A Survey-Based Exploration of Users' Awareness and Their Willingness to Protect Their Data with Smart Objects

Chathurangi Ishara Wickramasinghe, Delphine Reinhardt

A Survey-based Exploration of Users' Awareness and their Willingness to Protect their Data with Smart Objects

Chathurangi Ishara Wickramasinghe and Delphine Reinhardt

Georg-August-Universität Göttingen, 37073 Göttingen, Germany
{c.wickramasinghe@stud.uni-goettingen.de,reinhardt}@cs.uni-goettingen.de

Abstract. In the last years, the Internet of Things (IoT) and smart objects have become more and more popular in our everyday lives. While IoT contributes in making our everyday life more comfortable and easier, it also increases the threats to our privacy, as embedded sensors collect data about us and our environment. To foster the acceptance of IoT, privacy-preserving solutions are therefore necessary. While such solutions have already been proposed, most of them do not involve the users in their design. In this paper, we therefore adopt a user-centric approach and lay the ground for the future design of user-centric privacy-preserving solutions dedicated to smart home environments. To this end, we have designed and distributed a questionnaire fulfilled by 229 anonymous participants. Our objectives are two-fold: We aim at investigating (1) requirements for end user-involved privacy-preserving solutions and (2) users' readiness to be involved in their own privacy protection. Our results show that the majority of our participants are aware of the data collection happening as well as the associated privacy risks and would be ready to control and audit the collected data.

Keywords: Internet of Things · IoT · Social IoT · Privacy · Data protection · Data collection · Smart objects · Smart home · Smart environments

1 Introduction

In the last decade, the interest in IoT has tremendously increased, resulting in different products now available for and usable by the general public [7]. IoT is based on a network, where the physical objects of our environment, such as homes and workplaces, gain the ability to provide services and simultaneously play an active role in our environment via embedded systems [7]. The IoT is composed of different smart objects, which adapts to both, users' behavior and the environment. For example, smart objects include smart lamps, smart fridges, smart door locks, and smart parking systems [7]. This rapid technological development is foreseen to continue in the coming years, reaching billions of smart objects. These objects further contribute in improving our lives in different areas, including our homes and workplaces [7]. Note that smart objects do not only present advantages in households, such as helping in managing our energy

consumption, but also in companies, which can benefit from automated context-aware processes [7, 12, 16].

To provide these services, smart objects with embedded sensors continuously collect a vast amount of data about their environments and potential users, thus potentially endangering the privacy of their owners as well as of potential bystanders [36, 38]. Privacy issues especially arise when sensitive personal data are collected and disclosed to third parties without the users' consent by smart object providers [18, 27, 38]. Cyber attacks caused by security vulnerabilities [28, 38], which among others are enabled by the use of low-power hardware in smart objects, can also result in information leaks and endanger users' privacy [28, 38]. Recently, Bloomberg reported that thousands of the Amazon workers listen, how the users interact with Alexa, the virtual assistant in Amazon Echo devices [11]. Despite the phenomenon of the privacy paradox[1] [10, 20], laws [1] still call for more user involvement in their own privacy protection process, because (1) users have the fundamental right of protecting their personal data [1, 27] and (2) users' privacy behavior highly depends on the context [10].

Furthermore, the European General Data Protection Regulation (GDPR) with different rights, such as "Right for Access" and "Right to be Forgotten", calls for giving the users more transparency regarding the personal data processing, empowering users to be responsible and to have more control for the protection of the personal data processing [1]. Therefore, it is important to put the user in the center while designing usable privacy-preserving solutions for smart home environments.

Within the scope of this paper, we primarily focus on (1) the exploration of user's willingness to control the disclosure of their data and their need for transparency regarding the data collection. Based upon the results, we further focus on (2) identifying requirements in the form of user centric control mechanisms for privacy-preserving solutions for smart home environments.

The remaining paper is structured as follows. We first discuss related work in Sec. 2. We next detail the methodology of our empirical study in Sec. 3. In Sec. 4, we present the demographics and the results of our survey. In Sec. 5, we formulate design requirements based on the survey results for end user-centric-privacy-preserving solutions. Discussions and closing remarks conclude this paper in Sec. 6 and Sec. 7, respectively.

2 Related Work

Existing works can be classified as follows: (1) user surveys regarding privacy issues in IoT and (2) technical approaches allowing users to apply control mechanisms for their privacy protection.

The first category includes surveys, which are carried out with smart objects' consumers in order to find out users' perception and opinions regarding privacy issues in IoT. Based on interviews with eleven smart home owners, Zheng et

[1] Privacy paradox explains the discrepancy between the users' stated preferences with regard to privacy protection and their actual behavior.

al. outline that the users' primary motivation of using smart objects lies on the convenience and connectedness [37]. They recommend developers to focus on designing (mobile) applications, allowing the users to access and control the collected data [37]. In [36], Zeng et al. also encourage developers to design smart objects considering users' privacy needs. Additionally, the user study by Martin and Nissenbaum [23] outlines that users find that the usage of their data is more relevant to users' privacy opinion than the sensitivity level of the collected data. Moreover, few large-scale surveys [3, 21, 25] were also carried out in order to find out users' privacy preferences while using smart objects. The results of these studies confirm that privacy issues regarding IoT objects highly depend on the context [3, 25]. Some user studies also focus on privacy issues regarding smart watches and toys connected to the Internet [24, 30]. These studies investigate users' awareness of privacy issues while using such smart objects. They give hints for the designers and smart object providers how to deal with users' needs regarding such smart objects in order to increase the acceptance of smart objects. In comparison to the previous works, our questionnaire-based approach focuses on identifying control mechanisms that users want to have in the data collection and disclosure process of smart home environments. These control mechanisms should empower users to protect their own privacy in their smart home environment. Additionally, our study helps to understand, whether the users want to have the empowerment to control their personal data protection while living in smart home environments.

