Towards an integrated privacy protection framework for IoT: contextualising regulatory requirements with industry best practices

Robert Thorburn, Andrea Margheri, Federica Paci

University of Southampton, UK {r.h.thorburn, a.margheri,f.m.paci}@soton.ac.uk

Abstract

One of the main obstacles to the widespread adoption of IoT devices and services is consumers' privacy concerns related to personal data collection, processing and sharing with third parties. Indeed, many IoT devices have been found collecting consumers' personal data without their knowledge or consent. While frameworks for identifying and mitigating security concerns of IoT devices and services are available, there is a lack of frameworks that address privacy issues for IoT applications. In this paper we lay the foundations for the future development of such a framework, based on both the experimental analysis of data flows on an IoT Smart Home testbed and a systematic analysis of other frameworks.

Keywords: IoT Privacy, Privacy framework, Privacy-by-Design, Smart Home

1 Introduction

The advent of the Internet of Things (IoT) is revolutionising data collection, processing and sharing. IoT systems can access, process and manage high volumes of data, including highly sensitive data. This has raised consumers' privacy concerns and eroded their trust in IoT systems and services.

This erosion of trust is wholly unsurprising given the nature and prevalence of media reporting on Smart Home related privacy threats, including devices such as smart toys, baby monitors and voice assistants. For instance, a woman in Portland, Oregon found out that her family's home digital assistant, Amazon's Alexa, had recorded a conversation between her and her husband without their knowledge, and sent the audio recording to a person on their contacts list.[1]

Pressing privacy concerns, including those in IoT, have lead to the adoption of multiple privacy regulations, including the European General Data Protection Regulation (GDPR) [6]. This new legal framework poses complex challenges for processing personal data which have to be met by IoT device manufactures, vendors and third-party service providers. In particular, GDPR advocates Privacy-by-Design (PbD) which means

---

[1] https://www.theguardian.com/technology/2018/may/24/amazon-alexa-recorded-conversation

that privacy issues should be considered as part of the design and implementation of systems, services, products and business practices. Therefore, due to the pervasiveness of IoT, PbD should be applied to devices collecting personal information, as well as to networks, back-end systems and software applications that transmit and process data. However, there is still uncertainty about what PbD means in the context of IoT, and most of all how we can implement it.

Currently, two main initiatives have proposed a PbD approach targeted to IoT devices and services: the Privacy Impact Assessment framework from CNIL (the French Data Protection Authority) [5] and the 3P Framework for IoT Privacy-By-Design [3]. However, both initiatives propose a standard risk management process without fully contextualising it to IoT nor providing guidelines on how to integrate the process into the design of IoT devices and services.

Contribution. Our research proposes an integrated IoT privacy framework for the principled adoption of PbD which reconciles both *technical*—e.g. due to application requirements—and *compliance-driven*—e.g. due to GDPR—approaches. In this paper, we pave the way to this result by eliciting a set of integrated principles based on experimental analysis of an IoT Smart Home testbed, and on the assessment of available privacy compliance frameworks.

IoT Smart Home devices have been chosen over industrial applications due to the former's propensity for under-performing on privacy and security metrics. This is not to say that other IoT applications are without significant challenges [16], just that Smart Home applications are particularly susceptible to privacy issues [1, 13].

Structure of the paper. Section 2 provides background information and discusses related work. Section 3 formalises the methodology employed. Section 4 reports our key findings. Section 5 discusses a set of integrated principles. Section 6 concludes and delineates future work.

2 Background and Related Work

Privacy-by-Design requires data controllers to implement key data protection principles during data collection, processing

and dissemination, which should minimise the potential privacy harm to data subjects. To demonstrate compliance with these principles and reduce privacy risks, data controllers should follow a privacy impact assessment methodology.

In what follows, we first overview the fundamental principles at the heart of protecting personal data and the methodologies to ensure compliance with those principles. Then, we discuss privacy frameworks tailored to IoT.

Data protection principles. The GDPR introduced seven key principles that data controllers should comply with: 1) *lawfulness, fairness and transparency* requiring that data subjects must be clearly informed on all data collection and processing; 2) *purpose limitation* requiring data to be only collected and processed under stated purposes; 3) *data minimisation* which requires that only the needed data is collected; 4) *accuracy* mandating up to date and accurate keeping of data; 5) *storage limitation* restricting storage of data up to purpose or legal needs; 6) *data security* ensuring confidentiality, integrity and availability of data; 7) *accountability* demonstrates compliance with the above principles.

