. The rights granted to you in these Terms are subject to the following restrictions [...] (ii) you agree not to modify, make derivative works of, disassemble, reverse-compile or reverse-engineer any part of the Services or Products; [...] (vi) you agree not to interfere with, disrupt or attempt to gain unauthorised access to the servers or networks connected to the Services or violate the regulations, policies or procedures of such networks; (vii) you agree not to access (or attempt to access) any of the Services by means other than through the interface that is provided by [vendor]; [...]" [Excerpt from vendor's Terms of Service]

[84] See Solove, Daniel J. (2012). Introduction: Privacy self-management and the consent dilemma. Harvard Law Review 126, 1880-1903.

7 Conclusion and future work

This report focused on issues of transparency in the consumer IoT. We employed legal and technical mechanisms in an attempt to provide insight into the data processing practices associated with consumer IoT products.

Our experience indicates that IoT vendors generally perform inadequately when it comes to data transparency rights (Section 4). The response rates to data access and data portability requests were overall surprisingly low, with nine out of 43 product vendors not responding at all, and those that did respond often failed to directly nor fully address the points raised in our requests, even after reminders and follow-ups. We also found that the process of interacting with vendors to obtain a meaningful response to our rights requests was often cumbersome, while the form and format in which responses were returned was generally inconsistent, generic, or unnecessarily complex. Given that data transparency rights are a key aspect of data protection law, poor adherence to these rights by vendors in the consumer IoT is concerning and problematic.

Furthermore, we found that monitoring the data flows of IoT products revealed some noteworthy observations, but that the responses of vendors (or lack thereof) to our rights requests in the vast majority of cases did not meaningfully support, clarify, or supplement what we saw. That said, by analysing both the results from data flow monitoring and some of the responses to our rights requests, we observed that IoT product vendors do collect and hold a range of data (**§5.1**), with categories including account registration data, device specifications, sensor perceived data, usage logs, and inferences. We saw that many products send data to quite a number of IP addresses; on average we saw that a product interacted with 88 IP addresses, linked with 12 organisations and eight countries. While most data went to the US, UK, or EU, transfers to countries outside of these areas were regularly observed (**§5.2.2**). When asking vendors about this, we received some information on the categories of third parties with whom data has been disclosed, but only five vendors were specific about the recipients and the purposes of data sharing.

We saw that the use of IoT products resulted in data transmissions (**§5.3**). Those vendors that did respond to our requests provided information that helped us to confirm some intuitions and better understand the nature of the data flows we observed while using the products. However, the lack of meaningful, detailed responses by most vendors means significant uncertainties remain for those attempting to understand how their personal data is being used. We also found that vendors are generally unclear about the data storage and retention specifics for their products (**§5.4**).

Our findings indicate that in the consumer IoT, different types of data are collected and shared in various ways, yet the specifics of such often remain opaque. Furthermore, what happens to data after it is sent to others is unclear – despite the use of data transparency rights. This is concerning, given the personal and intimate domestic settings in which consumer IoT devices operate.

Finally, we considered the potential for blocking, which is a subject of ongoing research in this domain (Section 6). We found that blocking mechanisms could potentially allow users to prevent the flow of data to particular destinations. However, there are still outstanding technical, legal, and social aspects that need to be considered, and the longevity or even the appropriateness of such an approach remains questionable.

Of course, our findings are only indicative of some of the potential issues of the consumer IoT: we examined only a limited set of the ever-growing number of IoT products that are available; our approach in externally examining (or "adversarial auditing") products and their vendors provides but a narrow view over certain actions and behaviours, leaving some uncertainties and much to be inferred; and our work represents but a snapshot of the current state of affairs at the time of testing, and thus we expect things to evolve, particularly as data protection issues become more prominent and consumer IoT devices more commonplace. That said, there were clear commonalities and trends in our findings, which does indicate that there are some serious concerns with respect to personal data flows and data transparency rights in the consumer IoT. Therefore, our work serves as a foundation, to provide initial insights into some issues and presenting opportunities for future research and intervention.

7.1 Future work

To build on the findings of this report, we make some recommendations for future work. These aim at both increasing levels of understanding of the data processing practices in the consumer IoT, and possible approaches towards addressing the issues discussed in this report.

Explore how data rights requirements can be better implemented
We found that the implementation of the right of access and the right to data portability by IoT vendors was lacking. Interventions could be informed by building better understandings of the reasons why some organisations may fail to inadequately respond, and what guidance and support are needed by different types of vendors such that they properly respect and implement data rights.