In the second category, we consider technical solutions that allow users to apply control mechanisms for their privacy protection. Solutions such as [16, 17, 28, 35] aim at avoiding the misuse of IoT objects and collected data by attackers for burglaries. While [16] implements a strong password authentication policy in their smart home automation system, the approach in [35] includes a set of new security policies for detecting abnormal behavior of each device. In addition, the solution presented in [17] introduces a new context-based permission system, which allows the user to decide based on collected context information, whether an abnormal action will be performed. Perera et al. propose in [28] a Privacy-by-Design framework, allowing the evaluation of IoT applications and middleware platforms based on a set of guidelines. These guidelines can be categorized in four elements: (a) Minimizing data collection, storage and disclosure without users' consent; (b) reducing the data granularity and controlling data; (c) anonymizing data and encrypting data communication and processing; (d) publishing source code, data flow diagrams of IoT applications, certifications and fulfilled compliance. Few technical frameworks, such as [4, 8, 14, 15, 26], present Role Based Access Control (RBAC) including k-anonymity mechanisms and privacy preserved access control protocols for IoT environments. These frameworks include authentication protocols to identify the user and to allow users the event-based data sharing for user-defined roles, such as doctor, partner, etc. The functionality of the frameworks is mostly explained with the help of the collected sensor data based on smart healthcare systems and other devices, such as wearables as well as few home and hotel automation devices [4, 8, 14, 15, 26]. Further ap-

proaches introduce a privacy preserving policy, authentication protocols and data encryption methods in order to protect the collected sensor data and thus users' privacy [2, 5, 6, 9, 13, 22, 29, 31–34]. However, most of these solutions reduce the availability of original data with time delay [36]. Finally, in [19], Khan et. al. present a solution to improve the privacy concerns in case of ownership change of the smart objects. These considerations show us that the proposed technical solutions include less user involvement. In comparison to previous works, our survey thus focuses on deriving control mechanisms from the end user perspectives. The proposed technical solutions in this category can be considered in the technical implementation of the derived requirements of this paper.

To the best of our knowledge, the contribution of our research work to this body of literature is two-fold: (1) We show users' readiness to be involved in their own privacy protection, (2) we derive requirements for end user-centric-privacy-preserving solutions. This lays the ground for our future work.

3 Methodology

In order to gather insights regarding our main objectives, we carried out an online questionnaire based survey[2]. Our questionnaire including 22 questions is in English and available in Appendix A.

It is structured as follows. It gathers insights in participants' knowledge and experience with smart objects. Next, it addresses the potential participants' awareness of data collected and disclosed by smart objects and their related privacy risks. It then focuses on the participants' potential willingness to inform themselves and control the data collected and shared by smart objects, before analyzing their requirements and motivation to use privacy-preserving solutions.

We distributed our questionnaire on online social network platforms, such as Xing, LinkedIn, SurveyCircle, IoT Subreddit and the community platforms of several companies in order to reach frequent Internet users. No incentives were given to the participants. It required approximately ten to fifteen minutes to be answered and consisted of multiple choice and open-ended questions. Main goal of the survey is to conduct a preliminary study as a basis for future studies rather than collecting representative insights, which are valid for the whole population.

In total, 229 participants completed the questionnaire. We have discarded invalid data sets and this resulted in 209 valid data sets. Moreover, during our analysis we derived and tested five hypotheses based on Q_{16}, applied statistical tests, such as Mann-Whitney, multiple linear regression and correlation tests and carried

[2] At the beginning of the survey, we informed the participants that both data collection and processing take place anonymously. Note that the survey was carried out at the University of Bonn, which did not have an ethical board for reviewing user studies in our field at the time of the study. We have, however, limited the data collection to the minimum and conducted it anonymously. The participants were informed that they could opt out at any time and that their data would be removed. After agreeing to participate, each participant has been assigned a pseudonym and asked to answer a questionnaire to gather his/her demographics.

out comparisons of different participant groups in order to get more insight regarding user-centric control mechanisms for privacy-preserving solutions.

4 Results

4.1 Demographics

Our respondents are predominantly male (69%). Most of them are between 26 and 50 years old (58%). 16% are under 26 and 25% over 51. The majority are German citizens (74%) followed by US Americans (7%), Sri Lankans (5%), and British citizens (5%). The remaining citizenships are distributed among 15 other nationalities from all over the world. Among the 209 participants, 166 indicated their annual income range, which ranges between "less than 25.000 Euro" and "more than 100.000 Euro". However, most of these participants (34%) annually earn "between 40.000 and 75.000 Euro".

4.2 Knowledge and Experience

In our sample, about 93% of our participants indicated that they have already heard about IoT (Q_1, n_{Q1}=209). In order to get more insight, we asked our participants, in which context they have heard about IoT (Q_2). In Q_2, we also specified what we meant by IoT, by giving some examples for orientation, such as smart home, smart factory, smart city, etc. The mentioned answers were smart home (ca. 27%), Industry 4.0 (ca. 20%), smart/intelligent things (ca. 19%), smart city (ca. 19%), and smart factory (ca. 13%) (Q_2, n_{Q2}=209).
In the free text box further answers were given such as smart vehicles, smart clothes, wearables, smart meters, smart grids, smart supply chain, smart campus, smart agriculture, robotic machines, smart logistics, smart health devices, and predictive maintenance. Additionally, 89% of the participants mentioned that they know or use smart objects (Q_3, n_{Q3}=209). The most cited answers were: Controlling home technology apps (12%), smart voice control objects, like Amazon Echo (10%), smart health devices (8%), smart door/window locks (8%), smart bulbs (7.5%), smart fridge (7.2%), augmented /virtual reality glasses (7%), smart washing machine (6%), smart alarm clock (5.7%), smart toothbrush (4%), smart grid apps (3%) and smart scale (2.7%). The majority uses the specific smartphone apps for this purpose (71%), while 11% uses the associated web interface (Q_4). Regarding Q_5 with "How frequently do you use a device connected to the Internet, such as smart scale, fridge, wearables, watch, etc.? (smartphone, computer, smart TV does not count as smart devices in this question)", about 70% of the participants indicated that they use devices connected to the Internet frequently ($n_{Q_5} = 206$). Among the participants frequently using smart objects, 76% use them at least once per day, while the remaining 24% only use them occasionally. The cross tables grouped by gender and age groups show that male participants and participants in general aged between 26 and 50 years significantly use connected devices more frequently than others. Based on the answers to Q_5, we derived three user categories that we use in our further analysis.

