Preprint of
“Privacy in Mobile Participatory Sensing: Current Trends and Future Challenges”

By
Delphine Christin

Rheinische Friedrich-Wilhelms-Universität Bonn,
Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
and
Fraunhofer FKIE, Fraunhoferstr. 20
53343 Wachtberg, Germany

E-Mail: christin@cs.uni-bonn.de
http://net.cs.uni-bonn.de/wg/ubips/

Accepted for publication in:
Journal of Systems and Software
DOI: 10.1016/j.jss.2015.03.067

For the most recent version of this article see

The documents distributed by this server have been provided by the contributing authors as a means to ensure timely dissemination of scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the authors or by other copyright holders, not withstanding that they have offered their works here electronically. It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may not be reposted without the explicit permission of the copyright holder.
Privacy in Mobile Participatory Sensing: Current Trends and Future Challenges

Delphine Christin\textsuperscript{a,b}

\textsuperscript{a}Rheinische Friedrich-Wilhelms-Universität Bonn, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
\textsuperscript{b}Fraunhofer FKIE, Fraunhoferstr. 20, 53343 Wachtberg, Germany
E-Mail: christin@cs.uni-bonn.de

Abstract

Mobile participatory sensing has opened the doors to numerous sensing scenarios that were unimaginable few years ago. In absence of protection mechanisms, most of these applications may however endanger the privacy of the participants and end users. In this manuscript, we highlight both sources and targets of these threats to privacy and analyze how they are addressed in recent privacy-preserving mechanisms tailored to the characteristics of participatory sensing. We further provide an overview of current trends and future research challenges in this area.

Keywords: Mobile sensing, participatory sensing, privacy

1. Introduction

Mobile participatory sensing\textsuperscript{1} takes advantage of the sensing, processing, and storage resources available in current mobile phones to gain insights about the participants and their environment. The collected information enables a wide

\textsuperscript{1}Without loss of generality, we use the generic term mobile participatory sensing to designate applications using mobile phones as sensors (or as data sink for interfaced sensors) where participants contribute sensor data. The notion of participatory sensing therefore includes mobiscopes [1] and opportunistic sensing [2], spatial crowdsourcing [3], and mobile crowdsensing [4]. It also covers specific terminologies focusing on particular monitoring subjects, such as urban sensing [2], participatory urbanism [5], citizen sensing [6], people-centric sensing [1, 2], and community sensing [7].
range of innovative applications, ranging from people-centric to environmental-centric scenarios. With the help of people-centric applications, participants can monitor and document health-related issues, such as diet behaviors [8, 9, 10, 11], individual exposure and impact to air pollution [12, 13, 14, 15, 16], depression [17], physical activities [17, 18], sport experiences [19, 20, 21, 22, 23], and stress conditions [24]. Using the sensors embedded in their mobile phone, participants can also enhance social media [25, 26, 27] or contribute to price auditing applications [28, 29]. Additionally, they can contribute to monitor the mobility of crowds at large scale events, such as concerts or festivals [30, 31]. In contrast, environmental-centric applications crowdsource the collection of data about, e.g., urban air pollution [12, 5, 32, 33, 34, 35, 36], noise pollution [12, 37, 38, 39, 40, 41, 42, 43], weather conditions [44], events in the city [45], bus arrival times [46], or thermal columns [47].

Most of these applications have however been conceived as research prototypes and their real-world deployment still remains limited in terms of either number of participants or deployment duration as shown in Table 1. Nevertheless, comparable commercial products are increasingly gaining popularity among the population. For example, mobile applications, such as RunKeeper [48], Endomondo [49], and Nike+ Running [50] documenting running activities and MyFitnessPal [51] measuring calorie intakes totalize between 10 and 50 million installs only on the Google Play store. Connected bracelets, such as FitBit One [52], Jawbone Up [53], and Nike+ FuelBand [54], has represented a 330 million US dollars market in 2013 [55]. These statistics therefore confirm the existence and constant growth of the Quantified Self community [56, 57, 58, 59], which members rely on technology to better quantify, e.g., their fitness, performance, or sleep quality. Similarly, a commercial mobile micro-jobbing application called AppJobber [60] that relies on, e.g., pictures taken by participants counts over 100,000 registered users across different European countries that is more than most deployments compiled in Table 1. Consequently, mobile participatory sensing applications are not only limited to small-scale research prototypes, but are increasingly integrated in our daily life at a larger scale.
Both participants and end users can however put their privacy at risk when interacting with such applications. Since most collected sensor readings are annotated with spatiotemporal information, the participants’ whereabouts can be inferred [61]. The sensor readings themselves may also reveal sensitive information about the contributing participants. Even end users who query application results may disclose their current location and potential interests. In order to address these threats, different countermeasures can be applied as shown in [61]. Building on our previous work, we herewith aim at reporting the latest progress in this area by especially analyzing newly proposed solutions. For this purpose, we adopt the same classification based on the different architectural components of typical participatory sensing applications. Consequently, this manuscript extends our previous survey and makes the following new contributions:

