A User-Centric Privacy-Preserving Approach to Control Data Collection, Storage, and Disclosure in Own Smart Home Environments

Chathurangi Ishara Wickramasinghe(✉) and Delphine Reinhardt

University of Göttingen, Göttingen, Germany
c.wickramasinghe@stud.uni-goettingen.de, reinhardt@cs.uni-goettingen.de

Abstract. The smart environments around us collect a vast amount of data and disclose those data to third parties, thus potentially endangering our privacy. Research works and the European General Data Protection Regulation (GDPR) call for more user involvement in the privacy-preserving process. Existing privacy-preserving solutions do not present a solution for the entire data collection and disclosure process, while fully putting the users in the center. Therefore, in this paper, we address four main weaknesses of the existing solutions. This led us to derive a user-centric privacy-preserving approach, which allows the end users to control the entire data collection, storage, and disclosure process in smart home environments. Our approach includes: (1) applying different minimization and aggregation levels to control the data collection, (2) mechanisms helping users to assess the sensitivity level of the collected data types, (3) a model balancing privacy risks with benefits allows users to make decisions by considering their attitude towards data collection and sharing, and (4) an approach presenting privacy risks and advantages arising from sharing collected context-data allows users to make context-dependent data sharing decisions. Our paper also outlines how the proposed privacy-preserving approach can be implemented in the existing IoT system architecture in the future.

Keywords: Internet of Things · IoT · Social IoT and privacy · Usability · Data protection · Data collection · Smart objects · Smart home · Smart environments

1 Introduction

Technological progress has contributed to the fact that pervasive systems with their services have become an essential part of our everyday life. The main goal of pervasive computing systems is to enhance the quality of the end users' life without requiring extensive technical knowledge from the end users [24]. Smart environments, such as smart home, smart city, smart office, etc. are one of the

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 190–206, 2022.
https://doi.org/10.1007/978-3-030-94822-1_11

parts of the technological development in the context of pervasive computing systems. In this paper, we concentrate on smart home environments, in which the end users are interacting with various smart objects, such as smart bulbs, smart door locks, smart fridges, smart heater systems, etc. [8]. While those smart objects improve our lives in different areas, they also collect sensor-based data of their owners as well as their environment and disclose those data towards third parties [31,32]. To meet the arising privacy issues in this context, several privacy-preserving solutions have been proposed, such as [1,3,6,9,11,12,19,20]. Note that most of them do not involve the end users in their design in order to improve the acceptance of those solutions and do not allow end users to control the entire data collection and disclosure process.

Laws, such as the European law on data protection and privacy, GDPR, still call for more user involvement in the privacy protection process [21]. Additionally, the GDPR with different rights such as "Right for Access" and "Right to be Forgotten" [Art. 5, 12, 15, 17 and 19], also calls for designing privacy-preserving approaches for smart environments, which allow users to have more transparency and more control on the protection of the personal data processing [14,21,27]. Therefore, in a previous work, a questionnaire-based study was carried out, in which six **User-Centric-Control-Points** (UCCPs) were identified as requirements for user-centric privacy-preserving solutions for smart home environments [29]. The UCCPs allow end users to have more transparency and to control (a) which information is collected in which granularity, (b) what is disclosed to whom and (c) for which purpose, while considering the associated context-based privacy risks and (social or personal) advantages [29]. The six derived UCCPs from [29] are:

- **Data Object Tagging**: Allowing users to tag the smart objects as sensitive or non-sensitive according to their perception,
- **Data Minimization**: Allowing users to limit the data collection by the smart objects,
- **Data Granularity**: Allowing users to set the data collected granularity for their review,
- **Data Sharing**: Allowing users to balance the associated risks and social or personal advantages arising from sharing the collected context-data,
- **Data Disclosure Limitation**: Allowing users to control the data sharing with the options to share or to delete the collected data,
- **Data Access Limitations**: Allowing users to limit the data access and used purposes.

In this paper, we use these UCCPs in order to address the four main weaknesses of the existing privacy-preserving solutions and to derive a user-centric privacy-preserving approach for smart home environments. Addressing those four weaknesses helps end users to (1) apply different minimization and aggregation levels in order to control the data collection [21], (2) to assess the sensitivity level of the collected data types in order to make a conscious decision regarding data disclosure [14], (3) to balance privacy risks with benefits in order to make data sharing decisions by considering their own attitude towards data collection and