1. Frequent users: They use connected devices several times a day,

2. Average users: They use connected devices at most once a day or less,
3. Non users: They do not use any connected devices.

One of the questions we asked the participants using a 5-point Likert scale[3] was to indicate their degree of agreement regarding the statement: "In a few years, I believe that it will be difficult to live without using smart objects" (Q_{11}, $n_{Q_{11}} = 208$). About 87% of the participants agreed that it will be difficult to live without smart objects. A majority appreciated the potential advantages offered by smart objects (Q_{12}), while only 20% of the sample stated that there are no advantages offered by smart objects. The seven most frequently mentioned advantages can be summarized as follows: Facilitating to fulfill routine tasks, high comfort and convenience, low error rates, setting adjustments according to lifestyle, recording interesting personal information, outline the optimization potentials and specific things are automatically done.

4.3 Collection, Disclosure, and Privacy

A large majority (93%) of our participants believe that smart objects collect information about themselves and their environments (Q_6, $n_{Q_6} = 200$). However, only 58% agreed with the statement:"I believe that I know the information collected by smart objects."(Q_7, $n_{Q_7} = 193$). Most cited answers were location (29%) and health (25%) followed by browsing (24%) and personal data (19%), like bank details etc. (Q_8). With regard to the derived user profiles in Sec. 4.2, frequent and average users appear to be more aware of the data collection than the non users (Mann-Whitney test frequent users vs. non users: $p - value = 0.003 < 0.05$, r = 0.248, average users vs. non users: $p - value = 0.048 < 0.05$, r = 0.195). The boxplots in Fig. 1 confirm the above mentioned results. The outliers present the divergent answers from the frequently mentioned answers by most users and each number presents the data set of the correspondent anonymous respondent.
Additionally, only 24% of the sample believe in knowing the third parties, who receive the data collected by smart objects (Q_9, $n_{Q_9} = 209$: "I believe that I know the parties who have access to collected data and receive the collected data from my smart objects (Parties can be: hospital, doctors, insurance companies, institutes using data for statistics, etc.)"). As expected, the majority (72%) indicated that they do not know the third parties who get access to their collected data. Note that few participants mentioned some of the third parties. The mentioned parties can be clustered as follows: retail companies (like Amazon, Apple, etc.), service providers (like Google), cyber security firms, social media companies (such as Facebook, etc.), several smart object/telco providers, institutes/companies using data for statistics and analyses, (health) insurance companies, hospitals, doctors, manufacturers of the heating systems/cars, banks and government departments.
We further asked the participants to mention potential privacy issues and privacy risks in the context of IoT in a free text box (Q_{14}, n_{Q14}=209). About 55%

[3] A score of 1 corresponds to a strong disagreement, while a score of 5 to a strong agreement.

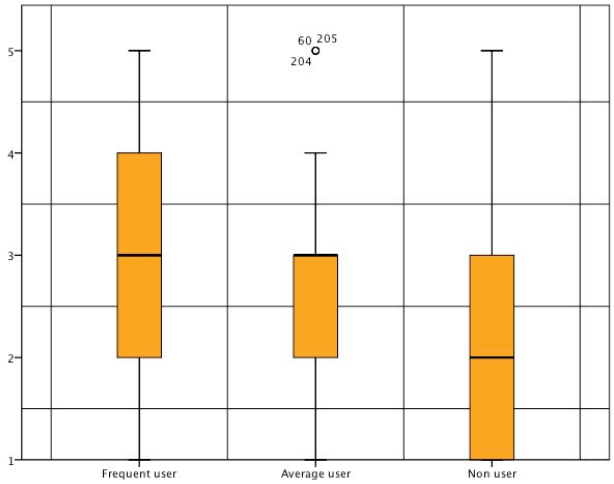

Fig. 1. Boxplots regarding "I believe that I know the information collected by smart objects."(Q_7) clustered by user profiles, derived based on the results of Q_5 (1 = strong disagreement, while 5 = strong agreement)

of the participants filled it. To sum up their statements, they mentioned that the smart objects on the one hand make their lives and every day routines easier, but on the other hand that those objects collect a vast amount of data and transfer those data to third parties, which are used for personalized services or offers, to create (more or less) detailed personal profiles of the users and to manipulate the smart object owners. Additionally, the participants also indicated that today they actually do not have any means to protect those data, before sharing it with third parties and that there is a lack of strict regulations regarding privacy aspects in smart environments. Moreover, approx. 93% of the participants agreed to the statement: "I believe that smart objects can endanger my privacy." (Q_{15}, $n_{Q_{15}} = 205$). The results of the Fisher's Exact test outline that there is no dependency between the participants' gender and their answers regarding Q_{15}. While 38% indicated that they take special measures to protect their privacy when using smart objects, 48% denied to do so (Q_{10}, $n_{Q_{10}} = 209$: "I take special measures (such as switching off some services etc.) to protect my privacy when using smart objects."). The mentioned measures are switching off the objects to avoid the data collection (35%), disuse of cloud connection, using local servers (6%), and checking all the privacy settings and disabling smart objects or features, which are not necessary (57%). One participant mentioned that s/he actually does not know any measures that really help to protect privacy (2%).

4.4 Information and Control Willingness

In the next step, respondents had to rate the following statements on a 5 point Likert scale. Based on the results regarding the statements, we investigate to which degree participants are willing to exercise control over the data collected

and shared by smart objects (Q_{16}). The statements and the distribution of the values regarding those statements are presented in Tab. 2 and in Fig. 2, respectively.

