PriMe: Human-centric Privacy Measurement based on User Preferences towards Data Sharing in Mobile Participatory Sensing Systems

Rui Liu, Jiannong Cao
Department of Computing
The Hong Kong Polytechnic University
{csrliu,csjcao}@comp.polyu.edu.hk

Sebastian VanSyckel
Business School
University of Mannheim, Germany
sebastian.vansyckel@uni-mannheim.de

Wenyu Gao
Department of Statistics
Virginia Tech
wenyu6@vt.edu

Abstract—**Mobile participatory sensing systems allow people with mobile devices to collect, interpret, and share data from their respective environments. One of the main obstacles for long-term participation in such systems is the users' privacy concerns. Due to the nature of these systems, users have to agree to provide some personalized information. Typically, however, people are reluctant to share any information, as it may be sensitive. This is especially the case if the content of the data in question is not completely transparent. In order to increase users' willingness to participate in such systems, we should help users identify which data they can share without violating their personal privacy policies. However, the perception of how sensitive a piece of information is may differ from user to user. In this paper, we propose the human-centric privacy measurement method *PriMe*, which quantifies privacy risks based on user preferences towards data sharing in participatory sensing systems. Further, we implemented and deployed PriMe in the real world as a user study for evaluation. The study shows that PriMe provides accurate ratings that fit users' individual perceptions of privacy, and is accepted by users as a trustworthy tool.**

I. Introduction

Ubiquitous and increasingly capable mobile devices bring forth so-called mobile participatory sensing systems. The idea behind these systems is that individuals and communities use mobile deceives to collect, analyze, and share data regarding their environments for use in discovery. Many such mobile participatory sensing systems have been developed over the years, and some also deployed in the real world [1]. One of the main obstacles for a long-term real-world deployment of such systems is privacy issues. Privacy in a participatory sensing system has particular characteristics. On the one hand, users have to provide their data in order to participate and keep the system running. On the other hand, users are generally ambivalent when it comes to sharing any information due to privacy concerns [2]. Some works on preserving privacy in participatory sensing systems have been published in recent years, e.g., [3]. However, we argue that privacy is not a static concept, but rather fluid and malleable as the perception of privacy differs from person to person. Users need to understand the implications of the data they are supposed to share regarding their personal privacy in order to make an informed decision about participating in sensing tasks or not. However, assessing the risk to one's personal privacy for every sensing task is very arduous. In order to automate this process, it is necessary to understand and model a user's privacy risk with regard to their personal perception.

In this paper, we propose the human-centric privacy measurement method *PriMe*. To the best of our knowledge, it is the first privacy measurement method for mobile participatory sensing systems that is based on the user's perception. For each sensing task, PriMe quantifies the privacy risks for each user individually based on his/her preferences towards sharing certain types of data. For this, we propose two intuitive properties of user preferences and regard them as metrics: 1) intrinsic sensitivity, i.e., the individual inherent sensitivity, and 2) extrinsic sensitivity, i.e., a person's sensitivity towards different data items due to data features. Then, we determine each users privacy risk by by quantifying and aggregating these two properties. To prove our proposed method, we implemented, deployed, and evaluated PriMe with real world users (65 recruited volunteers from different backgrounds). The results show that PriMe is able to make accurate measurements that satisfy users, and thus are widely accepted as a trustworthy tool.

II. Participatory Sensing and Privacy

Numerous participatory sensing systems have been developed over the years, not only in research, but also in industry. Some popular mobile applications, such as Waze, WeatherSignal and Cicada, have already appeared. A comprehensive survey can, for example, be found in [4]. Most of these systems follow the same design rationale and share a common general architecture, as shown in Fig. 1. Our privacy measurement approach PriMe is applicable to any participatory sensing system with this system model, which we briefly describe next.

Our work presented in this paper provides a privacy measurement method based on users' preferences that enables users to better understand their privacy risks in participatory sensing systems. For this, we first clarify the semantic of privacy in participatory sensing systems.

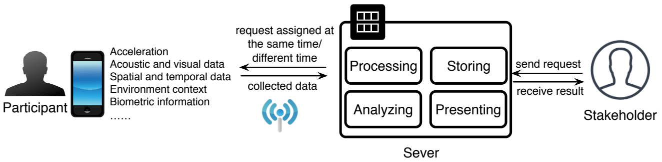


Fig. 1. Typical Architecture of Participatory Sensing Systems

We refer to two acknowledged privacy definitions as proposed in the literature: *"the right to be alone due to private life, habits, acts, and relations"* [5] and *"the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"* [6]. The definitions emphasize that privacy is the ability of an individual or a community to seclude themselves or conceal information about themselves from others. In effect, privacy should be driven by individual preference, instead of being one absolute for all. The same holds true in participatory sensing systems. Therefore, we believe that *privacy in participatory sensing depends on the participants' sensitivities towards the data in question.* We consider two types of sensitivity metric:

1) *intrinsic sensitivity*; This metric indicates the human nature about privacy. However, people's privacy concerns differ from person to person. Some people inherently have a higher/lower sensitivity than others. For example, one person may really care about his/her contacts, whereas others may not care about this information at all.
2) *extrinsic sensitivity*; This metric illustrates the data feature about privacy. Some data are more sensitive than others. For instance, location naturally can disclose more information than the data about your favourite food, no matter what kind of people you are. Extrinsic sensitivity also shows a person's sensitivity may vary with different data and scenarios. For example, a user may care less about his/her weight information in discussion with a health mentor, even though the weight information is very sensitive based on his/her intrinsic sensitivity.

