On the Design of a Privacy-Centered Data Lifecycle for Smart Living Spaces

Joseph Bugeja, Andreas Jacobsson

On the Design of a Privacy-Centered Data Lifecycle for Smart Living Spaces

Joseph Bugeja and Andreas Jacobsson

Internet of Things and People Research Center, Department of Computer Science and Media Technology, Malmö University, Sweden
{joseph.bugeja,andreas.jacobsson}@mau.se

Abstract. Many living spaces, such as homes, are becoming smarter and connected by using Internet of Things (IoT) technologies. Such systems should ideally be privacy-centered by design given the sensitive and personal data they commonly deal with. Nonetheless, few systematic methodologies exist that deal with privacy threats affecting IoT-based systems. In this paper, we capture the generic function of an IoT system to model privacy so that threats affecting such contexts can be identified and categorized at system design stage. In effect, we integrate an extension to so called Data Flow Diagrams (DFD) in the model, which provides the means to handle the privacy-specific threats in IoT systems. To validate the model, we apply it to the design of a realistic use-case involving Facebook Portal. We use that as a means to elicit the privacy threats and mitigations that can be adopted therein. Overall, we believe that the proposed extension and categorization of privacy threats provide a useful addition to IoT practitioners and researchers in support for the adoption of sound privacy-centered principles in the early stages of the smart living design process.

Keywords: IoT · data lifecycle · data flow diagrams · data privacy · privacy threats · smart connected home · smart living space · Facebook Portal.

1 Introduction

IoT products are widely being deployed enabling the development of new applications. By combining the input of environmental and user activity information with intelligent algorithms, activation mechanisms, and information feedback, traditional living spaces have been enhanced to smart living spaces. Smart living spaces involve different connected devices and services bringing up benefits such as improved convenience and experience for the users, optimized power efficiency, elderly telemonitoring, etc. [33]. The application of the IoT is broad, ranging from daily personal home applications to industrial automation applications or city transportation [13].

Despite their benefits, IoT technologies implemented in the home environment tend to generate and process a diverse amount of sensitive and personal

data from users and making such data accessible to different entities almost instantaneously from anywhere there is an Internet connection. Data typically includes the geographical position of individuals, movement patterns, and sensed data that may reveal the physical conditions, behaviors, and activities of individuals. Improper usage of such data can lead to undesired privacy harms to an individual, group, and society. Some examples of consequences that can impair the users of IoT technologies include unsolicited advertisements, identity theft, discrimination, and more [5].

Our right to privacy as a concept is not a new idea, as early as 1890, Warren and Brandeis described privacy as the "the right to be let alone" [30]. They identified it as the right that enables individuals to have personal autonomy, freedom of association, moments of reserve, solitude, intimacy, and independence. Westin, as the computer era was emerging, described privacy as the "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [31]. While Westin's notion of privacy has been expanded upon (e.g., by Altman [2]) privacy remains a nebulous concept which is not amenable to precise definition [18]. This makes it especially challenging to enable a focused discussion about privacy in the IoT. In this paper, we focus on "information privacy" – a term that we adopt to signify the ability of users to exercise control over personal data about themselves while also minimize future privacy risks by protecting data after it is no longer under a user's direct control [28]. With the pervasiveness of IoT deployments it is essential to identify recent privacy threats and their related countermeasures.

Privacy concerns have motivated the development of legal regulation, such as the European General Data Protection Regulation (GDPR) [17], standards and frameworks for governing the processing of personal data, e.g., the privacy framework (ISO/IEC 29100:2011) [21], and as well engineering approaches such as "Privacy by Design (PbD)" [8] that aim to safeguard privacy since the beginning of the development process. Accordingly, software engineers are increasingly expected to give appropriate consideration to privacy and data protection issues throughout the development lifecycle [22]. However, the mentioned approaches alone are insufficient raising the need for a more proactive and integrative approach to safeguard the privacy of data subjects (i.e., the natural person to whom the personal data belongs to) [1]. In particular, tools are needed to better assist IoT practitioners to address the complexity and variability of privacy issues throughout the development process.

To this aim, in this paper we provide an understanding of: i) what are the information privacy threats affecting a smart living space; ii) which data lifecycle phase each privacy threats affects; and iii) how can a privacy-preserving data lifecycle be modeled to reduce the exposure of end-users to such threats. Overall, the main contribution is an extension to a standard system modeling technique – Data Flow Diagrams (DFD) – with new processes and annotations that can be used to incorporate privacy principles into the software design process. In summary, our contribution benefits IoT practitioners by helping bridge

Fig. 1. DFD element types.

the gap between the technical design and concerns related to privacy compliance; and support the adoption of privacy principles in the early stages of the design process.