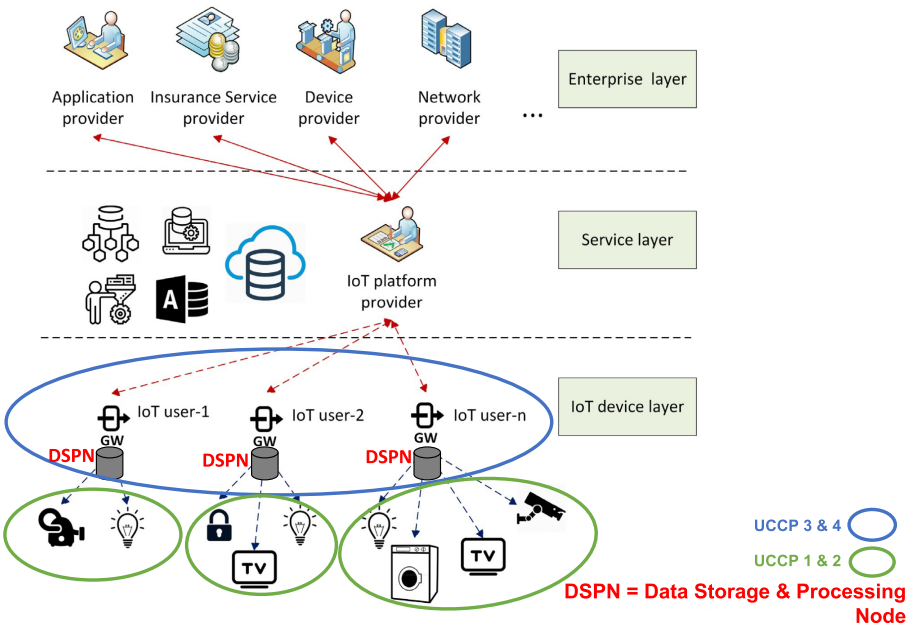

Fig. 1. Our user-centric privacy-preserving solution integrated in the IoT system architecture from [10]

sharing, and (4) to make data sharing decisions after considering the context-based privacy risks and advantages [14] arising from the different collected data types from all the smart objects in their own smart home environment at one point in time.

Our derived user-centric privacy-preserving approach can be implemented in already existing IoT system architectures. IoT system architectures include three layers: IoT device-, service- and enterprise layer [10]. While the IoT device layer consists of smart objects, gateways and Internet connection, the service layer includes the services of the IoT platform providers, such as data flow, processing, storage, and sharing tools, etc. [10]. The third layer, the enterprise layer, comprises business applications and service management technologies [10]. There are mainly two options on how to implement a user-centric privacy-preserving solution in such an IoT architecture. The options are (1) integrating a **Data Storage and Processing Node**, (*DSPN*), in the IoT device layer [2] and (2) supplying end users with a private cloud solution by an IoT platform provider [1] in the service layer. While the first option offers a storage node in their smart home environment (at the device layer), which collects and releases the collected context-data to the service layer according to users' setting in the user-centric privacy-preserving solution [2], the second option provides a data storage option in the service layer of the provider, which saves collected context-data of the smart home owners outside of their smart home environment. We recommend

to implement our proposed privacy-preserving solution including four UCCPs, as described in Sect. 2, in the IoT device layer in order to give end users the opportunity to control the entire data collection, storage, and disclosure process in their smart home environments, as illustrated in Fig. 1. Such integration would allow to meet the demanded requirements by the GDPR [21]. Note that the implementation of our approach is however out of scope of this manuscript.

To sum up, in comparison to previous works, our derived user-centric approach has the added value, that it enables the end users to use pervasive computing systems, in this case composed as intelligent objects in smart home environments, with having more transparency and control over the entire data collection and disclosure process without explicit awareness of the underlying communications and computing technologies. Additionally, our approach can be implemented in IoT system architectures in the future and will enable end users to control the entire data collection, storage, and disclosure process.

The remaining of this paper is structured as follows. We first present our user-centric privacy-preserving solution for smart home environments considering the above mentioned four weaknesses in Sect. 2 and its qualitative evaluation in Sect. 3. In Sect. 4, we discuss our proposed model. Related work and closing remarks conclude this paper in Sect. 5 and in Sect. 6, respectively.

2 Our Proposed User-Centric Privacy-Preserving Approach

2.1 Overview of the Entire Privacy-Preserving Model with UCCPs

Figure 2 presents our user-centric privacy-preserving approach including four UCCPs and their interrelationships. In the following, the functions of UCCPs are described in detail.

Our model proposes recommendations regarding data disclosure while considering the users' attitude related to data collection and sharing. While the user settings of the four UCCPs allow end users to control the entire data collection and disclosure process in their smart home environments, the integration of the proposed privacy-preserving solution in the IoT architecture allows end users to control the entire data storage until the data are released or deleted by the smart home owner. Table 1 summarizes the four user settings from UCCPs 1 to 4 with the time of their execution.

The time of execution includes three stages: Registration (R, comprises the first time users start setting up the smart object and initial utilization), Update (U, comprises the point of time when the smart object has been updated) and Aggregation Period (P, comprises the period set by the users for the review of the collected data). The details regarding each users' settings are presented in the corresponding UCCPs. In order to address the mentioned four weaknesses in Sect. 1 and to allow end users to control the entire data collection and disclosure process, our approach includes the following components. While UCCP 1 and 3 include mechanisms helping users to assess the sensitivity level of the collected