Privacy Impact Assessment Methodologies. Different parties have proposed methodologies that differ in their approach to implement PbD. Some methodologies assume that Privacy-by-Design means considering privacy risks early in the design process and selecting privacy enhancing technologies to mitigate these risks [7, 17]. For instance, the Multilateral Privacy Requirements Analysis Method (MPRAM) [7] elicits the different privacy goals of multiple stakeholders of a system and transforms them into privacy requirements. LINDDUN [17] is a privacy threat modelling technique that first identifies privacy threats and selects privacy enhancing technologies to mitigate the associated risk. The "LINDDUN" acronym is derived from the categories of privacy threat it identifies, namely: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of Information, Unawareness, and Non-compliance.

Other methodologies [15, 4] instead assume that PbD means demonstrating compliance with well-known data protection principles rather than on identifying privacy risks. For example, the German Standard Data protection Model (SDM) [15] provides appropriate mechanisms to transfer the data protection requirements of the GDPR into technical and organisational measures. In order to achieve this purpose, the SDM structures the GDPR requirements in terms of data protection goals: data minimisation, availability, integrity, confidentiality, transparency, unlinkability, and intervenability. The SDM uses these data protection goals to transfer the legal requirements of the GDPR into a catalogue of technical and organisational measures, which the regulation itself requires. The privacy impact assessment methodology proposed by CNIL, the French Data Protection Authority [4] provides a systematic process to build and demonstrate compliance with GDPR principles. The process consists of four main steps: 1) understand the data processing activities; 2) ensure the presence of controls implementing the GDPR principles, 3) assess the data security risks;

and 4) validate the results.

Privacy Frameworks for IoT. PbD is usually a neglected aspect in designing and building IoT devices and services. There are currently only two frameworks that address privacy concerns of IoT devices and smart services. The Proactive and Preventive (3P) [3] framework for IoT Privacy-By-Design provides a process to follow: 1) define IoT service design and operation blueprint; 2) develop the IoT data flows, application interfaces, infrastructure and network layouts based on stakeholder needs; 3) clarify, document and limit purposes for collecting and using personal data; 4) identify all security and privacy risks; 5) conduct privacy impact assessment of all IoT devices and data components; 6) build IoT privacy capabilities; 7) implement IoT security and privacy controls; and 8) continuously review the effectiveness of privacy controls and identify new privacy risks. Instead, CNIL [5] has applied its Privacy Impact Assessment (PIA) methodology to connected objects. However, neither of these initiatives provide direct application guidelines or a truly integrated approach.

3 Methodology

To investigate the current state of IoT implementations, we considered both *legal principles* and *technical aspects* as these are the two pillars of the integrated approach proposed herein. We used the data flows on our IoT testbed as a point of departure, from where we categorised the threats detected according to a standard privacy threat taxonomy.

Therefore, in order to propose a set of principles for building an integrated privacy framework, our methodology relies on the following steps:

1. *privacy threat identification*: pointing out privacy threats based on network traffic analysis and linking these to an established privacy threat taxonomy

2. *privacy risks*: comparing and contrasting available privacy frameworks to derive points of overlap and areas not yet addressed

3. *integrating principles*: defining integrated principles towards a new IoT privacy framework, based on the results of the preceding two steps

IoT Smart Home Testbed. A Smart Home is typically characterised by a mix of devices, but often contains a so-called starter kit with a few core devices from one supplier [10]. Our testbed devices were chosen to reflect this. The full list of devices is in Table 1, and includes four Withings devices, one Amazon Echo, a TP Link Plug and a Misfit tracker. Additionally, we used an Android smartphone for hosting device control apps.