- We first analyze different scenarios in which the privacy of participants and end users may be endangered. As a result, we extend and refine our previous threat model, which primarily focuses on threats coming from administrators of participatory sensing campaigns.

- We next classify and examine how the current state-of-the-art in privacy-preserving solutions tackles the aforementioned threats. In this analysis, we especially consider recent approaches that have emerged in the last three years and are thus not covered in our previous work.

- Based on the surveyed solutions, we finally study the research challenges
which have been addressed and identify those which still need to be tar-
geted in the future.

We present these contributions as follows. In Sec. 2, we recall the typical participatory sensing architecture. We next address privacy issues resulting from potential interactions with participatory sensing applications in Sec. 3, before detailing novel privacy-preserving solutions specially tailored for this scenario in Sec. 4. After examining past and current challenges in Sec. 5, we conclude this manuscript in Sec. 6.

2. Participatory Sensing Architecture

In what follows, we adopt the system model illustrated in Fig. 1, which serves as underlying basis for this manuscript. Participatory sensing applications include the participants’ mobile phones and one or several application servers maintained by the campaign administrators. End users (i.e., the participants themselves or any other persons interested in the campaign) can leverage additional devices, such as laptops or desktop computers, to access the results of the participatory sensing campaign. The different stakeholders typically interact with the system components as follows:

1. **Tasking**: The campaign administrators or end users first determine the sensing tasks to be executed, which are then distributed to the participants’ mobile phone.
2. **Sensing**: The participants’ mobile phone collects the sensor readings corresponding to the defined sensing tasks.
3. **Local processing and storage**: The collected sensor readings may be locally processed on the mobile phone to, e.g., extract interesting features. The processed sensor readings may be temporarily stored by the participants before being transmitted to the application server.
4. **Reporting**: The participants’ sensor readings are reported to the application server for further analysis and display.
5. **Centralized storage and processing**: The application server may first store all reported sensor readings before analyzing them. It can then, e.g., remove incorrect sensor readings, compute summaries, or analyze the reported sensor readings.

6. **Presentation**: End users can access the results of the sensing tasks made available by the campaign administrators. For example, they can query specific sensor modalities or results in a region of interest.

### 3. Privacy Issues

Although participatory sensing systems undeniably provide novel opportunities in terms of sensing, they can put the privacy of the participants and end users at stake. For example, we have observed in our previous survey that most applications collect spatiotemporal information about the participants. This information is usually used to annotate the collected sensor measurements, such as sound samples, pictures, pollution and biometric data or the phones’ acceleration. As a result, they can provide a wealth of insights about the participants, ranging from their current context to their behavior. We therefore highlight potential threats to privacy in Sec. 3.1. We next define a threat model in Sec. 3.2 summarizing possible privacy attacks conducted against the system described in Sec. 2.
3.1. Threats to Privacy

Participatory sensing campaigns first threaten the participants’ privacy. In absence of any protection mechanisms, the participants’ identity and location can be revealed when interacting with the system (e.g., when downloading the tasks from the application server or reporting sensor readings to the server). Assuming that the participants would use pseudonyms to contribute to participatory sensing applications, this would not efficiently protect their anonymity. Indeed, their identity (or at least a link between their different contribution) may still be inferred based on an analysis of the spatiotemporal annotations [62]. For example, a reverse look-up address search may reveal their name, as participants typically commute between their domicile and workplace. Alternatively, a cross-analysis of the participants’ mobility patterns could enable their re-identification based on their uniqueness [63]. Such mobility patterns may also reveal participants’ routines and habits, medical state based on frequent visits to hospitals and political affiliations [23].