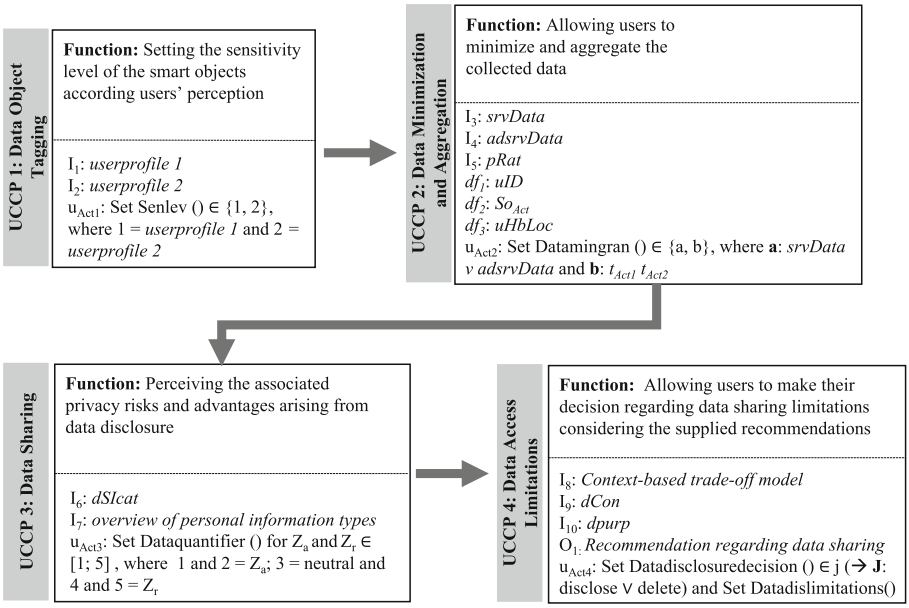

Fig. 2. Four UCCPs and their interrelationships

data, UCCP 2 includes different minimization and aggregation levels for different types of the collected data. Furthermore, UCCP 4 contains a model balancing privacy risks with personal and social benefits by considering the users' attitude regarding data collection and disclosure as well as the function helping end users to make context-based privacy-preserving data sharing decisions.

2.2 UCCPs for Controlling the Data Collection Process

UCCP 1, named as Data Object Tagging, and UCCP 2, named as Data Minimization and Aggregation, allow end users to control the data collection process. End users are asked to set the settings regarding these two UCCPs during the **registration process** of the smart objects. In order to capture users' attitude regarding data collection and privacy-preserving, end users are asked in the user setting u_{Act1}: *Set Senlev()* ϵ {1,2} of UCCP 1, to assign themselves to one of the two profiles described in this model, ($uprofile_1$, $uprofile_2$)[1]. If the users assign themselves to the $uprofile_1$, then the sensitivity level for the objects in smart

[1] **User profile one** ($uprofile_1$): Martha does not care which kind of data are collected by her smart objects and is ready to disclose all the data, including the personal data according to the definition of the Article 4 of the GDPR [21] (information directly or indirectly linkable to Martha), to different data consumers for different purposes.

User profile two ($uprofile_2$): Martha wants to know which data types are collected by her smart objects in order to supply the smart objects' services and which additional data types are collected. If these collected data include personal data according to the definition of the Article 4 of the GDPR [21] (information directly or indirectly linkable to Martha), then she likes to tag the smart objects as *sensitive* and otherwise she will tag the smart objects as *non sensitive*.

Table 1. Summary of the user actions of the approach *(X means user must set and (X) means user can set)*

User action	Registration (R)	Update (U)	Aggregation Period (P)
UCCP 1: *Set Senlev()*: Setting the sensitivity level	X	(X)	
UCCP 2: *Set Datamingran()*: Minimizing and aggregating the data	X	(X)	(X)
UCCP 3: *Set Dataquantifier()*: Weighting between privacy risks and advantages			X
UCCP 4: *Set Datadisclosuredecision ()* and *Set Datadislimitations ()*: Setting the data consumer and purposes by data sharing			X

home is set to the value 1 with the label *non sensitive* and if the users choose $uprofile_2$, then the sensitivity level for the objects in smart home is set to the value 2 with the label *sensitive*. By default or if, users feel in-between the given profiles, then the sensitivity level for the objects in smart home is also set to the value 2 with the label *sensitive*.

Complementary to UCCP 1, UCCP 2 allows end users to minimize and aggregate the data collected. With the user setting, u_{Act2}: *Set Datamingran()* ϵ {a,b} in UCCP 2, where **a**: *srvData* ν *adsrvData* and **b**: t_{Act1} ν t_{Act2}, end users are asked to set the minimization and aggregation options in order to control the data collection and data aggregation for data sharing. While *srvData* includes data collected to provide the objects' service, *adsrvData* includes additional data collected by smart objects' sensors. The setting options regarding data aggregation allow end users to choose between two options. The two options are (1) exact time of each action of the smart object for daily review (t_{Act1}) or (2) time period users want to aggregate and review the collected data by their smart objects (t_{Act2}), for example weekly, monthly, etc.[2]. In order to supply end users with more background information concerning the smart object providers the privacy ratings of the providers (*pRat*) is presented. The *pRat* is presented based on the approach in [34], which includes a 5-star-based rating system with the icon "i" next to it giving more explanation regarding the rating, for instance the data of your smart home environment are directly saved in the providers' cloud and you cannot be sure who can get access to your data and for which

[2] An example for t_{Act1} could be that the smart object owner is absent at 07:30 am on 5th of February and present again at 8 pm in the living room. He gets up at 06:30 am and switches on his smart bulbs in two rooms, namely bathroom and sleeping room. In contrast to this, an example for t_{Act2} could be that the smart object owner is available at home at various times per month and switches on his smart bulbs 200 times per month.

purpose [34][3]. In case, the user setting is missing, the default settings regarding *Set Datamingran()* are: a is assigned to $srvData$ and b is assigned to t_{Act2} including *monthly* as the aggregation period.

In order to simplify the applicability of our approach for end users, the UCCP 2 also includes three default settings regarding the variables, uID including detail information about users' personal identity, So_{Act} including setting the data aggregation layer, and $uHbLoc$ including the recording of users' availability at home and exact home-based location, for instance the user is present in the living room. The default settings are presented to the end users and if they like they are allowed to change these default settings. The default settings for uID is assigned to $status1$ including approximate and general information about the person and default settings for So_{Act} is assigned to So_{Act1}, which means that the granularity of the data is set at the layer of sensors. The default settings regarding $uHbLoc$ is assigned to $status1$, which ensures that no data regarding home-based location is collected. In case of updates, the end users get a push notification, which asks them, whether they want to adjust the settings in UCCP 2 regarding minimizing and aggregating the data collected by their smart objects. In addition, the users also can adjust the settings regarding the aggregation period (t_{Act1} ν t_{Act2}) during the review at the end of the previous aggregation period.