All devices were connected to a Raspberry Pi which acted as router for all traffic and, hence, as capture point for the

Device	Description
1 *Misfit Shine 2*	Wearable fitness tracker
2 *Withings Go*	Wearable fitness tracker
3 *Withings Thermo*	Connected thermometer
4 *Withings Body*	Scale and health tracker
5 *Withings Home*	Camera and air quality monitor
6 *TP Link Plug*	Connected plug and energy sensor
7 *Amazon Echo Plus*	Smart assistant

Table 1: IoT Smart-Home testbed

Privacy Threat	Device (from Table 2)						
	1	2	3	4	5	6	7
Information Collection							
Surveillance	✗	✗	✗	✗	✗	✗	✗
Information processing							
Aggregation	✗	✗	✗	✗	✗	✗	✗
Identification	✗	✗	✗	✗	✗	✗	✗
Secondary Use	✗	✗	✗	✗	✗	✓	✗
Exclusion	✗	✗	✗	✗	✗	✓	✗
Information dissemination							
Confidentiality breach	✓	✓	✓	✓	✓	✓	✓
Disclosure	✗	✗	✗	✗	✗	✓	✓
Exposure	✓	✓	✓	✓	✓	✓	✓
Increased accessibility	✓	✓	✓	✓	✓	✓	✓
Invasion							
Intrusion	✗	✗	✗	✗	✗	✓	✗

Table 2: Privacy threat assessment according to Solove's taxonomy [12] (where ✗ indicates the occurrence of a threat, while ✓ that under normal circumstances that threat does not occur)

associated data flows.[2] Further set up specifications can be found online [14]. Notably, this setup ensures full control of the network, its connected devices and all the generated traffic and also reduces the node count by not including a standalone router. As opposed to a filter based solution on a shared network, this approach makes sure that no device generated data are accidentally discounted during the capture process. By way of example, devices using non-standard ports may be overlooked by capture filters.

Privacy Threat Identification. The experimental evaluation of the testbed was based on the analysis of pcap files of traffic captures (using TCPdump) on the Rasberry Pi. Additional insight on the data flows was obtained via the Android app 'Lumen Privacy Monitor' [9], which tracks and reports on the activities of all the apps on a smartphone. These activities enable detecting the actual *data flows* in the IoT testbed.

To detect privacy regulation infringements, the data flows had to be analysed according to the devices and their context. Specifically, for each device and its functionality we took into account terms and conditions presented to data subjects (usually during device setup on a smartphone) on data collection, data processing, data access for data subjects and control systems for data subjects. This analysis points out experimentally assessed privacy regulation infringements. To further support this analysis we utilised the IoT implementation contained in CNIL's PIA.

As a result of this step, we were able to systematically categorise the infringements found according to established privacy threats. To this end, we used Solove's taxonomy [12]. Although originating from American tort law, the taxonomy is significantly close to the GDPR to offer common ground for privacy threat specification.

Privacy Risk. The analysis of the privacy methodologies was conducted using six focus areas and aims at pointing out both areas of overlap and existing gaps. One challenge facing such an endeavour is that privacy engineering in IoT is still a new and developing area [8] with significant need for such frameworks [2]. Consequently, the focus areas used here are based on current best practice drawn from the analysed frameworks, as well as the need to develop a single integrated framework for IoT. In detail, we have: 1) *integrated* measuring how frame-

works can deal with both compliance driven and technical concerns; 2) *execution* stating whether a framework provides guidance on the implementation of its directives; 3) *auditing* stating whether audit is a focus area; 4) *modelling* reporting on the need of using specialised system models; 5) *IoT inclusive* describing whether an IoT specific approach is presented; finally 6) *Privacy focus* stating what privacy protection approach is adopted.

We decided to compare and assess the following framework and methodologies, based on the above criteria: MPRAM [7], LINDDUN [17], SDM [15], and CNIL IoT [5].

4 Privacy Evaluation

The privacy evaluation aims first at pointing out (i) *privacy threats driven by experimental analysis* according to Solove's taxonomy, and then (ii) *privacy framework principles* by comparing existing privacy frameworks.

4.1 Privacy Threat Identification

The experimental analysis of the Smart Home testbed yielded some concerning results. These not only relate to data flows but also privacy protection procedures, specifically with regards to obtaining informed consent from data subjects and compliance with the GDPR's transparency principle.

These privacy threats can be classified as per Table 2 according to the Solove's taxonomy.[3] Specifically, the testbed analysis revealed multiple instances of data reuse, violations of the purpose limitation and data aggregation. Notably, we have anal-

---

[2] The tesbed was run over three months during Autumn 2018.