Furthermore, the current participants’ location may also be identified based on the collected sensor readings. For example, pictures, audio samples, and pollution data may include unique features, exposing the participants’ whereabouts. When looking closer at the sensor readings, those can reveal fine-grained details about the participants. For example, accelerometer data can reveal the participants’ current activities [64], their identity based on gait recognition [65], and their keyboard inputs, such as passwords [66, 67]. Additionally, participants can be re-identified based on the characteristics of the sensor readings they have collected [68].

Apart from the participants, end users’ privacy can also be endangered. Assuming that end users can send query to the application server to retrieve sensor readings or subscribe to different data streams. By doing so, end users may provide knowledge to the campaign administrators about themselves. Their location can be first disclosed when sending a query, but also based on the query’s content. In fact, it is likely that end users are interested in data collected
in their proximity. Similarly, queries and subscriptions can reveal insights about their personal interests.

3.2. Threat Model

In this section, we analyze different scenarios, in which the privacy of the participants and end users may be endangered. Note that this analysis extends the one conducted in [61], which primarily focuses on threats coming from the campaign administrators.

3.2.1. Internal Threats

We first consider possible threats coming from active stakeholders of participatory sensing applications, i.e., participants, campaign administrators, and end users. Theses threats can vary in terms of (1) degree of involvement, (2) sources and targets, (3) duration, and (4) background knowledge.

Concerning the degree of involvement, different scenarios are possible, ranging from involuntary to deliberate threats to others’ privacy. For example, assuming that a participant reports her precise location to the application server, she can reveal the locations of other participants having reported coarser locations based on the similarities between the sensor measurements. In this case, the disclosure is not the result of a voluntary action conducted by this participant. On the opposite, campaign administrators could, e.g., actively collude with participants, who would report their precise locations on purpose to reveal those of other participants. Between both extrema, we can also imagine honest-but-curious stakeholders. Whereas they would behave normally (i.e., not conducting an active attack), they would be interesting in inferring information about others based on the data they have access to.

Among the different scenarios, both sources and targets can belong to either the same stakeholders’ subset (i.e., participants, campaign administrators, or end users) or different ones. For example, participants may impersonate other participants and reveal sensitive data about them [69]. Alternatively, end users may attempt to deanonymize contributing participants based on the available
data. Moreover, the number of involved sources and targets may vary depending on the attack severity. Additionally, sources and targets can be either in physical proximity or randomly distributed [70, 71].

The duration of potential threats may also be different. For example, we can imagine continuous as well as time-limited attacks [70, 71]. Such attacks could be concentrated around particular times of interest, randomly distributed, or periodically programmed.

Malicious stakeholders may finally come with different background knowledge that they can combine with information available through the participatory sensing applications to ease their attacks or increase their severity [72]. For example, campaign administrators may use the registration information including their home address to re-identify participants using pseudonyms.

Note that we have previously focused on potential threats coming from participants, campaign administrators, and end users. We have deliberately not considered the network operator. The reason behind this exclusion is that the network operator needs to be trusted to respect the stakeholders' privacy, as it de facto has access to the unique phones’ identifier and their locations [73, 74, 75].

3.2.2. External Threats

Like most existing systems, participatory sensing applications can also be subject to external attacks. In this case, an external attacker may attempt to gain insights about the participants and/or the end users. For example, they may be interested in knowing their identity, location, current context, or personal interests. To this end, they may hence attack (1) the application server to get access to the stored data or (2) he communication between the participants and/or the end users with the application server. Again, these threats are not specific to participatory sensing and well-established solutions can be applied to protect the concerned stakeholders from these threats.

In summary, threats to privacy resulting from participatory sensing applications can be diverse. While this list is by no means exhaustive, it aims at
increasing the awareness of future application developers and providing pointers for future privacy-preserving mechanisms.

4. Current State-of-the-Art

We herein present recently published privacy-preserving solutions that address the threats to privacy highlighted in Sec. 3.2. For this purpose, we build upon our previous work and classify novel solutions based on the architecture introduced in Sec. 2. For each studied solution, Tab. 2 summarizes the origin and target of the addressed threats to privacy.

4.1. Tasking

The starting point of all participatory sensing applications is the distribution of the tasks to the participants’ mobile phone. While this first step may seem innocuous, it may already provide insights about the participants’ identity, device, as well as current location when, e.g., downloading the tasks to be executed.