Visible intervention and enforcement
Effective intervention and enforcement measures are important for highlighting and discouraging poor behaviour, and for encouraging vendors to improve their data rights and data processing practices. This could entail streamlined mechanisms for data subjects to report issues and support effective follow-ups on complaints, programmes for testing the compliance of IoT product and vendor behaviour, or other forms of monitoring for uncovering issues in the consumer IoT landscape.

Develop best practices regarding data rights and processing in the IoT
There is much inconsistency among IoT product vendors in the way they deal with rights requests and interact with data subjects, and in the format in which responses were returned. Beyond data rights, it was clear that there were differences in the data processing practices across vendors, some performing far better than others. This suggests there is real opportunity to encourage learnings and the development of best practices. Examples might include, for instance, suggesting approaches that do not require account registration for product use; enabling other forms of identity for verifying rights requests (such as in-app authorisations) rather than requesting government-issued photo ID; and encouraging data flow architectures that retain as much data as possible in the local network or device, among others.

Consider potential roles for dominant platforms in oversight and compliance
In our analysis, we saw that a small group of organisations – mostly cloud platforms – underlie much of the consumer IoT, providing infrastructure supporting many IoT products. Regulators could consider how to leverage the central position of these organisations for compliance and to encourage better general practices. One potential approach to consider, for example, is where large players (infrastructure providers) in the consumer IoT could perhaps have a responsibility to “know their customers”, or maintain certain standards and rules regarding data transactions in a manner consistent with good data protection practices – similar to how banks play a role in helping to identify and prevent criminal activity.

Explore mechanisms for visibility beyond primary recipients
Our technical analyses only considered the data flows directly with the product and external IP addresses. This is because the data flow monitoring infrastructure cannot see what happens to data after it is received by an address. Moreover, the rights responses from vendors generally did not make the picture clearer, particularly regarding any subsequent transmission of data to others. From the responses we received, only four vendors provided the identities of the organisations they interact with. However, we do note that in January 2023, the European Court of Justice (ECJ) has specified that the right of access requires that the data controller (here the vendor) communicate the specific identities of data recipients, rather than just providing the general category of recipients (which tended to be what we received, at best), save in exceptional cases.[85] Though the ECJ no longer binds the UK, given that it relates to a legal framework that is substantively the same as the one that applies in the UK, it may be persuasive for a UK court ruling on the same question. That said, although such rules have been clarified as being more stringent than many had thought, it is unclear how impactful such a clarification would be in practice, given that our experience is that compliance remains an outstanding area of concern.

Further examine the impact of transnational data transmission
We found that the data from IoT products flows to addresses associated with countries all over the world. This can have implications when data moves to countries that have different data protection standards, or where their authorities make use of “law enforcement” or “national security” as grounds to access data. It is therefore recommended to build a better understanding of the current and potential impact of transnational data transmissions, and how IoT vendor compliance with certain rules, rights, or standards may be affected by this.

---

[85] Case C-154/21 Österreichische Post, ECLI:*EU*:*C*:2023:3.

Explore mechanisms for making rights responses more meaningful and interpretable

We observed that the responses to our rights requests tended to be inconsistent among vendors. The responses to our questions (as part of exercising our data transparency rights) were seldom complete and often not tailored to our requests. Further, data was returned in varying formats and sometimes lacked explanations of relevant concepts and definitions to aid in interpreting such information – these findings consistent with previous research on data rights.[86] Therefore, there is clear scope for mechanisms encouraging responses that are more meaningful and easier to understand, and thereby more effective for data subjects.[87]

Focus on products of a particularly sensitive nature

We have seen that some devices designed for more sensitive uses, such as those for young children and those collecting data directly obtained from human bodies, can transmit data to many destinations. As such, it would be particularly interesting to undertake a deep-dive with IoT products that operate in sensitive scenarios and collect intimate data, such as fertility monitoring devices and safety trackers aimed at women, children or the elderly. Such products could potentially harm the rights of particular groups, including those vulnerable, and often entail “special category” data. Such devices therefore warrant closer monitoring. We are currently extending the work described in this report, by conducting some analyses with products of these types.

---

[86] See Ausloos, J., & Dewitte, P. (2018). Shattering One-Way Mirrors. Data Subject Access Rights in Practice. International Data Privacy Law 8(1), 4-28. https://papers.ssrn.com/abstract=3106632; Wong, J., & Henderson, T. (2018). How Portable is Portable?: Exercising the GDPR’s Right to Data Portability. Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 911–920. https://dl.acm.org/doi/10.1145/3267305.3274152

[87] Norval, C., Cornelius, K., Cobbe, J., & Singh, J. (2022). Disclosure by Design: Designing information disclosures to support meaningful transparency and accountability. 2022 ACM Conference on Fairness, Accountability, and Transparency, 679–690. https://doi.org/10.1145/3531146.3533133