2.3 UCCPs for Controlling the Data Disclosure Process

UCCPs 3, named as Data Sharing, and UCCPs 4, named as Data Access Limitations, allow end users to control the data disclosure process. End users are asked to set the settings regarding these two UCCPs during the **aggregation period** of the smart objects, which users set in the UCCP 2 with t_{Act1} ν t_{Act2}. In order to capture users' sense for associated privacy risks and advantages arising from data sharing, end users are asked in the user setting *Set Dataquantifier ()* of UCCP 3 to set their personal quantifier between their own risk sensitivity (Z_r) and their own sense of advantage (Z_a). If the sense of advantages (Z_a) is higher than the risk sensitivity (Z_r), then the quantifier can be assigned to values 1 or 2, if the risk sensitivity is higher, then the quantifier can be assigned to values 4 or 5 and if the sense of advantages and risk sensitivity are equal, then the quantifier can be assigned to value 3 (neutral). In order to allow end users to perceive the associated privacy risks and advantages and to assess the sensitivity of different information types in the context of data disclosure, end users are provided with data sharing information categories ($dSIcat$) in Table 2 and an overview of personal information types in Table 3 by smart object providers. While $dSIcat$ present the associated privacy risks and (personal and social) advantages arising from disclosing the collected data, the overview of personal information types including different information in various context assigned to types of the $dSIcat$ (from Table 2) helps end users to assess the sensitivity of the information types.

[3] The approach from Zimmermann et al. [34] is a 5-star-based rating system. This system is similar to a star-based product rating we know for example on Amazon. The 5-star-ratings of each provider result from the given information by each provider and user experiences with the corresponding smart object.

Our approach includes the presented categories in *dSIcat*, because those represent the privacy risks and advantages arising from data disclosure in smart home environments and the former papers, such as [5,6,12,16,33,34], also classify the categories from *dSIcat* as relevant categories in this context. Additionally, as in the other contexts [16,22], different types of personal data are also collected in smart home environments and our approach considers the relevant personal information types summarized in Table 2 in this context. The *dSIcat* and overview of personal information types must be supplied and updated at the latest by the time, when the end users start reviewing the collected data at the end of each pre-defined aggregation period. Furthermore, end users are also allowed to add categories to *dSIcat* during the review at the end of previous aggregation period after learning over a certain period. The default setting regarding UCCP 3 includes that the quantifier is set to the value 5, which means that the risk sensitivity (Z_a) is higher than the sense of advantage.

In addition to UCCP 3, UCCP 4 allows end users to make their decision regarding data sharing and limit the data access after considering the recommendations supplied by our context-based trade-off decision model. While the user setting u_{Act4}: *Set Datadisclosuredecision ()* ϵ j (where j can be assigned to disclose ν delete) allows the end users to choose between the two options ("disclose" or "delete"), the user setting *Set Datadislimitations ()* allows end users to limit the data consumers (*dCon*)[4] and usage purposes (*dPurp*)[5] in case of data sharing. *dCon* get a rating based on the 5-star-rating approach[6], which is based on the approach from [34]. In case, the user settings are missing, the default setting for *Set Datadisclosuredecision ()* is set to *delete*. The inputs regarding *dCon* and *dPurp* must be provided and updated from smart object providers at the latest by the start of end user reviewing at the end of each pre-defined aggregation period.

In order to support end users in their data disclosure decision making process, our approach supplies recommendations based on our context-based trade-off decision model, which considers the user settings of the previous UCCPs, 1, 2 and 3. The formula of our model bases on the Markowitz's risk-return model from [17]. The Markowitz's risk-return model is rated as one of the most popular risk models in finance [17]. The formula of the Markowitz's risk-return model is:

$$U(x) = E(x) - O \times Var(x) \tag{1}$$

[4] *dCon* include third parties getting access to disclosed data, such as doctors, insurance company, government agencies, etc.

[5] *dPurp* informs end users for which purpose, such as personal health plan, statistical purposes, etc., the shared data are used by the *dCon*.

[6] Each *dCon* has to answer several questions, for instance, where the data are saved, for which purpose the data are used, with which other companies/associations the data are shared, etc. Based on the answers of the *dCon*, they get a rating based on the 5-star-rating approach.

The original Markowitz's risk-return model formula[7], presented under formula 1, was adjusted for our approach. In the following the balancing formula of our proposed approach is explained briefly. In the formula of our approach, we use the following two additional abbreviations: *So* for *smart object* and *b* for *behaviour*. The operation $E(x)$ includes the following equation in our approach[8]. It considers the default and user settings from UCCP 2 and UCCP 3 regarding uID, $uHbLoc$ and quantifier between Z_a and Z_r.

$$E(x) = Z_a \times (uID \times uHbLoc \times So_{ia}) \tag{2}$$

Else it is:

$$E(x) = (uID \times uHbLoc \times So_{ia}) \tag{3}$$

The operation $O \times Var(x)$ includes the following equation in our approach[9]. It considers the default and user settings from UCCP 2 and UCCP 3 regarding uID, $uHbLoc$ and quantifier between Z_a and Z_r.