To prevent the campaign administrators from inferring this information, different solutions have already been proposed, such as using tasking beacons, attribute-based authentication, location privacy-preserving routing schemes [76], or downloading the tasks in densely populated locations [77]. In addition to these solutions, the novel PiRi scheme introduced in [78, 79] proposes to rely on the participants to distribute the tasks among them. To this end, each participant defines a region around her current location and merges it with the region of $k - 1$ other participants to obtain a larger region. Instead of transmitting all computed regions, only elected participants transmit their extended region. The tasks distributed by the campaign administrators to these elected participants are then redistributed between all participants according to the region they are able to cover. By doing so, the campaign administrators do not gain access to the individual participants’ location. Based on this approach, the same authors propose TAPAS [80], which aims at improving the quality of the

<table>
<thead>
<tr>
<th>Category</th>
<th>Solutions</th>
<th>Threats to</th>
<th>Threats from</th>
<th>Targeting</th>
<th>Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ID</td>
<td>L</td>
<td>SR</td>
<td>Pa</td>
</tr>
<tr>
<td>Tasking</td>
<td>[76]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[77]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[78, 79]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[80]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[81]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[82]</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[83]</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
<td>(x)</td>
</tr>
<tr>
<td>Sensing</td>
<td>[84]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[86]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
</tr>
<tr>
<td>Local</td>
<td>[87]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>storage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[77]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
</tr>
<tr>
<td></td>
<td>[88]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[89, 90]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
</tr>
<tr>
<td></td>
<td>[91]</td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[92]</td>
<td>x</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[93]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[94]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[95]</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[96]</td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Reporting</td>
<td>[97]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
</tr>
<tr>
<td></td>
<td>[98, 99]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[100]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[101]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[102]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[103]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[104, 105]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[106]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[107]</td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[108]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[109]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[110]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[111]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[112]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[113]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[114, 115]</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[116]</td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[117]</td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[118]</td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[119]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Storage &amp; access</td>
<td>[120, 121]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>control</td>
<td>[122, 123]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Presentation</td>
<td>[125]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[70, 71]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[126]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[73, 74]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>[75]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
collected sensor readings by optimizing the participants’ selection in proximity of a particular location of interest while still protecting their location privacy. In both cases, the privacy protection however fully depends on the participants’ trustworthiness and can hence be endangered as soon as the participants would, e.g., collude with the campaign administrators.

An alternative to PiRi and TAPAS is to use a central trusted entity to build the cloaked region based on the $k$ nearest participants to a point of interest as proposed in [81]. In this case, the participants however provide their location to the trusted entity. They therefore need to trust this entity to efficiently protect their data against external attacks and also not to disclose it to unauthorized entities.

As a result, participants can cloak their location using either distributed or centralized mechanisms based on other participants or a trusted entity, respectively. In both cases, the task distribution between participants using cloaked locations can be optimized. For example, the solution proposed in [127] aims at increasing the task fulfillment rate while simultaneously minimizing the distance to cover by the participants to fulfill these tasks.

Instead of relying on either other participants or a trusted entity, the network provider plays the role of a broker between the participants and the application server in [82]. The participants first transmit their location to the network provider. Note that involving the network provider in this scheme does not increase the risks for the participants’ privacy, as it already knows the participants’ location. The application server can then query the network provider to find participants willing to fulfill tasks in a region of interest. Instead of providing the identity and location of potential participants, the network provider adds random noise to the query results based on the concept of differential privacy [128]. As a result, the application server cannot infer whether a particular participant is located in a particular region based on its queries. Finally, the application server uses the answer of the network provider to contact potential participants using geocasting. If the contacted participants are willing to contribute to the task, they answer back to the application server, which hence
infers their identity. Otherwise, they remain silent and thus unknown from the application server. Consequently, the network provider simultaneously protects the participants’ location privacy by leveraging differential privacy, while allowing the application server to distribute sensing tasks to voluntary participants based on their location. Following the same idea of privacy brokers, the approach proposed in [83] enables end users to directly distribute tasks to the participants. In this case, mobile cloud agents are responsible for managing the distribution of the tasks in a decentralized fashion. Again, the participants need to trust them to respect and apply their privacy preferences.

4.2. Sensing

Assuming that the sensing tasks have been distributed to the participants’ mobile phone. Participants’ may then be able to control ex ante the degree of granularity at which the sensor readings are collected depending on their individual privacy preferences. As enabler for this control, we have proposed different interfaces allowing participants to choose between three degrees of granularity for the collection of location information, sound samples, pictures and acceleration data [84]. We have further introduced picture-based warnings based on the participants’ current privacy settings in [129]. These warnings aim at increasing the participants’ awareness about potential risks for their privacy and allow them to then accordingly adapt their settings if necessary. Note that both approaches can also be applied post ante before sharing the collected data with the application server without loss of generality.