Thus, the two types of sensitivity metric describe subjective and objective factors in a sense. That is the reason why we need to consider both of them. Based on this, PriMe measures each users' privacy risk by quantifying and combining the intrinsic and extrinsic sensitivities of the information in question. Next, we present our approach.

III. Privacy Measurement

We propose PriMe, which can quantify a participant's privacy. The result of PriMe's measurement is a floating point number from zero to one, with a higher number indicating a higher privacy risk. In this section, we elaborate the quantification of privacy by formalizing it in the context of a participatory sensing system.

We assume that there are N participants in a participatory sensing application. Each participant has L data items. Participants can set a privacy tag to each data item to present their willingness to share information associated with the data item. All privacy tags for participants form the $N \times L$ matrix M. At this moment, participants set their privacy tags by use of the dichotomous variable $\{0, 1\}$. More specifically, 0 denotes that the participant does not want to share that datum with anyone, whereas 1 means the participant allows the disclosure of that information. The rows of M correspond to participants, and the columns of M correspond to profile items, respectively. $M_{i,j} = 0$ represents that the profile item j of participant i is private, whereas $M_{i,j} = 1$ denotes the data item is public and can be shared with others. The following two examples further illustrate the situation.

Example 1: In a participatory sensing application, a participant i has three data items, $j_1 = \{email\,address\}$, $j_2 = \{current\,location\}$ and $j_3 = \{age\}$. Thus, $M_{i,j_3} = 1$ means that participant i is willing to share his/her age, and $M_{i,j_1} = 0$ indicates that the participant is reluctant to provide his/her email address. Disclosing the current location or the email address of a user is usually more threatening than revealing a user's age, because the location or the email address – which we then call *sensitive information* – can identify the exact person, whereas age cannot.

Example 2: Another participant i' has the same data items, $j_1 = \{email\,address\}$, $j_2 = \{current\,location\}$ and $j_3 = \{age\}$. However, the participant sets $M_{i,j_1} = 0$ and $M_{i,j_2} = 1$, which means that he/she cares greatly about the email address rather than current location. In this case, location is not a sensitive information any more, even though it is very sensitive to many other people. That is, this particular participant does not feel his/her privacy threatened, if the location information is leaked.

In light of these examples, it is easy to see that a person's privacy preference is a crucial property to determine the privacy threat. Thus, we consider the property of a participant's individual sensitivity (i.e., a participant's privacy preference) rather than the sensitivity of each data because the former is a more inherent property than the latter. Individual sensitivity is an inborn property and can be shaped by a long-term implication of the environment. However, the sensitivity of each data item differs from person to person, scenario to scenario. The general way to identify such sensitivity is to detect from a large sample of people but it still rely on a person's individual attitude towards certain data. According to the two properties in Section II, we define $\delta_{i,j}$ as participant i's extrinsic privacy preference of item j and β_i as the intrinsic sensitivity of participant i. $Pr_{i,j}$ denotes the privacy risk of information item j of participant i when item j is provided. Based on these two parameters, we quantify users' privacy by drawing inspiration from the Rasch Model [7], as shown in Equation 1.

Before we continue presenting our approach, we first discuss why we chose the Rasch Model to quantify the privacy risk in participatory sensing.

$$Pr_{i,j} = \frac{e^{\beta_i - \delta_{i,j}}}{1 + e^{\beta_i - \delta_{i,j}}} \tag{1}$$

The Rasch Model is a psychometric model for measuring/analyzing categorical data as a function of the trade-off between (a) the respondent's abilities and attitudes, and (b) the item difficulty to a particular respondent. A typical application of the Rasch Model is, for example, to estimate the probability of people answering questions correctly based on the ability of a person and the hardness of the question as perceived by the person.

We think the relationship between a users' privacy and their respective attitude towards each item fits this model. More specifically, based on willingness a participant is to reveal his/her information, we also can estimate the probability of a user perceiving a data as sensitive, which can be regarded as a privacy measurement. In the Rasch Model, β_i represents the ability of person i and $\delta_{i,j}$ denotes the difficulty of each question to a specific person, and the result is the probability of a correct response to a given assignment. In our scenario, we can map the two parameters exactly to the sensitivity metrics we describes previously. Thus, there are two parameters, β_i and $\delta_{i,j}$, that need to be computed.

Next, we show how to estimate β and δ based on the matrix M. For this, the maximum-likelihood estimation (MLE) method can be used because this method maximizes the likelihood, or the probability, of our observation and thus is naive and should be the first to think of.

Before we step into the likelihood function, we look deep into the data first. Any participant's decision will not affect the others. Thus the tags are independent across participants. For each participant, the j items can be grouped based on their relative sensitivity to the participant. Some items are similar to a participant so he/she would have the same probability to reveal information on these items. In our setting, $\delta_{i,j}$ would be the same. Thus we can believe that $Pr_{i,j}$ would be the same in one group. The classification of groups need not be the same for different participants. Suppose the j items are classified into G_i groups for participant i and any item falls into a group $g_{ik}, k = 1, 2, \cdots, G_i$. Choices among different groups should be independent for the same participant. Choices within groups should also be independent but identically follow a Bernoulli distribution with parameter $Pr_{i,g_{ik}}$. Therefore, the likelihood function for M is

$$\mathcal{L}(\beta, \delta|M) = \prod_{i=1}^{N} \prod_{k=1}^{G_i} \prod_{j \in g_{ik}} Pr_{i,g_{ik}}^{M_{i,j}} (1 - Pr_{i,g_{ik}})^{(1-M_{i,j})} \quad (2)$$

where $Pr_{i,g_{ik}} = Pr(M_{i,j}|\beta_i, \delta_{i,g_{ik}})$. The estimators are the ones that maximize the above likelihood function, i.e.,

$$(\hat{\beta_i}, \hat{\delta}_{i,g_{ik}}) = \arg \max_{\beta_i, \delta_{i,g_{ik}}} \mathcal{L}(M_{i,j}|\beta_i, \delta_{i,g_{ik}})$$