1.1 Data flow diagrams

DFDs are a visual notation that allows to model data flows in information systems in a structured way [29]. Graphically, the main DFD elements are represented in Figure 1, and their basic elements are:

- *External entity*: Represents a person, organization, or services that are external to the system but interact with it.
- *Process*: Represents an activity or a function, e.g., updating user profile or consulting a cloud endpoint, that is performed for some specific business purpose.
- *Complex process*: These are a logical representation of a process, e.g., a mobile application or web application, that performs many distinct operations and thus can be represented or refined through separate DFDs.
- *Data store*: A collection of data that are at rest, e.g., databases, files, or cache information.
- *Data flow*: These are a single piece of data or a logical collection of several pieces of information that are being communicated.
- *Trust boundary*: Represents the border between untrustworthy and trustworthy elements or a delineation between data moving from low to high trust and vice versa.

DFDs are widely used during the system analysis phase to capture the requirements of a software system including the identification of security threats (e.g., in STRIDE) and privacy threats (e.g., in LINDDUN) [10].

1.2 IoT entities, processes, and data flows

An IoT system can be viewed as a dynamic and distributed networked system, capable of producing and consuming information [13]. At a high-level, it consists of connected devices (e.g., sensors), services (e.g., cloud-based analytics software), network infrastructure (e.g., protocols such as Bluetooth), and users (e.g., the smart living space inhabitants).

For an IoT system, the external entities can be generally grouped into data subject, data controller, and data user [32][24][15]. Data subject is the human entity that generates the original raw data and whose personal data are processed by the IoT system. Data controller (sometimes also referred to as "data holder," "data curator," or "data processor") is the person or organization that collects, stores, processes, and releases the data. Data users represent the entities that access the released data. Typically, data users are the data subjects however they can also be other devices or systems. At a general level, there are four main data phases in an IoT system [34][6][20]:

- *Data generation*: Represents the activity where the data subject interacts with the IoT system, directly or indirectly, to create personal data. Personal data are any information that is related to an identified or identifiable natural person. This interaction is typically, done through end-user devices such as smartphone applications with the help of services.
- *Data collection*: Collection represents the act of acquiring personal data from data subjects, including external sources. Typically, this is done through sensors embedded inside the smart device. Information at this stage may be stored in the IoT system. The longevity of this may range from temporary (transient storage), e.g., in memory buffers, to persistent (persistent storage), e.g., inside databases; and may occur automatically without involving directly the end-user.
- *Data processing*: The IoT analyses the data stored in the cloud data centers or inside the devices to provide the smart services. Due to shifting levels of autonomy, IoT devices are typically capable of making some decisions on their own, i.e., without human intervention. There are different processing models, including: cloud-centric, gateway-centric, and edge-centric data processing models.
- *Data disclosure*: Represents the act of disseminating, making available or transmitting personal data for external use by third-parties. Commonly, when the disclosure is done to the data user, this phase is typically referred to as data presentation. The output from this phase tends to range from notification to actuation.

This IoT data lifecycle, is also accompanied by a separate lifecycle, where the actual (physical) IoT device is initially deployed, then operated, and finally retired (decommissioned). However, for this study, we focus on the logical, i.e., the data lifecycle, as this corresponds to "information privacy".

1.3 Organization of the paper

In Section 1, we provide an introduction of the research done, propose DFDs as a modeling tool, and describe the IoT data phases. Next, we provide a background overview and summarize the relevant related work. In Section 3, we identify the privacy protection goals, exemplify, and categorize the main IoT privacy threats directed to smart living spaces. Then, in Section 4, we provide a privacy-centered data lifecycle for smart living space applications using the proposed

extension to DFDs. The proposed additions are applied in Section 5 to a smart connected home setup using Facebook Portal as its smart living device. In Section 6, we discuss how the proposed extensions can help in addressing some of the privacy compliance requirements. Finally, in Section 7, we conclude this paper and identify some avenues for future work.

2 Background and related work

This work lies at the intersection of the research areas of privacy by design strategies, privacy threat modeling, and privacy threat analysis.

Privacy by design strategies. This is a design approach that aims to improve the overall privacy friendliness of IT systems [19]. PbD seeks to encourage the inclusion of privacy at the start rather than retrofitted into existing systems [8].

PbD was originally presented by Cavoukian and consists of seven principles including: "privacy as the default," "full functionality," and "privacy embedded into design" [9]. For example, users should not have to opt-in to receive data protection. Langheinrich [23] also proposed six principles for PbD. These included "choice/consent," "anonymity/pseudonymity," and "access/recourse". They were suggested for ubiquitous systems, but have found implementation in a range of fields. Recently, ENISA [11] summarized eight PbD strategies as derived by Hoepman [19] and applied them to the context of big data analytics. The strategies are: "minimize," "hide," "separate," "aggregate," "inform," "control," "enforce," and "demonstrate."