$$O \times Var(x) = Z_r \times (uID \times uHbLoc \times So_{ip}) \tag{4}$$

Else it is:

$$O \times Var(x) = (uID \times uHbLoc \times So_{ip}) \tag{5}$$

The output of this UCCP O_1 results from the following $if-clause$: $O_1 = if\ (U(x) = E(x)$ - $O \times Var(x) > 0)$, *then "Disclose the collected data of f(b)"*, *else "Do not disclose the collected data of f(b)"*. This $if - clause$ means that the context-based trade-off model of our approach suggests end users to disclose the collected data, concluded in the formula $f(b)$, if the associated advantages are weighted higher by the end users in comparison to the arising privacy risks $(U(x) > 0)$. Otherwise, the end users are suggested not to disclose the collected data summarized in $f(b)$. As mentioned above, the formula $f(b)$ summarizes the collected data in the interaction of all existing smart objects *So* in users' smart home environment according to the users' settings regarding the UCCPs 1 to 3. If the users set *Set Datamingran(b)* is assigned to t_{Act1} in UCCP 2, then all the smart objects' actions with their exact time of execution are summarized for the users' daily review:

[7] $U(x)$: Trade-off between expected payoff with main focus on the pure profit and the variability of the payoff (risk) of an investment x;
$E(x)$: Expected payoff with main focus on the pure profit of an investment x;
O: Risk attitude of the decision maker, O assigned to 0 means risk-neutral; $O > 0$ means risk averse and $O < 0$ means risk-seeking;
$Var(x)$: Variability of the payoff (risk) of an investment x.

[8] Summary of the **associated advantages** ia_x from $dSIcat$ regarding all the *So* the user owns in his/her smart home environment: $\text{So}_{ia} = (So_{1ia1} + So_{1ia2} + So_{2ia1} + So_{2ia2} + ... + So_{nia1} + So_{nia2})$.

[9] Summary of the **associated privacy risks** ip_x from $dSIcat$ regarding all the *So* the user owns in his smart home environment. At this point, the sensitivity level $SetSenlev()$ regarding the *So*, which users set in UCCP 1, is also considered:So_ip = Set Senlev $(So_{all}) \times [(So_{1ip1} + So_{1ip2} + ...) + (So_{2ip1} + So_{2ip2} + ...) + ... + (So_{nip1} + So_{nip2} + ...)]$

Table 2. I_6: Overview of the data sharing information categories ($dSIcat$) [5,6,12,16,33,34]

Abbreviation of category	Category name	Category description
ip_1: Associated privacy risks 1	Discrimination and Manipulation	Using to create special contract and discriminate the users, for example, manipulating the device owners with contracts
ip_2: Associated privacy risks 2	Burglaries and Misuse	Using the data to harm the users, for instance, breaking in after analysing data about home availability and smart door lock
ip_3: Associated privacy risks 3	Profiling	Using the data to track the users and manipulate the users and steal the users' identity
ip_4: Associated privacy risks 4	Carrier risks	Using the data to find out characteristics of the users, for example, analysing and disclosing such information can result in risks for future employers
ip_5: Associated privacy risks 5	Damaging	Using the data to damage the device owners, for instance identity theft based on the disclosed sensor data
ip_6: Associated privacy risks 6	Personal Exposure	Using the data to publish things users are doing, for example, data disclosure could result in being exposed because they had done something they did not want their friends and family to know about, maybe also to carry out Propaganda, etc.
i_{a1}: Associated social advantage 1	Personal Advantages	Using the data to provide the user specific contracts and users can earn money
i_{a2}: Associated social advantage 2	Social Advantages	Using for statistical aims, for example, using data for research works, market analysis

Table 3. I_7 - Extract from the overview of the personal information types and their associated $dSIcat$ [16,22]

Information type	Directly	Linkable	dSIcat					
			ip_1	ip_2	ip_3	ip_4	ip_5	ip_6
Body size	–	x	x	x	–	–	–	–
Voice print	–	x	x	–	–	–	–	–
Body and facial images	x	–	x	x	–	–	–	–
Biological characteristics/Biometrics	x	–	x	x	–	x	–	–
Recording of using health equipment	–	x	x	x	x	–	x	x

$$f(b) = \sum_{b_{t1\ Act1}}^{b_{tn\ Act1}} = [b_{t1\ Act1} + b_{t2\ Act1} + ... + b_{tn\ Act1}] \tag{6}$$

$$b_{t1\ Act1} = t_{1\ Act1} \times (So_{1\ Act1} + So_{2\ Act1} + ... + So_{n\ Act1}) \tag{7}$$

$$b_{tn\ Act1} = t_{n\ Act1} \times (So_{1\ Act1} + So_{2\ Act1} + ... + So_{n\ Act1}) \tag{8}$$

If the users assigned *Set Datamingran(b)* to t_{Act2} with weekly in UCCP 2, then all the smart objects' actions with rough information are aggregated for the users' review period, in this case weekly base:

$$f(b) = b_{Act2}; b_{Act2} = t_{Act2} \times (So_{1\ Act2} + So_{2\ Act2} + ... + So_{n\ Act2}) \tag{9}$$

To sum up, our approach also allows the end users to adopt the settings of the previous review period regarding the four UCCPs in the upcoming aggregation periods.