Q#	Statements
$Q_{16.1}$	I would like to have more information about the data collected by smart objects about me in a smart home environment. ($n_{Q16.1}$=206)
$Q_{16.2}$	I would like to have an overview of all the information collected by my smart objects. ($n_{Q16.2}$=206)
$Q_{16.3}$	I would like to have a summary of the collected data over a given period, e.g. daily, weekly, monthly, etc. ($n_{Q16.3}$=206)
$Q_{16.4}$	I would like to know in real-time about the data collected in my smart home environment. ($n_{Q16.4}$=205)
$Q_{16.5}$	I would like to have more information about the associated risks to my privacy by sharing the collected data. ($n_{Q16.5}$=205)
$Q_{16.6}$	I would like to have more information about the associated personal and social advantages by sharing the collected data from my smart home environment. ($n_{Q16.6}$=206)
$Q_{16.7}$	I would like to be able to control which information is collected about myself in a smart home environment. ($n_{Q16.7}$=206)
$Q_{16.8}$	I would like to be able to control the data shared by my smart objects. ($n_{Q16.8}$=206)
$Q_{16.9}$	I would like to be able to determine who is able to access my data. ($n_{Q16.9}$=206)
$Q_{16.10}$	I would like to be able to determine which information is used for which purpose. ($n_{Q16.10}$=206)
$Q_{16.11}$	I would be willing to spend time to audit the data collected about myself in a smart home environment. ($n_{Q16.11}$=205)
$Q_{16.12}$	I would prefer having an automated system taking privacy decisions for me after learning about my privacy risk awareness. ($n_{Q16.12}$=206)
$Q_{16.13}$	I would like to have clear policies with the provider regarding the collected data from my own smart home environment. ($n_{Q16.13}$=205)

Table 1. Submitted statements in the Q_{16}: "Please enter your answer regarding the following statements."

The outliers present the divergent answers from the frequently mentioned answers of the participants and each number presents the data set of the correspondent anonymous respondent.

About 94% of the participants indicated that they want to have more information about the data collected by smart objects about themselves in a smart home environment ($Q_{16.1}$). 96% also precised that they want to have an overview of all the information collected by used smart objects ($Q_{16.2}$). About 94% of the participants would like to see a summary of the collected data over a given period, such as daily, weekly, monthly ($Q_{16.3}$). Additionally, about 84% of the consumers want to have more information about collected data in their own smart home environments in real-time ($Q_{16.4}$).

In addition, for more transparency approx. 92% of our sample want to have more information about the associated risks to their privacy resulting from sharing the

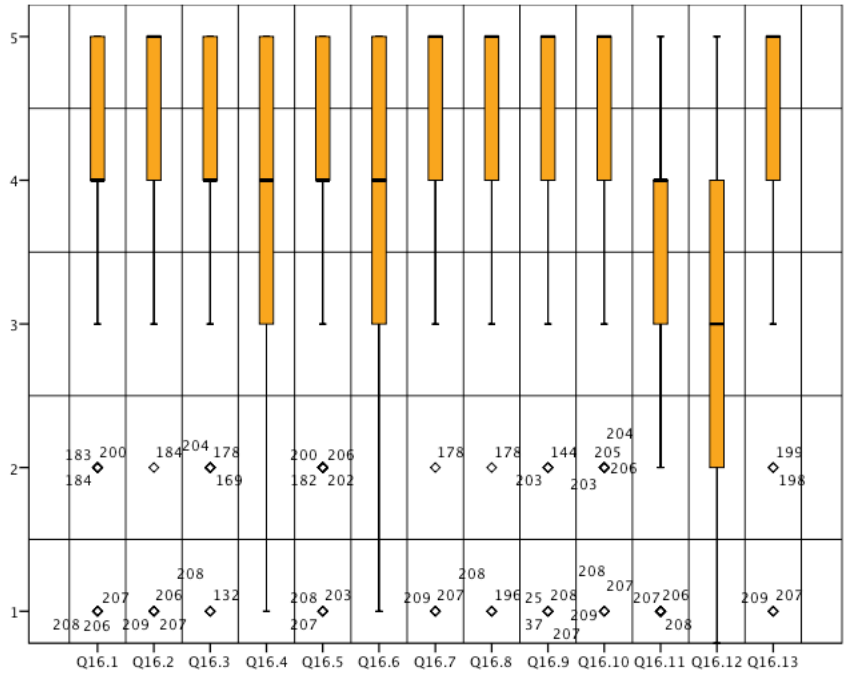

Fig. 2. Boxplots for submitted statements in the Q_{16} (1 = strong disagreement, while 5 = strong agreement)

collected data ($Q_{16.5}$). About 87% of the participants also want to have more information about the associated personal and social advantages by sharing the collected data of their own smart home environment ($Q_{16.6}$). Further analysis shows that there is a positive correlation between statements $Q_{16.5}$ and $Q_{16.6}$ (r = 0.608, significant at the 0.01 level - 2-tailed). This confirms that the users, who want to have information about associated risks to their privacy by data sharing, at the same time also want to have information about the associated personal and social advantages resulting from data sharing.

Almost all participants (97%) indicated that they would like to have control about the data collected and shared by smart objects ($Q_{16.7}$ and $Q_{16.8}$). Note that there are no statistically significant differences between the answers given by participants belonging to different user profiles (Q_5, Mann-Whitney test, $p - values > 0.05$) and users applying special measures to protect their privacy (Q_{10}), as shown by a multiple linear regression test ($p - values > 0.05$).

Additionally, a large majority would like to determine which third parties are able to access their collected data (95% for $Q_{16.9}$) and for which purpose (95% for $Q_{16.10}$). To exercise this control, only 86% of the participants are willing to spend time on auditing the collected data ($Q_{16.11}$). An automated system taking privacy decisions would be supported by 74% of the participants ($Q_{16.12}$). About 96% of the participants also mentioned that they are willing to have clear policies

with the data provider regarding the data collection in smart home environments ($Q_{16.13}$).