Privacy threat modeling. These approaches contribute to the realization of PbD strategies by providing a systematic, rigorous, and methodical approach towards privacy analysis.

LINDDUN [14] threat modelling framework uses a DFD to identify and model privacy threats. Privacy threats addressed by LINDDUN are: linkability, identifiability, non-repudiation, detectability, disclosure of information, unawareness, and non-compliance. Luna et al. [25] proposed QTMM (Quantitative Threat Modeling Methodology) to help draw objective conclusions about privacy-related attacks. It follows the same modeling steps as LINDDUN however it focuses on three privacy-specific threat categories: linkability, unawareness, and intervenability. Antignac et al. [3] proposed an extension to DFDs to support the inclusion of privacy concepts borrowed from the EU GDPR and ISO 29100 standard. This conceptual model is useful for making privacy concepts in the design model explicit.

Privacy threat analysis. Different scholars have proposed different taxonomies and catalogues to help identify and understand privacy threats.

The Antón-Earp taxonomy [4] identifies five privacy protection goals, e.g., "choice/consent," to safeguard the privacy of a customer's data, and seven corresponding categories of vulnerabilities, e.g., "information aggregation," that reflect a potential privacy violation. Solove's taxonomy [27] is a taxonomy of privacy harms that groups privacy threats, e.g., "surveillance," occurring at dif-

ferent system phases: information collection, information processing, information dissemination, and including invasions to the data subject. Ziegeldorf et al. [34] identified seven categories of privacy threats that affect an IoT-based system. Three examples of threats covered in [34] are identification, tracking, and profiling.

Main observations. In reviewing the existing work, we observe that most of the mentioned approaches have not evolved to cater for IoT technologies. For instance, the Antón-Earp taxonomy [4] and Solove's taxonomy [27] while generic enough were created before the advancement of IoT technologies. This could mean that these taxonomies may be more applicable for studying web-based or traditional information systems but not necessarily IoT-based systems. Similarly, LINDDUN has not been validated on IoT systems [26]. Lastly, PbD while useful as an approach, there are still uncertainties about what it means in the context of IoT and especially how it can be implemented.

Our contribution. Given that the focus of this paper is on IoT-based systems, we leverage the literature of Ziegeldorf et al. in [34] to identify privacy threats in smart living spaces. Different to that study, we organize the threats identified in that study according to the privacy protection goals being violated and identify corresponding data phases leading to those threats. This is similar to Solove's taxonomy [27] but with IoT specific phases. Furthermore, when it comes to modeling privacy we follow a similar conceptual modeling approach to Antignac et al. [3] but we focus on IoT-based systems and also propose different extensions and processes. These target both the data controller and data subject. Additionally, we suggest PbD strategies for mitigating the identified privacy threats at the different data lifecycle phases.

3 IoT privacy goals and threats

In order to safeguard end-users right to privacy, three main privacy-specific protection goals have been proposed that articulate what is being protected and from who. These goals are: unlinkability, transparency, and intervenability [25][12][35].

Unlinkability is defined as the property that data processing is operated in such a way that the privacy-relevant data cannot be linked to any other set of privacy-relevant data outside of the domain. Transparency ensures that privacy-relevant data processing can be understood and reconstructed at any time. Intervenability ensures that the parties involved in any privacy-relevant data processing, including the data subject, have the possibility to intervene where necessary.

To the above, we also add the goals of confidentiality and detectability. Confidentiality represents the goal of preventing unauthorized access to information or systems. Detectability corresponds to the goal of preventing an attacker from sufficiently distinguishing if an item of interest exists or not. While these goals are somewhat related to the unlinkability goal, we added them separately since some IoT privacy threats are primarily affecting those.

Table 1. Summary of the IoT privacy threats alongside the main privacy protection goal being violated by each and the corresponding data lifecycle phase during which each threat typically occurs. The symbol: ● indicates that the threat occurs often; ◐ indicates that the threat might occur; and ○ indicates that the threat rarely occurs.

Information privacy threats	**Protection goals**	**Data generation**	**Data collection**	**Data processing**	**Data disclosure**
Identification	Unlinkability	◐	◐	●	○
Localization and tracking	Unlinkability	◐	◐	●	○
Profiling	Unlinkability	○	◐	◐	●
Linkage	Unlinkability	○	◐	◐	●
Privacy-violating interaction and presentation	Confidentiality	●	○	○	●
Inventory attacks	Detectability	○	●	○	○
Lifecycle transitions	Transparency	○	●	○	○

Based on the work of Ziegeldorf et al. [34] and our observations, below we present a summary of IoT privacy threats applied to smart living systems and grouped according to the corresponding primary protection goal being violated by each. Table 1 outlines the different IoT privacy threats.