[3] For the sake of presentation, we do not report the categories *Interrogation*, *Insecurity*, *Blackmail*, *Appropriation*, *Distortion* and *Decisional Interference* as they do not apply to our testbed.

ysed the testbed adhering to devices' functionality without injecting any adversarial behaviour. We will however, explore such options as part of our future work.

We list the following additional details on the reported privacy threat assessment, with high-level definitions form Solove's taxonomy in italics.

Surveillance - *Leveraging monitoring to trigger change or prevent human behaviour.* All the devices employed surveillance techniques either directly on data subjects (e.g. loose weight, walk more or using less electricity), their environment or any individuals in that environment. Notably, even Amazon Echo Plus can suffer from this threat as its functioning can affect the behaviour of any individuals in its vicinity.

Aggregation - *Collection of various data unrelated to processing purpose, and from multiple sources.* All data subjects consented to data aggregation during device setup. However, all devices collect additional data besides their intended purpose, sometime outside given consent (Misfit Shine) or under unclear requirements (e.g. location data for Withings Thermo).

Identification - *Collecting data that allows the identification of individuals.* All IoT control apps compelled data subjects to identify themselves to access and manage their already harvested data. A privacy dashboard enabling anonymised data management is not used.

Secondary Use - *Usage of data for a purpose for which data subjects are not informed.* All the devices aside from the TP Link Plug, fall foul of this requirement either through the harvesting of location data, device (phone) data or behavioural data on the user. However, this secondary use is not clearly stated to the data subject, either through omission (Misfit Shine) or by way of overly lengthy consent documentation (Amazon Echo Plus). Also noteworthy is the behavioural tracking, which relates to a dedicated tracking app recording user metrics, which was detected by the Lumen privacy monitor.

Exclusion - *Failure to provide data subjects with full and direct control over their data.* All devices allow data subjects some level of access to their data. However, the access to secondary use data, if any, is neglected. The mechanisms for exercising that control is also highly convoluted in some cases.

Confidentiality Breach - *Untrustworthy action on individual data by data controllers.* Although data aggregation and the involvement of third-parties can be shown, we have not found any clear instances of personally identifiable data being untrustworthy released or shared.

Disclosure - *Disclosure to data controllers of individuals data beyond purpose limitation without anonymisation.* Data

beyond purpose limitation are accessed by all devices, except for the TP Link Plug and the Amazon Echo Plus (respectively, no additional data is accessed or the consent is so broad as to cover most anything).

Exposure - *Exposing private emotional or physical truths.* Most of the device gather data of a deeply personal nature, e.g. medical data (Withings Body) or individual pictures (Withings Home). We have not detected unintended exposure of this data.

Increased accessibility - *Making data of an individual easily accessible to the public.* This is not contingent on the data being private, but solely to the ease of access of third parties. No instances of this were detected.

Intrusion - *The presence of monitoring devices in private situations.* All devices rely on constant monitoring to function. Generally speaking, disabling devices' functions make the devices themselves lacking of purpose in terms of IoT functionality and in most cases leaves the device inoperable. However, the TP Link Plug's smart features can be disabled in order to act like a normal electric plug.

Additionally, a specific note has to be made on the use of *tracking software* for crash and behavioural reports. For the former, all devices (except the TP Link Plug) uses third-party tracking software for anonymous crash reports. While, for the latter, all Whitings devices uses third party behavioural tracking software installed by the control app. Even if such tracking is fully anonymised, the other concerns around informed consent and exclusion still remain.

The findings from our testbed not only draw clear attention to the privacy challenges faced in the IoT sphere, but are also in keeping with the pervading understanding of these challenges [1]. These include, but are not limited to, the pervasive and always-on nature of devices, large scale data aggregation and the need to manage these challenges as they develop.

4.2 Privacy Framework Analysis

We report in this section the findings on the assessment of privacy frameworks introduced in Section 3. Table 3 summarises the results.