We finally asked the participants to indicate their motivating factors to use smart home objects while having the control on the data collected and shared (Q_{17}, $n_{Q_{17}} = 209$) and most cited reasons were: "Having control about the usage of collected data about me" (32%) followed by "feeling myself secured and protected" (29%) as well as "avoiding to draw a digital biography" (22%) and "having information about the data consumer of my data" (15%). Two participants wrote in the free text box that nothing motivates them to use smart home objects while having the control on the data collected. Two further participants also mentioned that they are not going to use any smart objects because they believe that there is no security while using those objects. One participant explicitly indicated that s/he never wants to waste any time on validating or examining the collected data.

Derived Hypotheses: The analysis of participants' answers regarding Q_{16} make it obvious, that the majority of the participants want to have more transparency and associated data sharing information. It is still to be verified whether they want to have these information in order to consider this input while controlling the data sharing process. For this purpose, we derived and tested five hypotheses to verify, whether there is a significant dependency regarding the fact that the users want to have more transparency and associated data sharing information in order to consider this input while controlling the data sharing process. Fisher's Exact tests confirm the all five hypotheses with $p = 0.00 < 0.05$. This means that users want to control the information collected by smart objects (H_1) and are willing to have an overview of those information (H_3). The results also confirm that the users are willing to get information regarding the privacy risks arising from the publication of data (H_2). Similarly, the test results also confirm that the users want to determine who is able to access the shared data (H_4) and for which purpose the data are used (H_5), while controlling the data shared with third parties.

5 Derived Requirements for User-Centric-Privacy-Preserving solutions

We leverage the results of the survey to derive requirements in form of control mechanisms for user-centric-privacy-preserving solutions in what follows. We define the identified control mechanisms as *User-Centric-Control-Points (UCCP)*.

Data object tagging: Considering the results of the whole survey, we can derive that it will be useful to allow the user to tag his/her different smart objects as sensitive or non-sensitive object depending on the data the objects collect. For example, in one case a smart table mat could be non-sensitive, because it just collects information whether something is on the mat or not, while in another case smart fridge or calory scanner could be tagged as sensitive, because those objects collect data regarding the users' eating and living habits. The users can consider these tagging when they make their decision whether they want to share the data while considering the associated privacy risks and advantages

arising from sharing the collected data. These considerations allow us to derive the UCCP 1: Allowing the user to tag the smart object as a sensitive or non-sensitive object.

Data minimization: Our results in Sec. 4.3 and 4.4 show that the participants do not have transparency and vast experience regarding the data collection and disclosure process in smart home environments. Participants' answers also outline that they want to have more information regarding the data collection process in smart home environments. Furthermore, the results underline that the participants want to control which information are collected in their smart home environment. These results help us to derive the following UCCP 2: Allowing the user to select which information are collected by the used smart objects in his environment.

Data granularity: The results in Sec. 4.4 let us conclude that (1) the participants want to have an overview of all collected information and (2) that they want to review the collected data over a preferred period, such as weekly, monthly, and thus to determine the granularity of data collection. Based on these results, we can derive our next UCCP 3: Allowing the user to set in which granularity the data are collected for users' review.

Data sharing: The participants' answers in Sec 4.4 also outline that they want to have the opportunity to get more information regarding associated privacy risks, personal and social associated advantages resulting from sharing the collected data. This leads us to derive our next UCCP 4: Allowing the user to view the associated risks and social or personal advantages arising by sharing the collected context-data.

Data disclosure limitations: The results in Sec. 4.4 show that the users want to control the data shared. In this context, it is to be mentioned that the GDPR also demands to obtain consent for the processing of the personal data in understandable and simply accessible form from the users [1]. These results help us to derive another UCCP 5: Allowing the user to control the data sharing. This UCCP must include at least the following two options: Share the data or delete the data without any third parties getting access to the data.

Data access limitations: The results in Sec. 4.4 show that the participants want to have the control on who is able to access their data and for which purpose in case of data sharing. This leads us to derive the next UCCP 6: Allowing the users to determine who is able to access the data and for which purpose the data are used. This UCCP should also allow the users to set the settings, how the data are disclosed, anonymized or non-anonymized.

As listed above, the results of our survey allowed us to derive six UCCPs as requirements for end user-centric-privacy-preserving solutions for smart home environments. Furthermore, the derived UCCPs can also be categorized into three categories. These categories are transparency of data collection, data implications and data access. In the first category, UCCPs are summarized, which allow users to gain more transparency regarding data collection. The second category includes UCCPs, which provide the data sharing information for users.

The third category comprises UCCPs, which allow users to control the data sharing process.

Category	UCCPs	Short Description
Transparency of data collection	UCCP 1: Data object tagging	Allowing the user to tag the smart object as a sensitive or non-sensitive object
	UCCP 2: Data minimization	Allowing the user to set which information are collected by the used smart objects in his environment(Q_7, Q_8, Q_9, $Q_{16.1}$, $Q_{16.7}$)
	UCCP 3: Data granularity	Allowing the user to set in which granularity the data are collected and saved for users' review ($Q_{16.2}$, $Q_{16.3}$)
Data implication	UCCP 4: Data sharing	Allowing the user to view the associated risks and social or personal advantages arising by sharing the collected context-data ($Q_{16.5}$, $Q_{16.6}$)
Data access	UCCP 5: Data disclosure limitations	Allowing the user to control the data sharing: Share the data or delete data without any third party getting access to the data ($Q_{16.8}$)
	UCCP 6: Data access limitations	Allowing the users to determine who is able to access the data and for which purpose the data are used and how the data are disclosed, as anonymised or non-anonymised data ($Q_{16.9}$, $Q_{16.10}$)

Table 2. Derived UCCPs for user-centric-privacy-preserving solutions based on the results

In future work, the privacy-preserving solutions with integrated UCCPs must be investigated in terms of their usability and applicability in everyday life.