Unlinkability threats

– *Identification*: Identification characterizes the risk of associating a persistent identifier, e.g., name and address, with a data subject and thus revealing the identity of the individual [34]. Typically, all IoT control apps compel data subjects to identify themselves, e.g., by entering account details during system setup or dynamically through technologies such as facial recognition. This threat is dominant in the data processing phase but may also occur at the data generation and data collection phase [34].

– *Localization and tracking*: This threat allows for the recording of a person's location [34]. Different monitoring techniques are employed by IoT devices, e.g., built-in motion sensors, and as well reading the location directly from the smartphone. This threat is dominant in the data processing phase but may also occur at the data generation and data collection phase [34].

– *Profiling*: Represents the threat of collecting and correlating information about individual activities in order to generate new information from the original data [34]. For instance, a smart thermostat can collect temperature, humidity, and ambient light data of the location where it is being used and then makes temperature adjustments for different situations accordingly. This may then be used to automatically infer that a person is home at a certain time and consequently offer targeted advertisements, e.g., food

and drink adverts. Profiling threats mostly appear in the disclosure phase where information is forwarded to third-parties [34]. Nonetheless, these may also occur during data collection and processing.

– *Linkage*: This threat consists in linking different separated systems such that the combination of data sources reveals information that the subject did not disclose or intended to [34]. As an example, if an IoT device, such as a smart lock, is integrated with another device from a different manufacturer, e.g., a connected doorbell, through a cloud-based service such as IFTTT (if this then that), then the doorbell might acquire information, e.g., about successful/unsuccessful attempts, from the smart lock that it was not originally intended to process. This threat of linkage primarily appears in the data disclosure phase [34]. However, similar to profiling, it may also occur during data collection and processing.

Confidentiality threats

– *Privacy-violating interaction and presentation*: Exposing personally identifiable information to individuals who are not supposed to have access to it [34]. As an example, information may be disclosed through a public medium, e.g., smart speaker system, to an unwanted audience, e.g., to temporary visitors in a smart home. This threat is dominant in the data generation and disclosure phase, in particular when information is presented to the users [34].

Detectability threats

– *Inventory attacks*: These refer to the unauthorized collection of information about the existence and characteristics of personal things [34]. As an example, given the wireless nature of most of the smart devices, a malicious threat agent may deduce the presence of a certain medical device, e.g., insulin pump, and thereby inferring that a person is suffering from a certain medical condition, e.g., diabetes. This threat primarily occurs at the data collection phase [34].

Transparency threats

– *Lifecycle transitions*: Occurs when users' private information collected during the IoT device's lifetime is disclosed during changes to the device's control spheres in their lifecycle [34]. As an example, sensors may collect private information about a user, then transmit it to the cloud for further analysis and returning the result. This transfer of data between different phases may disclose sensitive information about the user. This threat primarily occurs at the data collection phase [34].

4 Privacy-centered smart living data lifecycle

In order to mitigate the identified threats, various mitigations have been proposed in scholarly literature, e.g., [19], and industry reports, e.g., [11].

For the unlinkability threats, we find in particular data-oriented strategies such as minimize and hide. Minimize, example by adopting a select-before-collect approach, reduces the amount of processed personal data to the minimal amount possible [19][11]. Hide, for example achieved through data encryption, anonymization, or attribute-based credentials, reduces personal data and their interrelations from plain view [19][11]. It is also useful as a strategy to reduce detectability threats. For the transparency threat, we refer to process-oriented strategies with protection strategies such as inform and control. As a strategy, the inform process notifies the data subjects whenever personal data are processed [19][11]. Acting as a counterpart to the inform strategy, is the control strategy that gives the data subjects agency over the processing of their personal data [19][11]. The inform strategy is also useful for mitigating the confidentiality threat. Additionally, the data-oriented strategy aggregate, is beneficial here. Aggregate, for example achieved through data provenance, can help restrict the amount of detail in the personal data that remains, e.g., before being sent or published as part of the data disclosure phase [19][11].

Data subject controls. Data subjects are the entities responsible for the data generation phase. For data subjects, new processes have to be added to offer them the capability of accessing, reviewing, and destroying stored personal data. Data access represents the act of specifying, retrieving or consulting personal data values that are stored. Data review signifies the act of implementing the access right and rectifying personal data values by data subjects to ensure that their data is accurate, complete, and up-to-date. Data destroy characterizes the act of erasing, redacting, or disposing of personal data.

Furthermore, at the different lifecycle phases, it would be appropriate to have mechanisms in place that allow the data subjects the facility to be informed about data collection and use, and to have control over that. Overall, the successful implementation of these processes reduce the threat of identification and tracking, and as well confidentiality threats.