Feature	MPRAM	LINDDUN	SDM	CNIL
Integrated	Limited	No	Yes	Limited
Execution	No	Yes	Yes	Yes
Auditing	No	No	Limited	Yes
Modelling	Yes	Yes	No	No
IoT inclusive	No	No	No	Yes
Privacy focus	Security goals	Threats	PbD	PbD

Table 3: Assessment results of IoT privacy compliance frameworks (where PbD stands for Privacy-by-Design)

The primary focus for MPRAM is on security goals, hence considering privacy as a security need in the requirement analysis process. By following several refinement steps involving the modelling of privacy and security needs, MPRAM formalises integrated security and privacy goals. However, it does not explicitly address IoT, does not focus on auditing and does not dictate how security goals should be acted (executed) upon. As a result, risk driven concerns can be addressed while compliance ones cannot; we then deem it as partially integrated.

LINDDUN, like MPRAM, does not explicitly target regulatory compliance, related auditing or IoT implementation. Unlike MPRAM though, it relies on Data Flow Diagram (DFD) modelling to locate threats and determine the mitigating actions (execution) to take. A key strength of DFD-focused approaches, is the clear direction they provide for the implementation of remedial or preventative actions.

Differently from the previous, both SDM and CNIL directly target GDPR compliance and accordingly follow PbD as their privacy focus. Although the SDM seeks to provide a level playing field for service providers and regulators alike, it does not specifically address auditing nor IoT. CNIL on the other hand provides an assessment based framework and as such is specifically focused on auditing. CNIL also includes documentation on IoT compliance auditing but this does not extend to non-regulatory best practice or an integrated and iterative design process; thus we deem it as partially integrated.

5 Towards an IoT Data Protection Framework

To guide the formulation of an integrated IoT privacy framework, we propose a set of principles that aim at reconciling the compliance driven and technical requirement driven approaches pointed out in the previous sections, whilst incorporating a PbD approach. To this end, we adopt the standard practice to first find areas of overlap or commonality, and then to use these as a point of departure to address any existing gaps [11].

All the frameworks investigated, as well as the Solove's taxonomy, have as a critical prerequisite the need for understanding the proverbial *lay of the land*. Whether by conducting focus groups, deriving data flow diagrams or conducting an audit, all these approaches need to have clear input on what is being assessed or designed. Thus, an understanding of data movement is focal to ensure accurate and complete threat location, system analysis and compliance assessment. Our testbed analyses therefore directly informs the work at this point.

P1 - Data Flow. *The use of data flow diagrams in design and assessment phases, including machine to machine, human to machine and human to human communications.*

To ensure principled organisation of different design and deployment phases (e.g., from modelling of data flows and security goals, to description of regulatory compliance procedure),

there is an inherent need for a taxonomy. Although Solove's can be applied, it overlooks specific features of IoT and PbD, e.g. lack of threats on data subject rights and reference to the device-to-device communications of IoT systems.

P2 - IoT Privacy Taxonomy. *The need for a formalised and integrated taxonomy, inclusive of machine to machine communication and focused on IoT devices and systems.*

Privacy must be addressed not just in terms of static regulatory requirements but also in terms of developing best practices for IoT industry. As such, industry practices must start taking into account IoT specific features, e.g. device to device communication, data anonymisation and data aggregation risks. Here too, we see the testbed analyses directly informing our proposed methodology.

P3 - Privacy-by-Design Focus. *The need for iterative application of PbD principles for both device and system development, inclusive of both static regulatory requirements and developing industry best practice.*

Current frameworks differ in the approach to formal auditing. By taking inspiration from SDM and CNIL, it is essential to provide clear and upfront information to both IoT provider and compliance auditor.

P4 - Audit. *The use of an easily auditable framework both in terms of preparation and compliance checks.*

Besides auditing, there is a clear need for an implementation process. Vagueness on implementation directly counteracts the principle of producing devices and systems which are auditable for compliance. This also needs to balance the fast changing technology landscape with a relatively fixed regulatory regime.

P5 - Implementation. *The need for dictating the expected outcomes of a compliant system and how to approach iterative design in a compliant manner.*

Finally, there is the need for a single framework integrating both technical and compliance driven approaches, hence overcoming current requirements of using more than one.

P6 - Compliance and risk driven. *The need for integrating technical and compliance driven elements to ensure the adoption of a single integrated framework.*

Despite the high-level principles, we can already deduce their benefit by referring back to the privacy threats on the testbed. For instance, an IoT privacy taxonomy would shed light on more nuanced IoT privacy aspects including surveillance, data aggregation and individual identification carried out by third parties, which data subject may be unaware of.