6 Discussion

The answers to the questions on information collection and disclosure ($Q_{6,7,9}$) represent an interesting aspect. Regarding the results of Q_6 a majority (93%) of the participants are aware about the data collection in smart environments, but only 58% agreed in Q_7 that they know the information collected by smart objects. The comparison of the user profiles shows that the frequent and average users are more aware of the data collection than the non users. Furthermore, only 24% of the participants indicated in Q_9 that they know the third parties, who receive the data collected by smart objects. By considering these answers, it becomes obvious that the users have a lack of knowledge regarding the sensitive data collected by the smart objects and the third parties receiving those sensitive data without users' consent. This might have two reasons: (1) users have less transparency about the collected data and/or (2) users put less effort in finding out which information are collected, because they do not receive such information in an understandable way. Although approx. 93% indicated in Q_{15} that they "...believe that smart objects can endanger my privacy", only 55% in Q_{14} mentioned potential privacy risks in IoT-context and only 38% mentioned in Q_{10}

that they are taking special measures to protect their own privacy. Regardless of the derived user profiles and users applying special privacy preserving measures, later in Q_{16} a majority indicated that they want to have control over data collection and disclosure in their smart home environment. This might mean that the missing transparency about collected as well as disclosed data and missing opportunities for the users to control the data collection and disclosure process give only limited permission for the users to be responsible for their own privacy protection. Additionally, it is not clear whether the 38% of the respondents (Q_{10}) apply those measures regularly or just once in a while. If the measures are applied regularly, then it is clear that those participants actively protect their privacy. Furthermore, there were also few participants, who mentioned that they do not have any motivation to deal with user-centric-privacy-preserving solutions, because they believe that there is no privacy in today's data-driven world.

Additionally, testing the five hypotheses (presented at the end of Sec. 4.4) helps us to conclude that users want to have more transparency and information regarding data collection and disclosure process in their smart home environment in order to consciously control the disclosure of the collected data. The results of $Q_{16.11}$ also provide the insight that the users want to be involved in their privacy protection while living in smart home environments. These results are not surprising and further emphasize that efficient user-centric-privacy-preserving solutions for data control are necessary. Consequently, the majority of our participants mentioned in Q_{17} that they are motivated to live in smart home environments while having control over the data collected and shared. In contrast, few participants pointed out that they will not use such devices due to privacy issues. Furthermore, based on the results regarding Q_{16} we were able to derive six UCCPs as requirements for user-centric-privacy-preserving solutions for smart home environments, explained in Sec 5. The derived UCCPs underline the aspects of GDPR [1] and can only provide added value if they are considered in the whole lifecycle of the personal data processing in smart home environments. Furthermore, the presented technical approaches in the second category in Sec. 2 can be considered in the implementation of the derived UCCPs, for instance RBAC mechanisms in the implementation of UCCP 6. The derived UCCPs based on users' answers represent their stated opinion and must be evaluated in a real smart home environment scenario. This will help us to find out whether the users accept to spend their time with such solutions in their everyday lives in order to protect their own privacy, as mentioned in $Q_{16.11}$.

Finally, our questionnaire-based survey has few limitations: As already preluded, the answers of the participants represent their opinion, but not necessarily their actual behavior. The participants may also be biased and not representative of the whole population. Indeed, the fact that they voluntarily answered the questionnaire, which was published on several Internet platforms and invitations sent by emails, may indicate that they may be more altruistic or that they are strongly willing to live or to deal with smart objects and environments than those who have not answered it. Ultimately, our findings mainly reflect the views of participants, who have access to the Internet.

7 Conclusions and Outlook

Within the scope of this paper, we have investigated based on the questionnaire-based survey (1) requirements for end user-centric privacy-preserving solutions and (2) users' readiness to be involved in their own privacy protection. Overall, our participants have indicated that they would like to have more transparency regarding data collection and more control over data collection and disclosure in smart home environments. Based on their answers, we have developed a set of requirements called UCCPs for privacy-preserving solutions in smart home environments that would allow users to exercise a control over their personal data. Our findings also underline that the participants want to be involved in their own privacy protection.

In future work, we plan to conduct user studies to investigate possible discrepancies between users' real behavior and stated opinion regarding the utilization of privacy-preserving solution with integrated UCCPs in smart home environments. We further plan to investigate the usability aspects of such solutions. Finally, further research work is needed to develop clear policy frameworks regarding the personal data processing in smart home environments, which have to be taken into account by the smart objects' providers.

8 Acknowledgement

We would like to thank Michael Friedewald for his helpful comments and the survey participants. Furthermore, we would like to thank Daniel Franke for providing us feedback on early versions of our questionnaire as well as Birgit Schuhbauer for proofreadings.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ **L119/1**, pp. 1 – 88 (2016)
2. Alcaide, A., Palomar, E., Montero-Castillo, J., Ribagorda, A.: Anonymous Authentication for Privacy-Preserving IoT Target-Triven Applications. Computers & Security **37**, pp. 111–123 (2013)
3. Apthorpe, N., Shvartzshnaider, Y., Mathur, A., Reisman, D., Feamster, N.: Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **2**(2, Article 59), pp. 1–23 (2018)
4. Barhamgi, M., Yang, M., Yu, C.M., Yu, Y., Bandara, A.K., Benslimane, D., Nuseibeh, B.: Enabling End-Users to Protect their Privacy. In: Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security. pp. 905–907 (2017)
5. Cao, H., Liu, S., Guan, Z., Wu, L., Deng, H., Du, X.: An Efficient Privacy-Preserving Algorithm Based on Randomized Response in IoT-based Smart Grid. In: 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. pp. 881–886 (2018)