3 Evaluation

The proposed user-centric privacy-preserving model addresses weaknesses of the already existing privacy-preserving solutions in IoT, as mentioned in Sect. 1. In this section, we evaluate our proposed model qualitatively. In order to outline the added value of our model, we evaluate our approach with already existing approaches, [1,3,6,12]. The existing approaches from [1,3,6,12] are relevant works in this area and provide the basis for our approach. While [6] presents a reference architecture including a trade-off decision component, [12] proposes a Role Based Access Control framework, which can be applied in the context of data sharing in smart home environments. However, the detailed and qualitative evaluation of these approaches shows that both approaches do not allow the users to control the entire data collection, storage, and disclosure process. Furthermore, the papers [1,3] present a security and privacy-preserving solution as well as privacy negotiation mechanisms for IoT environments. Although these approaches provide the basis for our proposed approach, they do not give sufficient options for the user to control the privacy-preserving data collection, storage, and disclosure.

The evaluation metrics are organized in two categories: (1) Privacy-preserving functionalities of the proposed solutions and (2) the rights of the smart object owners. The qualitative evaluation metrics for each category are derived from [15,18,21,29]. Table 4 outlines the results of this qualitative evaluation. Each evaluation category contains five metrics, which are evaluated by using the following rating scale: ○ = no possibility; ◑ = partially possible; and ● = possible.

4 Discussion and Limitations

4.1 Discussion

With our proposed privacy-preserving approach, we (1) address four weaknesses of the previous approaches, as mentioned in Sect. 1, (2) empower end users

Table 4. Results of the qualitative evaluation

Evaluation category	Metrics	Our model	Model 2 [6]	Model 3 [12]	Model 4 [1]	Model 5 [3]
Privacy-preserving functionalities	Data minimization [18]	●	○	○	○	○
	Limited data processing for specific purposes (smart object service) [18]	●	○	○	◑	○
	Data aggregation [18]	●	○	○	◑	○
	Privacy-preserving data storage [18,21]	●	◑	○	◑	●
	Transparent processing of collected personal data	●	●	○	○	◑
Rights of the smart object owners	End users to limit the data collection [21,29]	●	○	○	○	○
	End users to limit the data access [21,29] by limiting the data consumers and usage purposes [29]	●	◑	◑	◑	●
	End users to assess the sensitivity level of the collected data types [29]	●	◑	○	○	◑
	End users to have transparency about data processing [15,21], for instance, arising privacy risks by disclosing	●	◑	○	○	○
	End users to evaluate the context-data before data disclosing [29]	●	○	○	○	◑

to control the entire data collection, storage, and disclosure process, and (3) help end users to make context-based privacy-preserving data sharing decisions. By proposing a user-centric privacy-preserving approach, which can be implemented in the future in existing IoT architectures, we address the demands from the GDPR [21], especially rights such as "Right for Access" and "Right to be Forgotten" (Art. 5, 12, 15, 17 and 19). Addressing the mentioned weaknesses from [6,14] and therefore including minimization and aggregation levels in our proposed approach for smart home environment allows end users to control the entire data collection process. Additionally, the related work with already proposed privacy-preserving solutions in this context, mentioned in Sect. 5, outlines that those solutions do not include mechanisms enabling end users to assess the sensitivity level of the collected data. Those mechanisms help to increase the transparency in this context, which is also required by the GDPR [21]. There are few privacy-preserving approaches including context-based permission systems [13], security frameworks for smart objects [23,25], and privacy risk

trade-off/negotiation models [1,3,5,6,12], but they still include weaknesses, which we address in our approach as outlined in Table 4, for instance presenting the associated privacy risks and advantages arising from sharing the collected data types from all the smart objects in smart home environments at one point in time and giving end users the opportunity to make the data disclosure decision consciously after assessing the sensitivity level of the context-based collected data.

As mentioned above, the included context-based trade-off decision model in our proposed approach provides end users with recommendations regarding the data disclosure based on their attitude regarding data collection and disclosure. However, it must be investigated in the future, whether end users are willing to have such trade-off decision-making model based recommendations. Moreover, they are asked to make settings considering their attitude towards data collection and disclosure. Therefore, they get several information as inputs in our approach, such as $pRat$, $dSIcat$, $dCon$, $dPurp$, etc. These inputs must be tested with end users within a user study to find out whether they want to have the predefined inputs and/or whether they are willing to have more information than the existing inputs of the proposed approach in order to consider those information in their data-disclosure decision-making process. It should be emphasised, that it is essential to give end users the opportunity to assess the sensitivity level of their collected data in order to avoid inappropriate decisions regarding data disclosure. The derived data sharing information categories ($dSIcat$) in Table 2, the overview of the personal information types in Table 3, and their classification to $dSIcat$ in our model may help end users to assess the sensitivity level of the collected data. These input information are based on a literature review and must be quantified with the help of an online questionnaire, which we will address in our future research work. In this context, it would be interesting to find out whether such overviews with personal information types assigned to categories, such as $dSIcat$, help the end users to assess the sensitivity level of their own personal data.

4.2 Limitations

Finally, our work has few limitations. The findings are mainly based on a literature review and on the results of a previously carried out questionnaire-based survey [29]. The derived model in Sect. 2 must be validated with a user questionnaire and studies, which we target to address in the near future.