The values can be achieved simply by taking derivatives of the above likelihood function with respect to the two parameters β and δ, and then setting them to zero. Since the logarithm function is monotonically increasing, we can take the derivatives of the logarithm of the likelihood function and then set them to zero. We follow the steps for each $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, L$. By use of Equation 1, we finally get

$$(1 - Pr_{i,g_{ik}}) \sum_{j \in g_{ik}} M_{i,j} - Pr_{i,g_{ik}}(|g_{ik}| - \sum_{j \in g_{ik}} M_{i,j}) = 0 \quad \text{for } \forall k = 1, \cdots, G_1 \quad (3)$$

where $|g_{ik}|$ is the number of elements in group g_{ik}. Therefore, we achieve

$$Pr_{i,g_{ik}} = \frac{e^{\beta_i - \delta_{i,g_{ik}}}}{1 + e^{\beta_i - \delta_{i,g_{ik}}}} = \frac{1}{|g_{ik}|} \sum_{j \in g_{ik}} M_{i,j} = \bar{M}_{i,g_{ik}} \quad (4)$$

That is,

$$\beta_i - \delta_{i,g_{ik}} = \log \frac{\bar{M}_{i,g_{ik}}}{1 - \bar{M}_{i,g_{ik}}}. \quad (5)$$

Equation 4 makes sense based on our premises. From our settings, participants inherently make the same decisions on items in the same group. The inherent probability that participant i will reveal information on items in this group g_{ik} is $Pr_{i,g_{ik}}$, which can therefore be estimated by the average values of tags in this group.

We only get one equation but we have two unknown parameters. Therefore, we need to seek other methods to estimate the parameters. From Equation 5 we can see that if we know either $\vec{\beta}$ or $\vec{\delta}$ then we will know the other one. Yet it is easy to see that if $\vec{\delta}$ is given then $\vec{\beta}$ will be simpler to solve. Thus we do the estimation by iteration starting from estimating $\vec{\beta}$ given $\vec{\delta}$.

First, we give some initial values to $\vec{\delta_i}$, and then estimate β_i using the Bayesian method. When we get the posterior probability of β_i, it is trivial to estimate β_i by its mode. By standard convention [7], β_i would have a Gaussian prior distribution, with some mean μ and variance σ^2. The posterior probability for β_i is

$$\begin{aligned} &\mathbb{P}(\beta_i|M_{i,j}, j = 1, \cdots, L, \vec{\delta_i}) \\ &= \frac{\mathbb{P}(M_{i,j}, j = 1, \cdots, L|\beta_i, \vec{\delta_i}) f(\beta_i)}{\int \mathbb{P}(M_{i,j}, j = 1, \cdots, L|\beta_i, \vec{\delta_i}) f(\beta_i) d\beta_i} \\ &\propto \mathbb{P}(M_{i,j}, j = 1, \cdots, L|\beta_i, \vec{\delta_i}) f(\beta_i) \\ &\propto \prod_{k=1}^{G_i} \prod_{j \in g_{ik}} \frac{e^{\beta_i - \delta_{i,j}}}{1 + e^{\beta_i - \delta_{i,j}}} e^{-\frac{(\beta_i - \mu)^2}{2\sigma^2}} \end{aligned} \quad (6)$$

Thus, the estimated β_i is

$$\hat{\beta_i} = \arg \max_{\beta_i} \prod_{k=1}^{G_i} (\frac{e^{\beta_i - \delta_{i,g_{ik}}}}{1 + e^{\beta_i - \delta_{i,g_{ik}}}})^{|g_{ik}|} e^{-\frac{(\beta_i - \mu)^2}{2\sigma^2}}. \quad (7)$$

Plugging the result into Equation 5, we can update the estimated $\hat{\delta}_{i,g_{ik}}$ and then begin our iteration until we converge. Thus, the whole procedure of privacy measurement is shown in Algorithm 1. After successfully estimating the parameters β and δ, it is trivial to quantify the individual privacy risk of a data item.

Algorithm 1: Privacy measurement in participatory sensing systems

```
Input: Dichotomous matrix $M$
Output: $\hat{\delta}, \hat{\beta}, Pr_{i,j}$
for $i = 1$ to $N$ do
    Classify the $L$ items into $G_i$ groups;
    for $k = 1$ to $G_i$ do
        $\delta_{i,g_{ik}}$ = initial_value;
    end
    $\vec{\delta} = \{\delta_{i,1}, \delta_{i,2}, \cdots, \delta_{i,L}\}$;
    while convergence do
        Calculate $\hat{\beta}_i$ using Equation 7;
        for $k = 1$ to $G_i$ do
            Calculate $\delta_{i,g_{ik}}$ using Equation 5;
        end
    end
    for $j = 1$ to $L$ do
        Calculate $Pr_{i,j}$ using Equation 1;
    end
end
```

IV. Study Methodology

In this section, we describe the study methodology of our work. First, we present the implementation of PriMe based on the system design, from the App and the server side, respectively. Then, we describe the study procedure to demonstrate our experiments.