Data controller controls. Data controllers are the entities responsible for the data collection, data processing, and data disclosure phase. For the data controllers, especially, at the data collection phase, the practice of data minimization is core to reduce personal data from being inappropriately disclosed to unauthorized entities. Strategies should be designed to minimize the amount of data types collected or requested by an IoT application, the volume and granularity of data stored by an IoT application, and raw data acquired by the system. The effective implementation of data minimization reduces the overall effect of the different IoT privacy threats.

Furthermore, when personal data need to be transmitted to the different external entities, in particular to data users as part of the data disclosure phase,

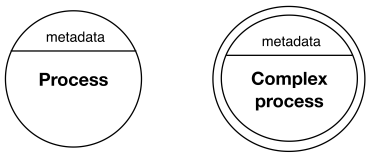

Fig. 2. A DFD extension to the process and complex process symbol. The metadata may represent a purpose statement indicating a reason for a process to collect and use personal data.

these should be sent encrypted, ideally end-to-end starting from the data subject, in the network, and then off to the service provider backend infrastructure. Encrypted data communication reduces the potential privacy risks due to unauthorized access during data transfer between components. Encryption can also be applied when storing and processing data to reduce privacy violations due to malicious attacks and unauthorized access to personal data. Moreover, instances of personal data, especially if these are to be shared with another data controller, should ideally be transformed at source to anonymize the identity of the source or target device or user. Implemented properly these controls help reduce detectability and unlinkability threats.

DFD privacy-centered extensions. To capture the aforementioned tactics in the actual IoT system design, the identified data lifecycle phases have to be accordingly modified. With this, the IoT practitioner would be offered the possibility of introducing additional processes before and after each data lifecycle phase. This is to mitigate the different threats identified at each phase and to empower the data subject with control, choice, and flexibility over the processing of personal data.

We do so by extending the standard DFD notation with an annotation as shown in Figure 2. The annotation allows for the direct specification of metadata. Through this, IoT practitioners can explicitly declare the purpose for collecting, processing, and disclosing the personal data of data subjects, and likewise to specify the duration for which personal data are stored in the smart living application. Overall, the different processes intended for the data subject and data controller are depicted in Figure 3.

Collectively, when the proposed privacy protection measures are applied to the data collection, data processing, and data disclosure phases, we refer to these processes as: secure data collection, secure data processing, and secure data disclosure, respectively. These can be represented as complex processes; that are in the realm of the data controller; with privacy-preserving data being their output data flow. In terms of the data subject processes we aggregate these and represent them as a complex process – privacy manager. Essentially, this can be compared to an end-user privacy toolbox but it can also take the form of a separate physical device that is fully controllable by the data subject.

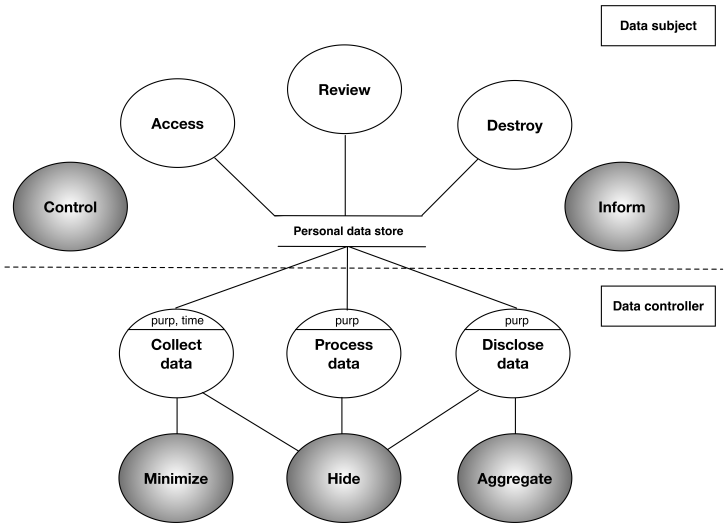

Fig. 3. Securing the IoT data lifecycle against IoT privacy threats. The data subject has the possibility to access, review, and remove any personal data retained by the system, and subsequently to get informed and have control over data lifecycle processes. The data controller updates the IoT data lifecycle processes with functions to minimize, hide, and aggregate information, and to specify a purpose (*purp*) for each corresponding phase and duration (*time*) for data retention. Processes highlighted in gray colour represent PbD strategies.

5 Application of the privacy-centered smart data lifecycle

Let us assume a smart connected home setup consisting of Facebook Portal (Portal)[1] as the main IoT home device. Portal is primarily a smart screen and speaker device embedded with cameras, microphones, and Artificial Intelligence (AI), allowing for video-calling and advanced entertainment support.