5 Related Work

In previous works, different approaches have been proposed regarding privacy-preserving solutions for IoT, such as [5,6,12]. While [12] presents a Role Based Access Control (RBAC) framework including k-anonymity mechanisms, in [5,6] Barhamgi et al. present a reference data sharing architecture for privacy engineering in environments with smart health care devices. The proposed architecture includes trade-off data sharing decision components, which allow the smart

object owners to make pragmatic data sharing decision balancing privacy risks and potential benefits. Both solutions aim at giving the opportunity to users to manage the data disclosure of their smart health care device. Moreover, [1] and [3] propose further privacy-preserving solutions. While [1] presents a security and privacy-preserving IoT architecture for smart home environments, [3] proposes privacy negotiation mechanism for IoT environments. Both solutions address the privacy-preserving issues in IoT environments, specially in the context of smart homes, for instance, ensuring privacy-preserving data storage, allowing end users to limit the data access, etc. However, these approaches [1,3,5,6,12] include few limitations, such as (1) the end users cannot control the entire data collection, storage, and disclosure process in their smart home environment, (2) users cannot make context-based data-disclosure decisions considering all the data collected by the smart objects in a smart home environment, (3) users have no possibilities to control the data collection process, and (4) users do not have the possibility to apply data minimization and aggregation strategies as well as data usage limitations. Furthermore, in [34] a smart home configurator is introduced, which supports the end users' decision-making process regarding smart home technologies when buying the smart objects. This configurator depicts the smart home data processes and informs end users about the implications regarding privacy and security in order to increase transparency and reduce the lack of clarity in the decision-making process during the purchase process. This approach [34] does not allow the end users to control the entire data collection, storage, and disclosure process while using those smart objects in their own smart home environment. Moreover, few user studies are carried out, such as [31,33], in order to analyse users' mental and threat models of privacy consequences and obstacles for the privacy protection. These user studies deliver valuable hints regarding user-centric privacy-preserving solutions, such as giving users the transparency about privacy consequences in an understandable way and limiting data recipients, etc., but do not present a general user-centric privacy-preserving approach regarding the disclosure of context-based data in smart home environments with different smart objects. Moreover, some prior works, such as [4,7,13,26,28,30], present further privacy-preserving standalone solutions including context-based permission systems, privacy-preserving policies, authentication protocols and data encryption methods in order to protect the collected sensor data by smart objects. Additionally, there are few works, such as [16,22], which propose information sensitivity typologies. In their survey results, they summarize different information types from various contexts to derived sensitivity levels in an understandable way [16,22].

In comparison to all the above-mentioned previous work, our approach concentrates on user-centric control points integrated privacy-preserving approach closing four main weaknesses of existing approaches, such as data minimization and aggregation levels, evaluation mechanisms for data sensitivity level, context-based trade-off decision model supplying data sharing recommendations while considering end users' settings regarding their attitude towards data collection and disclosure as well as balancing privacy risks and personal and social advantages.

6 Conclusions and Future Work

Within the scope of this paper, we have derived a user-centric privacy-preserving approach considering user-centric-control-points, called UCCPs. To sum up, it allows the end users to control the entire data collection, storage, and disclosure process by considering different minimization and aggregation levels, applying mechanisms helping users to assess the sensitivity level of the collected data types, by balancing the arising privacy risks and (social and personal) advantages and applying functions helping end users to make context-based privacy-preserving data sharing decisions. With this approach we address the existing weaknesses of the previous proposed solutions and empower end users to have more control over the processing of their personal data in the smart home environment.

In future work, we plan (1) to implement the proposed approach in a smart home environment and to carry out (2) user studies to investigate its performance as well as users' acceptance and derive further requirements regarding the proposed approach in this paper. Furthermore, we also plan (3) to conduct questionnaire-based studies in order to find out whether the overview of personal information types assigned to *dSIcat* categories helps end users to assess the sensitivity level of their collected personal data.

Acknowledgments. We thank the anonymous reviewers for their feedback and Alexander Richter for his support. Furthermore, we would like to thank Daniel Franke and Birgit Schuhbauer for proofreading.

References

1. Abu-Tair, M., et al.: Towards secure and privacy-preserving IoT enabled smart home: architecture and experimental study. Sensors **20**(21), 1–14 (2020)
2. Aïvodji, U.M., Gambs, S., Martin, A.: IOTFLA: a secured and privacy-preserving smart home architecture implementing federated learning: a secured and privacy-preserving smart home architecture implementing federated learning. In: Proceedings of 2019 IEEE Security and Privacy Workshops (SPW), pp. 175–180 (2019)
3. Alanezi, K., Mishra, S.: Incorporating individual and group privacy preferences in the Internet of Things. J. Ambient Intell. Humanized Comput. 1–16 (2021). https://doi.org/10.1007/s12652-021-02959-7
4. Alcaide, A., Palomar, E., Montero-Castillo, J., Ribagorda, A.: Anonymous authentication for privacy-preserving IoT target-driven applications. Comput. Secur. **37**, 111–123 (2013)
5. Barhamgi, M., et al.: POSTER: Enabling end-users to protect their privacy, pp. 905–907 (2017)
6. Barhamgi, M., et al.: Enabling end-users to protect their privacy. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, pp. 905–907. ACM (2017)
7. Cao, J., Carminati, B., Ferrari, E., Tan, K.L.: Castle: continuously anonymizing data streams. IEEE Trans. Dependable Secure Comput. **8**(3), 337–352 (2010)
8. Carretero, J., García, J.D.: The Internet of Things: connecting the world. Pers. Ubiquit. Comput. **18**(2), 445–447 (2014)