A. System Implementation

To evaluate our proposed method, we conduct a user study, in which we focus on environmental noise monitoring as the participatory sensing application. The implementation architecture of the system for the user study is depicted in Fig. 2. The system consists of an App part and a server part. Throughout the paper, we refer to this App part as the *PriMe App*. In the study, the participants are required to install the Android application, to provide them with the participatory sensing function and collect their preferences and feedback towards the privacy assessment of each shared data.

The PriMe App is developed on Android platform. Our prototype of the application is implemented on Android 4.0.3 - 5.1.1 and runs on the Google/LG Nexus 4 handset. There are two design premises of the PriMe App. First, it is a participatory sensing application in nature, so it should receive the sensing tasks and allow the users to provide their collected data. Second, the PriMe App also allows users to express their preferences towards data sharing in the mobile participatory sensing system. According to the premises, we implement and deploy the PriMe App. The key functions are depicted in Fig. 3, which is composed by snapshots of the PriMe App on a Nexus 4 phone.

Fig. 3(a) and Fig. 3(b) show example environment noise sensing acts. In this example, the users make use of PriMe to collect noise data and provide it to the server. They also consider to reject the task due to privacy concerns, as this task requires fine-grained location and calendar information – next to the collected noise data itself. Fig. 3(c) shows the notification to a user when he/she is assigned a new task. The description of the task and the required data are listed in these notifications. Fig. 3(d) shows that PriMe provides an interface for users to express their preferences towards different data items by switching buttons. In the spirit of participatory sensing, users should not set all data types as sensitive, but only those, which they really care about. According to the users' preferences, PriMe then quantifies the privacy of each users' data, as shown in Fig. 3(e).

The server side is designed to receive the collected data from the users and implement our privacy measurement algorithm due to the resource restrictions of mobile devices. As shown in Fig. 2, there are three key components in the server. In the collection part, the server mainly focuses on cleaning and structuring the collected data. The privacy measurement is based on the collected user preferences towards data sharing. In the presentation and reinforcement part, the server can summarize the collected data and revise the measurements' results based on updated user preferences.

B. Study Procedure

We deploy PriMe in the real world in order to study its applicability and perception by its users. For the study, we recruited 65 volunteers for the duration of three weeks. As described in the previous sections, the sensing tasks for monitoring the noise levels in specific areas of Hong Kong are assigned to these participants.

The number of participants is determined according to various influential existing works, e.g., [8] and [9]. Even though our study's size was adequate to evaluate our proposed method with statistical significance, we still plan on conducting a larger study by recruiting participants online in the future. The participants in our user study are from The Hong Kong Polytechnic University, either part time students, full time students, or faculty members. In order to avoid statistical bias and make our results trustworthy, the participants were selected from different backgrounds, genders, and age groups. More specifically, 22 participants are full time students at the university, from various departments. 20 are part-time students, who are also employees in different industries. The remaining 23 participants are faculty members at the university, also from different areas. Details about the participants are shown in Table I.

During the user study, the noise monitoring tasks were randomly assigned to the participants, who were then asked to record the noise signal and upload the data to the server. Further, we asked them to share additional personal information in order to simulate various participatory sensing applications besides noise monitoring. The participant can also decline the task due to their privacy concerns. We did not consider the underlying effect of the decline since we assume the decline is caused by participants' sensitivity. Once PriMe measures a participant's privacy, the server does not send a