6. Cao, J., Carminati, B., Ferrari, E., Tan, K.L.: Castle: Continuously Anonymizing Data Streams. IEEE Trans. Dependable Secure Comput. **8**(3), pp. 337–352 (2010)
7. Carretero, J., García, J.D.: The Internet of Things: Connecting the World. Pers. Ubiquit. Comp. **18**(2), pp. 445–447 (2014)
8. Chakravorty, A., Wlodarczyk, T., Rong, C.: Privacy Preserving Data Analytics for Smart Homes. In: 2013 IEEE Security and Privacy Workshops. pp. 23–27 (2013)
9. Chan, E.M., Lam, P.E., Mitchell, J.C.: Understanding the Challenges with Medical Data Segmentation for Privacy. In: Usenix Conference on Safety, Security, Privacy and Interoperability of Health Information Technologies. pp. 1–10 (2013)
10. Coopamootoo, K., Gross, T.: Why Privacy Is All But Forgotten. Proceedings on Privacy Enhancing Technologies **4**, pp. 97 – 118 (2017)
11. Day, M., Turner, G., Drozdiak, N.: Amazon workers are listening to what you tell Alexa https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio
12. Friedewald, M., Da Costa, O., Punie, Y., Alahuhta, P., Heinonen, S.: Perspectives of ambient intelligence in home environment. Telemat. Inform. **22**, pp. 221–238 (2005)
13. Guo, L., Dong, M., Ota, K., Li, Q., Ye, T., Wu, J., Li, J.: A Secure Mechanism for Big Data Collection in Large Scale Internet of Vehicle. IEEE Internet Things J. **4**(2), pp. 601–610 (2017)
14. Huang, X., Craig, P., Lin, H., Yan, Z.: SecIoT: a Security Framework for The Internet of Things. Secur. Commun. Netw. **9**(16), pp. 3083–3094 (2016)
15. Huang, X., Fu, R., Chen, B., Zhang, T., Roscoe, A.: User Interactive Internet of Things Privacy Preserved Access Control. In: 2012 International Conference for Internet Technology And Secured Transactions. pp. 597–602 (2012)
16. Hussain, S.H., Geetha, S., Prabhakar, M.A.: Design and Implementation of an Adaptive Model for Sustainable Home Automation using Internet of Things (IoT). Int. J. Adv. Eng. Tech. **VII**(1), pp. 827–829 (2016)
17. Jia, Y.J., Chen, Q.A., Wang, S., Rahmati, A., Fernandes, E., Mao, Z.M., Prakash, A., Unviersity, S.J.: ContexIoT: Towards Providing Contextual Integrity to Appified IoT Platforms. Network and Distributed System Security Symposium (NDSS) pp. 1–15 (2017)
18. Karaboga, M., Matzner, T., Morlok, T., Pittroff, F., Nebel, M., Obersteller, H., Ochs, C., von Pape, T., Pörschke, J.V., Schütz, P., Simo Fhom, H.: Das versteckte Internet: Zu Hause - im Auto - am Körper. White paper, Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt (2015)
19. Khan, M.S.N., Marchel, S., Buchegger, S., Asokan, N.: ChownIoT: Enhancingi IoT by Automated Handling of Ownershio Change. In: Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data (IFIP AICT 547). pp. 1– 15 (2018)
20. Kokolakis, S.: Privacy Attitudes and Privacy Behaviour: A Review of current Research on the Privacy Paradox Phenomenon. Comput. Secur. **64**, pp. 122–134 (2017)
21. Lee, H., Kobsa, A.: Understanding User Privacy in Internet of Things Environments. In: 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT). pp. 407–412 (2016)
22. Li, X., Niu, J., Bhuiyan, M.Z.A., Wu, F., Karuppiah, M., Kumari, S.: A Robust ECC-based Provable Secure Authentication Protocol with Privacy Preserving for Industrial Internet of Things. IEEE Trans. Ind. Informat. **14**(8), pp. 3599–3609 (2017)

23. Martin, K., Nissenbaum, H.: Measuring Privacy: an Empirical Test Using Context to Expose Confounding Variables. Colum. Sci. & Tech. L. Rev. **18**, pp. 176–218 (2016)
24. McReynolds, E., Hubbard, S., Lau, T., Saraf, A., Cakmak, M., Roesner, F.: Toys that listen: A study of Parents, Children, and Internet-Connected Toys. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 5197–5207 (2017)
25. Naeini, P.E., Bhagavatula, S., Habib, H., Degeling, M., Bauer, L., Cranor, L.F., Sadeh, N.: Privacy Expectations and Preferences in an IoT World. In: Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017). pp. 399–412 (2017)
26. 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), pp. 5943–5964 (2016)
27. Pasquale, F.: The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press (2015)
28. 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 (ACM). pp. 83–92 (2016)
29. 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. Future Gener. Comp. Sys. **33**, pp. 11–18 (2014)
30. Udoh, E.S., Alkharashi, A.: Privacy Risk Awareness and the Behavior of Smartwatch Users: A Case Study of Indiana University Students. In: 2016 Future Technologies Conference (FTC). pp. 926–931 (2016)
31. 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)
32. Yang, J.C., Fang, B.X.: Security Model and Key Technologies for the Internet of Things. The Journal of China Universities of Posts and Telecommunications **18**, pp. 109–112 (2011)
33. Yang, L., Humayed, A., Li, F.: A Multi-Cloud based Privacy-Preserving Data Publishing Scheme for the Internet of Things. In: Proceedings of the 32nd Annual Conference on Computer Security Applications (ACM). pp. 30–39 (2016)
34. Yang, W., Li, N., Qi, Y., Qardaji, W., McLaughlin, S., McDaniel, P.: Minimizing private data disclosures in the smart grid. In: Proceedings of the 2012 ACM conference on Computer and communications security. pp. 415–427 (2012)
35. Yu, T., Sekar, V., Seshan, S., Agarwal, Y., Xu, C.: Handling a Trillion (unfixable) Flaws on a Billion Devices: Rethinking Network Security for the Internet-of-Things. In: Proceedings of the 14th ACM Workshop on Hot Topics in Networks. No. Article 5 (2015)
36. Zeng, E., Mare, S., Roesner, F.: End User Security and Privacy Concerns with Smart Homes. In: Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security (SOUPS '17). pp. 65–80 (2017)
37. Zheng, S., Apthorpe, N., Chetty, M., Feamster, N.: User Perceptions of Smart Home IoT Privacy. Proceedings of the ACM on Human-Computer Interaction **2**(CSCW 200), pp. 1–20 (2018)
38. 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), pp. 1606–1616 (2019)