Privacy threat identification. Using the previously established map of possible IoT privacy threats as detailed in Table 1 as a guide, together with Portal's privacy policy[2] and its feature list as supporting documentation, we identify a number of possible privacy threats in Portal.

Identification is a potential threat since an account is needed to be able to benefit from Portal. Tracking is a possible threat as the data subject's location can be inferred from the smartphone device when enabled. Linkage is a potential threat especially since Amazon Alexa account can be connected to Portal for voice-based interactions. Privacy-violating interaction and presentation is a threat that is common to Portal and similar device types, in particular as photos

[1] https://portal.facebook.com [accessed December 13, 2019].
[2] https://portal.facebook.com/legal/data-policy [accessed December 13, 2019].

and videos may be pulled from Facebook and perhaps displayed unintentionally to guests or temporary visitors inside the smart living space. Inventory attacks may be a threat if an attack is for instance able to detect Portal's presence, e.g., by observing the device's response to a specific wake word or for instance by detecting its unique fingerprint. Profiling may be an eventual threat especially if the device is in future integrated with other connected systems such as WhatsApp and Instagram.

Privacy threat mitigation. In order to mitigate the identified privacy threats, we evolve the data lifecycle into a privacy-centered data lifecycle using the processes identified in Section 4. This version uses the secure processes: privacy manager, secure data collection, secure data processing, and secure data disclosure, which implemented properly reduce the identified privacy threats discussed.

More concretely, the privacy manager can be deployed to the data subject as a user-friendly application available over a smartphone. This application, might for instance give the user the access rights to view or delete their voice history, update applications permissions, and get notified about how Facebook (and Amazon for the voice-processing) use their data. Likewise, it could offer users ways to control, e.g., disable collection of personal data, when needed.

When it comes to the data controller, the secure data collection process might for instance refrain from asking data subjects certain particulars, for instance, about the geographical data where the device is installed in; secure data processing process might store data inside the actual device encrypted using a strong and approved cryptographic algorithm; and secure data disclosure process might obscure the real identity of the data subject before transmitting it, e.g., alongside the search criteria, to Amazon.

Graphically, the privacy-centered data lifecycle as applied to Portal is depicted in Figure 4. Here, the proposed secure processes are indicated in the actual DFD, including being annotated with metadata indicating a specific purpose (e.g., collecting location data to only get calls when users are at home) and retention time (e.g., seven days) for processing activities.

6 Discussion

The successful implementation of the privacy-centered data lifecycle together with its secure processes helps data subjects attain different rights over their personal data. Based on the privacy principles and regulations of data processing in ISO/IEC 29100:2011 and GDPR, we outline the data subject's rights that the model is grounded upon:

- *Right to be forgotten:* The data subject shall have the right to obtain from the data controller the erasure of personal data concerning him or her without undue delay and the data controller shall have the obligation to delete personal data [GDPR Article 17, Clause 1; Individual participation and access (ISO/IEC 29100:2011)]. In our case, this right is implemented through

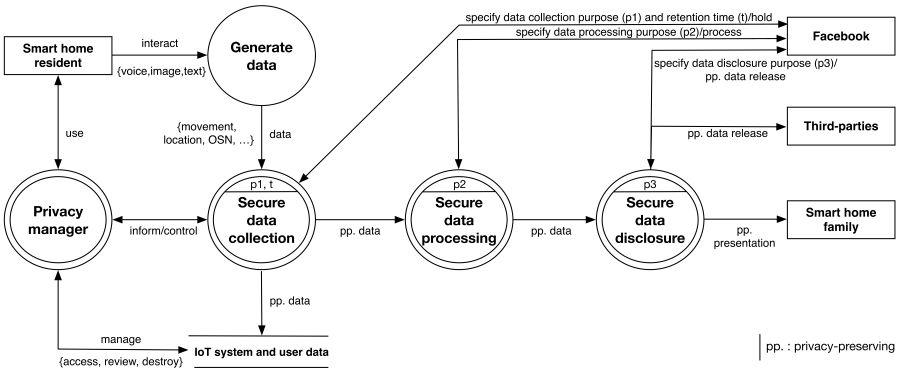

Fig. 4. Securing the IoT data lifecycle against IoT privacy threats using Facebook Portal as a use-case. As data is traveling across the different phases it is being outputted in a privacy-preserving (*pp*) manner. At the same time, the smart home residents can manage their collected personal data, and can get notified about it.

the privacy manager, specifically through the data destroy process, and by having the retention period for data attributes specified explicitly in the secure data collection phase.