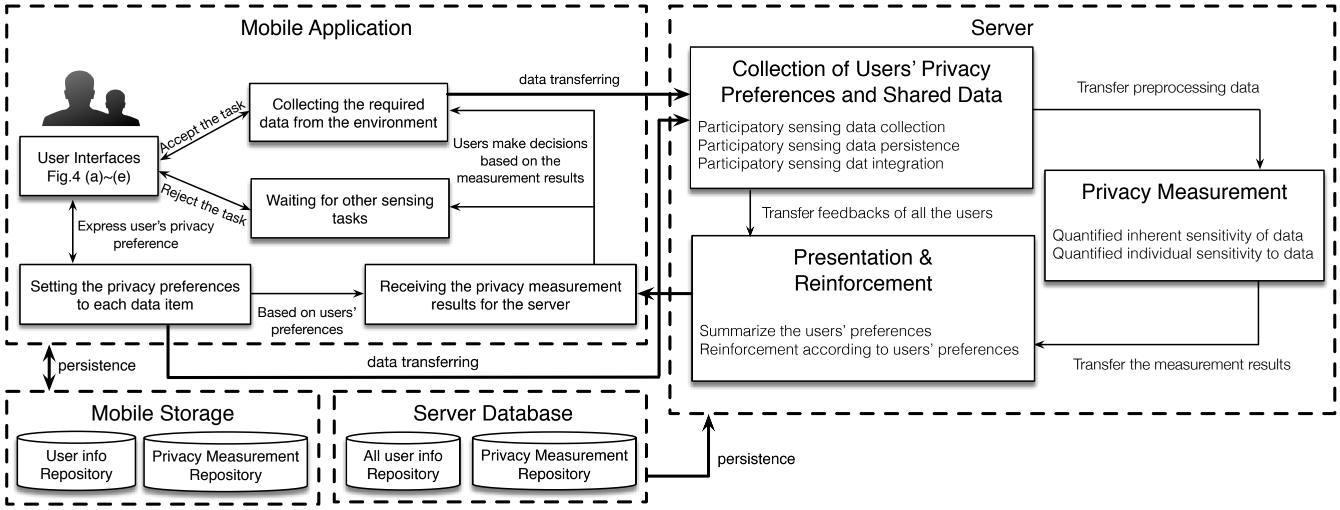
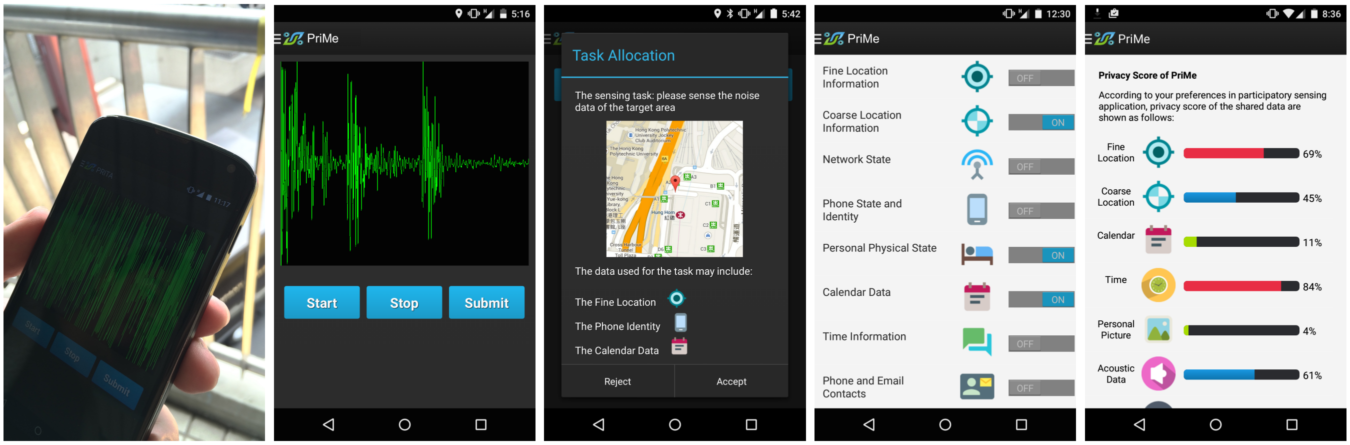

Mobile Application
Server
User Interfaces Fig.4 (a)~(e)
Accept the task
Reject the task
Collecting the required data from the environment
Waiting for other sensing tasks
data transferring
Users make decisions based on the measurement results
Express user's privacy preference
Setting the privacy preferences to each data item
Based on users' preferences
Receiving the privacy measurement results for the server
Collection of Users' Privacy Preferences and Shared Data
Participatory sensing data collection
Participatory sensing data persistence
Participatory sensing dat integration
Transfer preprocessing data
Privacy Measurement
Quantified inherent sensitivity of data
Quantified individual sensitivity to data
Transfer feedbacks of all the users
Presentation & Reinforcement
Summarize the users' preferences
Reinforcement according to users' preferences
Transfer the measurement results
persistence
data transferring
Mobile Storage
User info Repository
Privacy Measurement Repository
Server Database
All user info Repository
Privacy Measurement Repository
persistence

Fig. 2. Overview of the Architecture of PriMe, including Mobile App and Server.


PriMe
5:16
Start
Stop
Submit
5:42
12:30
8:36
Hong Kong Polytechnic University Jockey Club Auditorium
The Hong Kong Polytechnic University
Hung Hom
Cross Harbour Tunnel Toll Plaza

Task Allocation

The sensing task: please sense the noise data of the target area

The data used for the task may include:

The Fine Location

The Phone Identity

The Calendar Data

Reject Accept

Fine Location Information OFF

Coarse Location Information ON

Network State OFF

Phone State and Identity OFF

Personal Physical State ON

Calendar Data ON

Time Information OFF

Phone and Email Contacts OFF

Privacy Score of PriMe

According to your preferences in participatory sensing application, privacy score of the shared data are shown as follows:

Fine Location	69%
Coarse Location	45%
Calendar	11%
Time	84%
Personal Picture	4%
Acoustic Data	61%

(a) (b) (c) (d) (e)

Fig. 3. Screenshots of the PriMe App on a Nexus 4. The example participatory sensing application in the study is to monitor the noise in Hong Kong. (a) Participants use the App to monitor noise from the environment. (b) In this case, the sensing data is the environment's noise level for a specific time period. (c) Sensing tasks are assigned to participants through the App, including target area and required data. (d) PriMe provides a user interface for participants to choose their privacy preference for each data item. (e) PriMe details the privacy scores of different data items, with higher scores indicating a higher sensitivity of the user towards the data (the original result is from 0 to 1, we present it in percentage).

task which requires highly sensitive data, as perceived by this participant, to this participant, in order to minimize declines. For compiling a list of the most interesting data in terms of frequent use in participatory sensing applications as well as their privacy issues, we studied articles on the Internet, e.g., [10], research papers on privacy in participatory sensing, e.g., [3, 11], as well as tips for security and privacy from official guidelines. Table II lists the resulting set of data, and describes their respective potential privacy risks. In the study, we showed these different data types to the participants in the PriMe App (see Fig. 3(d) and 3(e)), and asked them to express their respective sensitivities.