A Appendix - Survey questions

Q_1: Have you heard about the Internet of Things (IoT)?
Possible answers: *Yes / No / I prefer not to answer this question*

Q_2: If yes, in which context have you heard about IoT?
Possible answers: *smart home / Industry 4.0 / smart factory / smart city / smart things / Others: free text box for participants / I prefer not to answer this question*

Q_3: Indicate the smart objects or applications (apps) that you know or use in your everyday life? (multiple choice possible)
Possible answers: *smart fridge / controlling home technology apps / smart grid apps / smart bulbs / smart alarm clock / smart toothbrush / smart washing machine / smart voice control objects, such as Amazon echo / augmented/virtual reality glasses / smart scale / smart health devices / smart door/window locks / smartphone / Others: free text box for participants / I do not utilize smart objects / I prefer not to answer this question*

Q_4: If you already use smart objects, how do you get access to the collected data from your smart objects, through an app or web interface? (multiple choice possible and please click on respective smart object to choose the option between app and web interface)
Possible answers: *smart fridge / controlling home technology apps / smart grid apps / smart bulbs / smart alarm clock / smart toothbrush / smart washing machine / smart voice control objects, such as Amazon echo / augmented/virtual reality glasses / smart scale / smart health devices / smart door/window locks / smartphone / I prefer not to answer this question*

Q_5: How frequently do you use a device connected to the Internet, such as smart scale, fridge, wearables, watch, etc.? (smartphone, computer, smart TV does not count as smart devices in this question).
Possible answers: *more than 20 times per day / less than 20 times per day / once per day / very rare / other options: free text box for participants / I do not use any smart objects / I prefer not to answer this question*

Q_6: When using smart objects, I believe that those objects collect information about myself and my environment.
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_7: I believe that I know the information collected by smart objects. 5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_8: Please indicate which kind of information you believe the smart objects collect about you. (multiple choice possible)
Possible answers: *location data / browsing data / data about your health, such as weight, movements, food purchase / personal data, e.g. bank details, relationships from voice control, finger prints from door locks etc. / other information: free text box for participants / I prefer not to answer this question*

Q_9: I believe that I know the parties who have access to collected data and receive the collected data from my smart objects. (Parties can be: hospital, doctors, insurance companies, institutes using data for statistics, etc.)
Possible answers: *I know / If you know, please mention in short key points the parties / I do not know / I prefer not to answer this question*

Q_{10}: I take special measures (such as switching off some services etc.) to protect my privacy when using smart objects.
Possible answers: *Yes / If yes, please indicate the measures you usually take and the conditions / No / I prefer not to answer this question*

Q_{11}: In a few years, I believe that it will be difficult to live without using smart objects.
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_{12}: What do you think about the advantages you have by using smart home objects? (Participants had the possibility to rate on each statement by using the following Likert Scale.)
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

- facilitating to fulfill the everyday and routine tasks.
- high comfort and convenience
- low error rates (humans make mistakes more easily and frequently than machines)
- automatic adjustments of settings regarding my current lifestyle
- the smart objects record interesting information about myself and my surroundings.
- smart objects outline the optimization potentials regarding my everyday work or my health plan etc.
- specific things are automatically done by smart objects and releasing you from these tasks so you can spend time for more important things.
- no advantages
- Others: free text box for participants

Q_{13}: Please enter the answer for the following question to make sure, that you are not a robot: 150 + (2x2) =

Q_{14}: What do you know about privacy issues in the context of Internet of Things, specifically, to what extent do you understand potential privacy risks? (such as third parties get access to your data / to your house or to your bank account etc.).
Possible answers: *Please enter your answer here / I prefer not to answer this question*

Q_{15}: I believe that smart objects can endanger my privacy.
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_{16}: Please enter your answer regarding the following statements. (Participants had the possibility to rate on each statement by using the following Likert Scale.)
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

- $Q_{16.1}$: I would like to have more information about the data collected by smart objects about me in a smart home environment.
- $Q_{16.2}$: I would like to have an overview of all the information collected by my smart objects.
- $Q_{16.3}$: I would like to be able to control which information is collected about myself in a smart home environment.
- $Q_{16.4}$: I would like to know in real-time about the data collected in my smart home environment.
- $Q_{16.5}$: I would like to have a summary of the collected data over a given period, e.g. daily, weekly, monthly, etc.
- $Q_{16.6}$: I would like to have more information about the associated risks to my privacy by sharing the collected data.
- $Q_{16.7}$: I would like to have more information about the associated personal and social advantages by sharing the collected data from my smart home environment.
- $Q_{16.8}$: I would like to be able to control the data shared by my smart objects.
- $Q_{16.9}$: I would like to be able to determine which information is used for which purpose.
- $Q_{16.10}$: I would like to be able to determine who is able to access my data.
- $Q_{16.11}$: I would be willing to spend time to audit the data collected about myself in a smart home environment.
- $Q_{16.12}$: I would prefer having an automated system taking privacy decisions for me after learning about my privacy risk awareness.
- $Q_{16.13}$: I would like to have clear policies with the provider regarding the collected data from my own smart home environment.

Q_{17}: Indicate the factors that motivate you to use smart home objects while having the control about the data collected.
Possible answers: *Feeling myself secured and protected / It is not possible for third party data consumers to draw a digital biography from my daily routine / Having control about the usage of collected data about me / Others: free text box for participants / I prefer not to answer this question*

Q_{18}: How old are you?

Q_{19}: What is your gender?
Possible answers: *Male / female / I prefer not to answer this question*

Q_{20}: What is your nationality?

Q_{21}: What is your annual income range? (Euro values or equivalent in local currency)?
Possible answers: *Less than 25.000 Euro / 25.000 Euro - 40.000 Euro / 40.000 Euro - 75.000 Euro / 75.000 Euro - 100.000 Euro / More than 100.000 Euro / I prefer not to answer this question*