- *Right to restrict processing:* The data subject shall have the right to obtain from the data controller restriction of processing where one of the following applies: (a) the accuracy of the personal data; (b) the processing is unlawful; (c) the data controller no longer needs the personal data; (d) the data subject has objected to processing [GDPR Article 18, Clause 1; Use, retention and disclosure (ISO/IEC 29100:2011)]. In our case, this right is implemented through the privacy manager, through the data review process, and by having the data processing purpose specified explicitly as metadata in the secure data processing phase.
- *Right to data portability:* The data subject shall have the right to receive the personal data concerning him or her, that it provided to a data controller and has the right to transmit that data to another controller without hindrance [GDPR Article 20, Clause 1; Individual participation and access (ISO/IEC 29100:2011)]. In our case, this right is implemented through the privacy manager, through the data access process.
- *Right to object:* Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her [GDPR Recital 32; Consent and choice (ISO/IEC 29100:2011)]. In our case, this right is implemented through the privacy manager, through the inform and control strategy.

Through the use of the privacy-preserving processes and annotations, the choices of the IoT designer with regards to privacy are made more explicit in the design model. Nonetheless, in practice this needs to be accompanied by sound

governance processes, standards, and practices that ensure that data controllers abide with privacy regulations, and thus safeguard individual's privacy rights accordingly. Here, we note that the GDPR, the forthcoming ePrivacy Regulation (ePR) [16], and the recent California Consumer Privacy Act (CCPA) [7], are key examples of good privacy regulations in this regard. Nevertheless, the fact that there are multiple regulations with overlapping sections raises the need for an international standard and consolidated guidelines for achieving data privacy.

When it comes to threat modeling DFDs provide a straightforward tool for doing so, but there are other threat modeling techniques that can be useful. For instance, Activity Diagrams (ADs) which are part of the Unified Modeling Language (UML) diagrams in UML 2.0. ADs are arguably more expressive than DFDs while retaining similar functionality towards threat modeling. Their expressiveness lies in the fact that they have a guard condition for the activity element, while DFD have no counterpart for this element. However, these extra elements while useful can make a diagram unnecessary complicated possibly resulting in a failure to identify threats in the system. Moreover, given that DFDs are focused on data, they provide a convenient choice for privacy analysis. Nonetheless, it would be useful to have a thorough evaluation of the proposed extensions. This could for instance be done by means of different empirical studies taking the perspective of requirements engineers and software architects.

Moving towards the system implementation phase, a data controller may leverage different technologies to realize the proposed mitigations. For instance, to address the threat of identification, localization and tracking, profiling, and linkage, data anonymization techniques or protocols, including data obfuscation, data encryption, data masking, and de-identification can be used. On the other hand, for reducing the threat of privacy-violating interaction and presentation, inventory attacks, and lifecycle transitions, access control with user-defined privacy policies or increasing user awareness are practical solutions. Additionally, there may be other PbD strategies that may be relevant depending on the particular context or use case.

7 Conclusions and future work

IoT devices have brought added efficiencies and conveniences to smart living users. Nonetheless, connected devices bring unprecedented privacy threats to users.

Recognizing this, we categorized and exemplified the different privacy threats affecting smart living spaces occurring at each data lifecycle phase. Noting the threats, we provided an extension to DFDs and a selection of PbD strategies to handle privacy-specific threats in IoT systems. Through the proposed privacy-centered lifecycle, a data controller can better plan in implementing privacy measures early-on in the software development process; and data subjects are empowered with improved control over their personal data.

For future work, it would be useful to develop a tool that provides IoT practitioners the facility to automatically elicit the privacy threats arising at each

data lifecycle phase and how each can be mitigated through the use of the newly identified secure processes. This helps evaluate the presented DFD extensions but also from a privacy compliance perspective for instance serving as evidence that privacy measures have been thought through. Finally, it would be beneficial to develop an approach that can quantitatively measure the privacy risk exposure of an IoT-based system to different malicious threat agents. These can be represented as a new type of external entity, e.g., data privacy attacker, that aims to compromise the smart living privacy requirements. Possibly, after the system is modeled, the privacy risk exposure can be automatically calculated based on the agent's capabilities and the system's vulnerabilities.

Acknowledgments. This work has been carried out within the research profile "Internet of Things and People," funded by the Knowledge Foundation and Malmö University in collaboration with 10 industrial partners.