At the end of the study, the participants were given a final questionnaire to ask for their feedback. Next, we discuss the results of the study as well as the questionnaire.

V. Findings

In this section, we discuss the results of our study, such as the performance of the proposed method, as well as other interesting findings.

A. *Participant Sensitivities*

During the study period of three weeks, we not only delivered sensing tasks to the participants, but also asked them to set their preferences towards sharing certain data types. Fig. 4 shows the sensitivity measurement results of an example participant. The higher the score, the more sensitive the participant is to the data type in question. For example, we can see that the participant considers fine grained location information as well as calendar information as the most sensitive. Meanwhile, Fig. 5 shows a plot of the sensitivity

TABLE I
STATISTICS ABOUT THE PARTICIPANTS IN THE STUDY

Category	Participants	Amount	Percentage
Gender	Male	43	66.2%
	Female	22	33.8%
Age	10-19	1	1.5%
	20-24	15	23.1%
	25-29	22	33.8%
	30-40	20	30.8%
	40+	7	10.8%
Background	Energy	3	4.6%
	Materials	1	1.5%
	Industrials	6	9.2%
	Consumer Discretionary	4	6.2%
	Consumer Staples	4	6.2%
	Health Care	8	12.3%
	Finance	9	13.8%
	IT in Security & Privacy	9	13.8%
	IT(except Security & Privacy)	13	20%
	Tele Services	5	7.7%
	Utilities	3	4.6%
Time users spent on smartphones	Rarely (0 1hr)	4	6%
	Sometimes (1 2hr)	14	21.5%
	Frequently (2 4 hr)	27	41.5%
	Very often (4+ hr)	20	31%
Attitudes towards study	Seriously completed	43	66%
	Normally completed	20	31%
	Hastily completed	2	3%

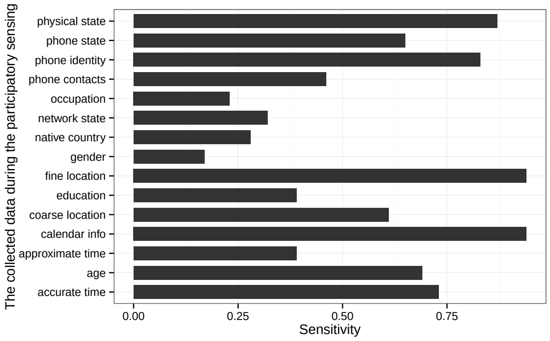


Fig. 4. The sensitivities of one participant to the set of data types.

of all participants to a specific data type – coarse location information in this instance. It becomes clear that the individual sensitivities of the participants regarding the same data type may differ greatly. For example, participant no. 32 has the lowest sensitivity towards coarse location information with around 0.35, whereas participant no. 28 has the highest sensitivity with almost 0.9. This rather large spread confirms our belief that privacy is fluid and its perception can strongly differ from person to person, making personalized privacy measurement approaches necessary.

B. Accuracy

Next, we evaluate the accuracy of the PriMe approach. For this, we compare the privacy measurement results generated by our approach with the participants' sensitivity statements (we discussed in the previous section). The similar the two results are, the higher the accuracy of PriMe is. To quantify the accuracy, we apply the Normalized Distance-based Performance Measure (NDPM) approach [12].

Fig 6 shows the results of the NDPM analysis, from week one to week three. The plots show that PriMe's accuracy increased over the duration of the study, achieving a good accuracy after the third week. Even in the worst case, PriMe

TABLE II
OVERVIEW OF THE MOST COLLECTED DATA IN PARTICIPATORY SENSING AND THEIR POTENTIAL PRIVACY RISKS.

Data Type	Description
Time	Some participatory sensing applications require current time, the format can be shown as 09183302202015. This data will identify temporal information and disclose privacy when other data is combined.
Location	Some participatory sensing applications require current location information, fine-grained location is provided by GPS, coarse-grained location is provided by WiFi or the cellular network. These information will reveal a user's location.
Picture & Video	Pictures and videos are also asked by participatory sensing applications, like taking photos of consumed meals and recording a short video with your family. The content of contributed pictures and collected videos also can reveal personal information about the participants and their environment.
Sound	Sound signals can be captured by smartphones for participatory sensing applications. Given a participant's sound signal, it may allow third parties to determine his/her current context.
Acceleration	Acceleration data is recorded intentionally or automatically during participatory sensing tasks. The data may appear less threatening, but it always can show some clues to leak a participant's privacy.
Environmental Data	Environmental data is often collected since a lot of participatory sensing applications focus on the environment. All the environmental data can indicate spatio-temporal information of the user.
Biometric Data	Biometric data can be used for diagnosis activities in participatory sensing applications. Biometric data normally includes a participant's current physiological state and personal information, such as age and gender. Therefore, privacy will be leaked if the biometric data is identified.