9. Chakravorty, A., Wlodarczyk, T., Rong, C.: Privacy preserving data analytics for smart homes. In: Proceedings of 2013 IEEE Security and Privacy Workshops, pp. 23–27 (2013)
10. Firoozjaei, M.D., Lu, R., Ghorbani, A.A.: An evaluation framework for privacy-preserving solutions applicable for blockchain-based Internet-of-Things platforms. Secur. Priv. **3**(6), 1–28 (2020)
11. Huang, X., Craig, P., Lin, H., Yan, Z.: SecIoT: a security framework for the Internet of Things. Secur. Commun. Netw. **9**(16), 3083–3094 (2016)
12. Huang, X., Fu, R., Chen, B., Zhang, T., Roscoe, A.: User interactive Internet of Things privacy preserved access control. In: International Conference for Internet Technology and Secured Transactions, pp. 597–602 (2012)
13. Jia, Y.J., et al.: ContexIoT: towards providing contextual integrity to applied IoT platforms. In: Network and Distributed System Security Symposium (NDSS), pp. 1–15 (2017)
14. Kounoudes, A.D., Kapitsaki, G.M.: A mapping of IoT user-centric privacy preserving approaches to the GDPR. Internet Things **11**, 100179 (2020)
15. Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., Zhao, W.: A survey on Internet of Things: architecture, enabling technologies, security and privacy, and applications. IEEE Internet Things J. **4**(5), 1125–1142 (2017)
16. Milne, G., Pettinico, G., Hajjat, F., Markos, E.: Information sensitivity typology: mapping the degree and type of risk consumers perceive in personal data sharing. J. Consum. Aff. **51**(1), 133–161 (2016)
17. Nagengast, A.J., Braun, D.A., Wolpert, D.M.: Risk-sensitivity and the mean-variance trade-off: decision making in sensorimotor control. Proc. Roy. Soc. B Biol. Sci. **278**, 2325–2332 (2011)
18. Oetzel, M.C., Spiekermann, S.: A systematic methodology for privacy impact assessments: a design science approach. Eur. J. Inf. Syst. **23**(2), 126–150 (2014)
19. Ouaddah, A., Abou Elkalam, A., Ait Ouahman, A.: FairAccess: a new blockchain-based access control framework for the Internet of Things. Secur. Commun. Netw. **9**(18), 5943–5964 (2016)
20. Perera, C., McCormick, C., Bandara, A.K., Price, B.A., Nuseibeh, B.: Privacy-by-design framework for assessing Internet of Things applications and platforms. In: Proceedings of the 6th International Conference on the Internet of Things, pp. 83–92. ACM (2016)
21. 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). Official Journal of the European Union L119/1, pp. 1–88 (2016)
22. Rumbold, J., Pierscionek, B.: What are data? A categorization of the data sensitivity spectrum. Big Data Res. **12**, 49–59 (2018)
23. Sachidananda, V., Siboni, S., Shabtai, A., Toh, J., Bhairav, S., Elovici, Y.: Let the cat out of the bag: a holistic approach towards security analysis of the Internet of Things. In: Proceedings of the 3rd ACM International Workshop on IoT Privacy, Trust, and Security (IoTPTS), pp. 3–10 (2017)
24. Satyanarayanan, M.: Pervasive computing: vision and challenges. IEEE Pers. Commun. **8**(4), 10–17 (2001)
25. Siboni, S., Shabtai, A., Tippenhauer, N.O., Lee, J., Elovici, Y.: Advanced security testbed framework for wearable IoT devices. ACM Trans. Internet Technol. **16**(4), 1–25 (2016)

26. Su, J., Cao, D., Zhao, B., Wang, X., You, I.: ePASS: an expressive attribute-based signature scheme with privacy and an unforgeability guarantee for the Internet of Things. Futur. Gener. Comput. Syst. **33**, 11–18 (2014)
27. Tabassum, M., Kosinski, T., Lipford, H.R.: 'I don't own the data': end user perceptions of smart home device data practices and risks. In: Proceedings of SOUPS 2015, Symposium on Usable Privacy and Security, pp. 435–450 (2019)
28. Wang, X., Zhang, J., Schooler, E.M., Ion, M.: Performance evaluation of attribute-based encryption: toward data privacy in the IoT. In: 2014 IEEE International Conference on Communications (ICC), pp. 725–730 (2014)
29. Wickramasinghe, C.I., Reinhardt, D.: A survey-based exploration of users' awareness and their willingness to protect their data with smart objects. In: Friedewald, M., Önen, M., Lievens, E., Krenn, S., Fricker, S. (eds.) Privacy and Identity 2019. IAICT, vol. 576, pp. 427–446. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-42504-3_27
30. Yang, J.C., Fang, B.X.: Security model and key technologies for the Internet of Things. J. Chin. Univ. Posts Telecommun. **18**, 109–112 (2011)
31. Zeng, E., Mare, S., Roesner, F.: End user security and privacy concerns with smart homes. In: Proceedings of SOUPS 2013, Symposium on Usable Privacy and Security, pp. 65–80 (2017)
32. Zhou, W., Jia, Y., Peng, A., Zhang, Y., Liu, P.: The effect of IoT new features on security and privacy: new threats, existing solutions, and challenges yet to be solved. IEEE Internet Things J. **6**(2), 1606–1616 (2019)
33. Zimmermann, V., Bennighof, M., Edel, M., Hoffmann, O., Jung, J., Wick, M.: 'Home, smart home' - exploring end users' mental models of smart homes. In: Mensch und Computer 2018-Workshop Band, pp. 401–417 (2018)
34. Zimmermann, V., Dickhaut, E., Gerber, P., Vogt, J.: Vision: shining light on smart homes - supporting informed decision-making of end users. In: Proceedings of 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS PW), pp. 149–153 (2019)