References

1. Alshammari, M., Simpson, A.: Privacy architectural strategies: An approach for achieving various levels of privacy protection. In: Proceedings of the 2018 Workshop on Privacy in the Electronic Society. pp. 143–154. ACM (2018)
2. Altman, I.: The environment and social behavior: Privacy, personal space, territory, and crowding. (1975)
3. Antignac, T., Scandariato, R., Schneider, G.: A privacy-aware conceptual model for handling personal data. In: International Symposium on Leveraging Applications of Formal Methods. pp. 942–957. Springer (2016)
4. Antón, A.I., Earp, J.B.: A requirements taxonomy for reducing web site privacy vulnerabilities. Requirements Engineering **9**(3), 169–185 (2004)
5. Bettini, C., Riboni, D.: Privacy protection in pervasive systems: State of the art and technical challenges. Pervasive and Mobile Computing **17**(PB), 159–174 (2015)
6. Bugeja, J., Jacobsson, A., Davidsson, P.: An Empirical Analysis of Smart Connected Home Data pp. 1–15 (2018)
7. California Senate Judiciary Committee et al.: California consumer privacy act: Ab 375 legislative history (2018)
8. Cavoukian, A.: Privacy by Design. Tech. rep. (2009), `http://www.ontla.on.ca/library/repository/mon/23002/289982.pdf`
9. Cavoukian, A.: Privacy by design in law, policy and practice. A white paper for regulators, decision-makers and policy-makers (2011)
10. Chen, Y.T., Huang, C.C.: Determining information security threats for an iot-based energy internet by adopting software engineering and risk management approaches. Inventions **4**(3), 53 (2019)
11. D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y.A., Bourka, A.: Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics. arXiv preprint arXiv:1512.06000 (2015)
12. Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.H., Metayer, D.L., Tirtea, R., Schiffner, S.: Privacy and data protection by design-from policy to engineering. arXiv preprint arXiv:1501.03726 (2015)
13. Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, I.C.: Internet of Things: Vision, application areas and research challenges **10**, 1497–1516 (2012)

14. Deng, M., Wuyts, K., Scandariato, R., Preneel, B., Joosen, W.: A privacy threat analysis framework : supporting the elicitation and fulfillment of privacy requirements pp. 3–32 (2011)
15. Dwork, C., Roth, A.: The Algorithmic Foundations of Differential Privacy. Foundations and Trends® in Theoretical Computer Science **9**(3-4), 211–407 (2014)
16. European Commission: Proposal for a Regulation on Privacy and Electronic Communications (ePrivacy Regulation) (2017), `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017PC0010`
17. European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. Tech. rep. (2016), `https://bit.ly/2Cxy5yP`
18. Friedewald, M., Wright, D., Gutwirth, S., Mordini, E.: Privacy, data protection and emerging sciences and technologies: towards a common framework. Innovation–The european journal of social science research **23**(1), 61–67 (2010)
19. Hoepman, J.h.: Privacy Design Strategies pp. 446–459 (2014)
20. Hu, F., Jeyanthi, N.: Internet of Things (IoT) as Interconnection of Threats (IoT). In: Security and Privacy in Internet of Things (IoTs) (2016)
21. ISO: ISO 29100 Privacy Framework **2011**, 1–21 (2011)
22. Jacobsson, A., Boldt, M., Carlsson, B.: A risk analysis of a smart home automation system. Future Generation Computer Systems **56**, 719–733 (2016)
23. Langheinrich, M.: Privacy by design—principles of privacy-aware ubiquitous systems. In: International conference on Ubiquitous Computing. pp. 273–291. Springer (2001)
24. Li, C., Palanisamy, B.: Privacy in Internet of Things: from Principles to Technologies. IEEE Internet of Things Journal pp. 1–18 (2018)
25. Luna, J., Suri, N., Krontiris, I.: Privacy-by-design based on quantitative threat modeling. In: 2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS). pp. 1–8. IEEE (2012)
26. Perera, C., Mccormick, C., Bandara, A.K., Price, B.A., Nuseibeh, B.: Privacy-by-Design Framework for Assessing Internet of Things Applications and Platforms (2016)
27. Solove, D.J.: A taxonomy of privacy. U. Pa. L. Rev. **154**, 477 (2005)
28. Spiekermann, S., Cranor, L.: Privacy engineering. IEEE Transactions on Software Engineering **35**(1), 67–82 (2009)
29. Tao, Y., Kung, C.: Formal definition and verification of data flow diagrams. The Journal of Systems and Software **16**(1), 29–36 (1991)
30. Warren, S.D., Brandeis, L.D.: The right to privacy. Wadsworth Publ. Co. (1985)
31. Westin, A.F.: Privacy and freedom. Washington and Lee Law Review **25**(1), 166 (1968)
32. Yu, S.: Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data. IEEE Access **4**, 2751–2763 (2016)
33. Zhou, Bo et al.: The carpet knows: Identifying people in a smart environment from a single step. In: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). pp. 527–532. IEEE (2017)
34. Ziegeldorf, J.H., Morchon, O.G., Wehrle, K.: Privacy in the internet of things: Threats and challenges. Security and Communication Networks **7**(12), 2728–2742 (2013)
35. Zwingelberg, H., Hansen, M.: Privacy protection goals and their implications for eID systems. IFIP Advances in Information and Communication Technology **375 AICT**, 245–260 (2012)