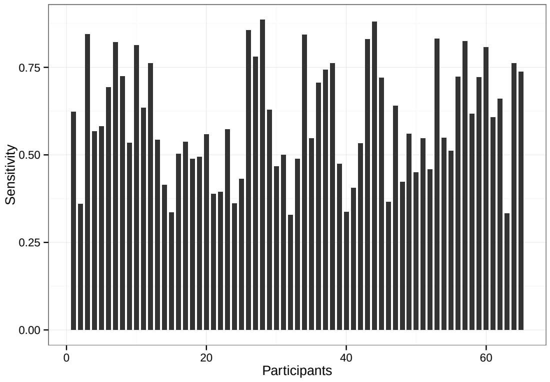
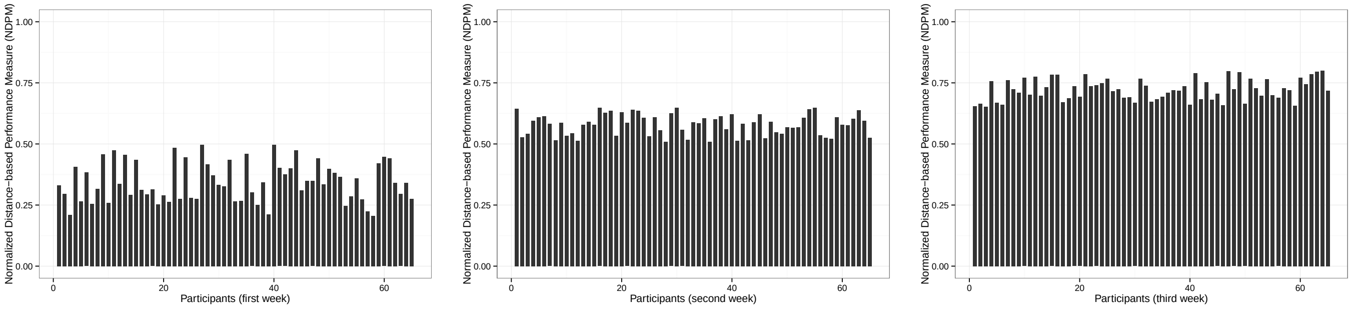


Fig. 5. The participants' sensitivity towards sharing coarse location data.



(a)

(b)

(c)

Fig. 6. The accuracy of PriMe's results compared to the participants' statements using NDPM accumulated by study week.

predicts more than 60% of the participants' rankings. More specifically, the average accuracy is low in the first week due to the limited amount of feedback in the early phase of the study. However, the measurement results steadily increase as we obtain more user feedbacks with regard to their preferences over time. In the third week, we found the accuracy became good to high. In future work, we plan on conducting a longer study in order to further examine the accuracy improvement.

C. Trustworthiness

To test whether the participants trusted PriMe's assessment, we added a proxy function to the PriMe App that allows it to accept sensing tasks automatically on behalf of the users. This proxy function can be activated or deactivated at any time, which at least implicitly indicates the level of trust in the system. Fig. 7 shows the results of the proxy activation recordings. More than 50% of the participants activated the proxy function in the second week. This means they trust the results generated by PriMe after using it for a while. Further, more participants enabled the proxy function in the third week than disabled it. Approximately 18% of the participants did not use the function during the study.

Finally, after the study, we asked the participants to answer a questionnaire on how they felt with regard to their privacy in participatory sensing. Twenty-one participants responded. Many participants noted that PriMe App's explicit listing of which data is needed in order to fulfill a sensing task increased their awareness of their privacy concerns. As examples, we would like to share the following two characteristic comments from participants:

> ”At the beginning, I didn't care about my privacy at all when I accepted sensing tasks, but when I saw the data in my screen, I realized that some sensitive information may be disclosed.”
>
> ”The participatory sensing App looks interesting and I also wanted to publish some tasks using it, but the big privacy risks really discouraged me.”

VI. Related Work

To the best of our knowledge, PriMe is the first privacy measurement method in participatory sensing systems [13].

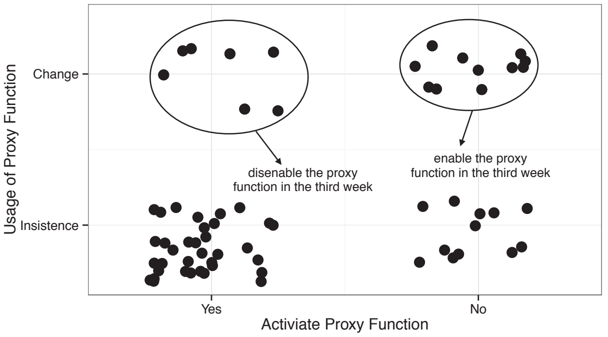


Fig. 7. Scatter plot showing the distribution of participants using the proxy function.

However, some privacy measurements in other systems, e.g., [14–16], as well as privacy preserving mechanisms for participatory sensing, e.g., [17–19], have been proposed. Hence, we discuss works in these two categories next.

A. Privacy Measurements

Privacy measurement has been proposed in various systems for figuring out privacy issues. In [16], in order to quantify the privacy in social networking sites, social cognitive and protection motivation serve as the foundation of the measurement, and a structural equation model is used to analyze the collected data. Emotional reaction is another concern of privacy measurement, as directly asking about a privacy issue may result in an emotional reaction and a biased response. [14] presents indirect techniques for measuring content privacy concerns through surveys in order to diminish emotional responses.

Currently, a lot of research focuses on privacy, and many different metrics have been established. However, how to measure and quantify privacy within the different contexts still needs to be addressed [15].

B. Privacy Preservation

A plethora of works on privacy preserving mechanism has been proposed. PiRi [20] is a privacy-aware framework for participatory sensing that addresses the privacy issues based on an untrusted central data server model and enables participation of the users without compromising their privacy.