6 Conclusions

Privacy-by-Design is a neglected aspect in the design of IoT devices and related services. While there are many guidelines and frameworks to implement Security-by-Design in IoT[4], there are few guidelines and methodologies on how to design IoT devices and services that comply with the principles of Privacy-by-Design.

In this paper, we have investigated which are the main privacy threats emerging in Smart Homes and then compared and assessed the extent to which existing methodologies would have helped in identifying these threats. Based on the analysis we have formulated a set of principles to guide the future development of a more effective methodology to implement PbD into IoT devices and services.

Future works. Firstly, we plan on widening the validation of the proposed principles by taking into account additional frameworks and taxonomies, e.g. those developed by The Open Web Application Security Project (OWASP). This will be further supported by retesting the devices on the testbed 12 months after the initial test. Thereby allowing us to gauge any impact of updates on the control apps, device firmware and policies and procedures. Consequently, the refined guiding principles can be used to develop, and experimentally validate, a new framework for engineering PbD into IoT devices and systems.

Acknowledgements

References

[1] Noah Apthorpe, Dillon Reisman, and Nick Feamster. A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic. *arXiv preprint arXiv:1705.06805*, 2017.

[2] Ian Brown. Britain's Smart Meter Programme: A Case Study in Privacy by Design. *International Review of Law, Computers & Technology*, 28(2):172–184, 2014.

[3] Abhik Chaudhuri and Ann Cavoukian. The proactive and preventive privacy (3p) framework for iot privacy by design. *EDPACS*, 57(1):1–16, 2018.

[4] CNIL. CNIL publishes an update of its PIA Guides. Technical report, 2018. https://www.cnil.fr/en/cnil-publishes-update-its-pia-guides.

[5] CNIL. Privacy Impact Assessment Application to IoT Devices. Technical report, Commission Nationale de l'Informatique et des Libertés, Paris, 2018.

[6] European Union. Regulation 2016/679 of the European parliament and the Council of the European Union, 2016.

[7] Seda Gürses, Bettina Berendt, and Thomas Santen. Multilateral Security Requirements Analysis for Preserving Privacy in Ubiquitous Environments. In *Proceedings of the UKDU Workshop*, pages 51–64, 2006.

[8] Charith Perera, Mahmoud Barhamgi, Arosha K. Bandara, Muhammad Ajmal, Blaine Price, and Bashar Nuseibeh. Designing Privacy-aware Internet of Things Applications. *arXiv preprint arXiv:1703.03892*, pages 1–35, 2017.

[9] The Haystack Project. The ICSI Haystack Project, 2017. https://www.haystack.mobi/.

[10] Irina Ioana Pătru, Mihai Carabaş, Mihai Bărbulescu, and Laura Gheorghe. Smart home IoT system. In *Networking in Education and Research: RoEduNet International Conference 15th Edition, RoEduNet 2016 - Proceedings*, pages 1–6, 2016.

[11] Allen F. Repko. *Interdisciplinary Research*. Sage Publications Inc, London, 2008.

[12] Daniel J. Solove. A Taxonomy of Privacy. *U. Pa. L. Rev.*, 154:477–560, 2006.

[13] Emmeline Taylor and Katina Michael. Smart Toys that are the Stuff of Nightmares [Editorial]. *IEEE Technology and Society Magazine*, 35(1):8–10, 2016.

[14] Robert Thorburn. Setting up a raspberry pi based iot testbed for traffic analysis, 2018. https://medium.com/cybersoton/.

[15] Unabhängiges Landeszentrum für Datenschutz. The Standard Data Protection Model. Technical report, 2016.

[16] Jacob Wurm, Khoa Hoang, Orlando Arias, Ahmad-Reza Sadeghi, and Yier Jin. Security analysis on consumer and industrial IoT devices. In *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 519–524, 2016.

[17] Kim Wuyts. LIND(D)UN privacy threat tree catalog. Technical Report September, 2014. http://www2.cs.kuleuven.be/publicaties/rapporten/cw/CW675.pdf.

---

[4] Such as the Department for Digital, Culture, Media and Sport's Secure by Design Guidance: https://www.gov.uk/government/publications/secure-by-design