In [17], besides the framework, a privacy-preserving participatory sensing scheme for multidimensional data which uses negative surveys has been presented. In this scheme, the server can reconstruct the probability density functions of the original distributions of sensed values, without knowing the participants' actual data. Location privacy is an important concern in participatory sensing systems. [19] presents a decentralized mechanism to preserve location privacy during the collection of sensor readings, which exchanges the sensor readings of users in physical proximity, in order to jumble the location information. The user-side privacy-protection scheme in [18] can adaptively adjust the parameters of participatory sensing for satisfying individual location privacy protection requirements against adversaries in a measurable manner.

VII. Conclusion and Future Work

In this paper, we presented PriMe, a personalized privacy measurement method for mobile participatory sensing systems. Based on the proposed properties of privacy in participatory sensing, we measure the privacy from the perspective of an individual's attitude, which is represented by two intuitive properties, namely the intrinsic sensitivity, i.e., the individual inherent sensitivity, and the extrinsic sensitivity, i.e., the individual sensitivity to different data in different scenarios. The real world study with 65 users shows that PriMe provides reasonable and accurate results, and that the participants, in turn, trust the system to a high degree.

Although we have conducted a system to measure privacy of each user in mobile participatory sensing systems and tested it based on a user study, we acknowledge that it is not the final step for this research. The ultimate objective is to help users to accept or reject tasks automatically based on their privacy concerns. To achieve this, we will refine our system and conduct more large-scale real-world tests (e.g., by releasing our App in App Stores) to get more reliable results. We will also try to provide more options, not only yes or no, to users to collect their preferences, which could increase the accuracy of PriMe with regard to the users' more fine-grained perception of privacy.

VIII. Acknowledgments

We thank our shepherd Prof. Ali Hurson and the anonymous reviewers for their valuable suggestions and feedback. This work is supported by the Germany/HK Joint Research Scheme No. G-PolyU508/13, NSFC/RGC Joint Research Scheme No. N_PolyU519/12, and the National Natural Science Foundation of China with Grant 61332004.

References

[1] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 402–427, 2013.

[2] R. Liu, J. Cao, L. Yang, and K. Zhang, "Priwe: Recommendation for privacy settings of mobile apps based on crowdsourced users' expectations," in *2015 IEEE International Conference on Mobile Services (MS)*, 2015, pp. 150–157.

[3] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *Journal of Systems and Software*, vol. 84, no. 11, pp. 1928–1946, 2011.

[4] P. Y. Cao, G. Li, G. Chen, and B. Chen, "Mobile data collection frameworks: A survey," in *Proceedings of the 2015 Workshop on Mobile Big Data*. ACM, 2015, pp. 25–30.

[5] S. D. Warren and L. D. Brandeis, "The right to privacy," *Harvard law review*, vol. 4, no. 5, pp. 193–220, 1890.

[6] A. F. Westin, "Privacy and freedom," *Washington and Lee Law Review*, vol. 25, no. 1, p. 166, 1968.

[7] T. G. Bond and C. M. Fox, *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press, 2013.

[8] R. Kawajiri, M. Shimosaka, and H. Kahima, "Steered crowdsensing: incentive design towards quality-oriented place-centric crowdsensing," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 691–701.

[9] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "Greengps: a participatory sensing fuel-efficient maps application," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 2010, pp. 151–164.

[10] "12 Most Abused Android App Permissions." http://about-threats.trendmicro.com/us/library/image-gallery/12-most-abused-android-app-permissions, 2013.

[11] J. Freudiger, R. Shokri, and J.-P. Hubaux, "Evaluating the privacy risk of location-based services," in *Financial Cryptography and Data Security*. Springer, 2012, pp. 31–46.

[12] Y. Yao, "Measuring retrieval effectiveness based on user preference of documents," *JASIS*, vol. 46, no. 2, pp. 133–145, 1995.

[13] R. Liu, J. Cao, and L. Yang, "Smartphone privacy in mobile computing: Issues, methods and systems," *DBSJ journal*, vol. 13, no. 1, pp. 1–13, 2015.

[14] A. Braunstein, L. Granka, and J. Staddon, "Indirect content privacy surveys: Measuring privacy without asking about it," in *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS)*, 2011, pp. 15:1–15:14.

[15] X. Page, K. Tang, F. Stutzman, and A. Lampinen, "Measuring networked social privacy," in *Proceedings of the Conference on Computer Supported Cooperative Work Companion (CSCW)*, 2013, pp. 315–320.

[16] N. Mohamed and I. H. Ahmad, "Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from malaysia," *Computers in Human Behavior*, vol. 28, no. 6, pp. 2366–2375, 2012.

[17] M. Groat, B. Edwards, J. Horey, W. He, and S. Forrest, "Enhancing privacy in participatory sensing applications with multidimensional data," in *2012 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2012, pp. 144–152.

[18] B. Agir, T. G. Papaioannou, R. Narendula, K. Aberer, and J.-P. Hubaux, "User-side adaptive protection of location privacy in participatory sensing," *GeoInformatica*, vol. 18, no. 1, pp. 165–191, 2014.

[19] D. Christin, J. Guillemet, A. Reinhardt, M. Hollick, and S. Kanhere, "Privacy-preserving collaborative path hiding for participatory sensing applications," in *2011 IEEE 8th International Conference on Mobile Adhoc and Sensor Systems (MASS)*, Oct 2011, pp. 341–350.

[20] L. Kazemi and C. Shahabi, "Towards preserving privacy in participatory sensing," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 328–331.