Additionally, the participants can define zones in which no sensor readings are either collected or reported to the application server. Setting such zones may however not be sufficient to efficiently protect the participants’ privacy. For example, they may only contain the participant’s home and hence indirectly reveal her identity. In order to prevent this issue, the solution proposed in [85] includes an underlying mechanism that automatically adapts the zone to cover at least \( k \) buildings. As a result, the so called silent zones dynamically optimize
the existing tradeoff between privacy protection and data granularity based on
the density of surrounding buildings.

4.3. Local Processing and Storage

After the collection of the sensor readings, local processing can be applied
on the participants’ phones to remove privacy-sensitive features. For example,
sound samples including human voices may be discarded. After processing,
the remaining sensor readings may then be stored on either the participants’
mobile phones or individual repositories. For example, participants can securely
store their collected on cloud servers using the approach introduced in [87]. By
encrypting them using proxy re-encryption and homomorphic encryption, the
sensor readings are not accessible by untrusted cloud providers, but can be
accessed by participatory campaigns selected by the participants.

4.4. Data Reporting

When reporting sensor readings to the application server, the following tech-
niques can be applied to protect the participants’ privacy.

4.4.1. Anonymity

To ensure the participants’ anonymity, several solutions leverage the con-
cept of mix networks [130] or mix zones [131]. The key idea is that the partici-
pants’ contributions are mixed with others when sent to the application server
to prevent the campaign administrators from identifying their original source.
Multiple servers ensure this function in [77] by mixing different contributions,
rerouting them, and introducing delays. In [88], the participants themselves
serve as routers and a multi-hop route is built between the participants, along
which each participant only knows her predecessor and successor. When the
sensor readings reach the tenth participant, they are uploaded to the applica-
tion server. This scheme however requires maintaining a peer-to-peer network
between the participants and coping with, e.g., disconnections.

An alternative is to leverage opportunistic encounters between participants
to exchange collected sensor readings when being in physical proximity [89, 90].
This scheme called path jumbling however fully depends on the participants’ trustworthiness and collaboration. To quantify the trust level of the participants and quarantine untrustworthy ones, TrustMeter approach has therefore been introduced in [132]. Participants can leverage it associated with dedicated user interfaces detailed in [133] to set the minimum trust level required by other participants to be able to exchange sensor readings with them. Moreover, the participants can select different exchange strategies between exchanging: (1) all sensor readings, (2) a random number of sensor readings, or (3) a number of sensor readings agreed between both exchange partners. Instead of exchanging full or partial sets of sensor readings, SLICER [92] inspired by [91] proposes to either exchange only one sensor reading at each encounter or select a subset of participants to share more sensor readings with and hence optimize the reporting process.

Instead of considering individual sensor readings, the following mechanisms focus on the participants’ full trajectories. By using [93, 94], the participants can conceal their own trajectories with those of other participants based on a trusted third party. In [93], the trusted third party merges the sensor readings belonging to different participants to build new equivalent trajectories. Alternatively, the trusted third party maps the identities and the participants’ trajectories entering and leaving predefined sensitive zones in TrPF [94]. By doing so, the mixing function can be customized and optimized depending on the desired privacy protection. While this control can be an advantage as compared to the concept of physical mixes, the solution however requires that the participants provide their precise location to this third party.

To prevent the third party from linking the participants’ identity to their location, the participants first collaborate by relaying the other participants’ sensor readings before uploading them to the third party in [95]. The third party then anonymizes the sensor readings and sends them back to the participants following the same route. Since the sensor readings have been previously anonymized, the participants can report them to the application server using their real identity. Note that the number of participants is however unlimited.
as compared to [88].

Such centralized architectures however present a single point of failure, making them vulnerable to malfunctions and external attacks. LOCATE [96] therefore adopts a distributed approach by leveraging a direct collaboration between the participants. In this case, not a third party needs to be trusted, but other participants. To mitigate this trust in other participants, two sets of the original participants’ trajectories are built. Assuming that a user frequently visited $n$ locations. The first set consists of trajectories including $n' < n$ of the frequently visited locations. In the second set, none trajectory contains any of the $n$ locations. The number of trajectories belonging to each set is computed based on the resulting entropy. The participants alternatively exchange trajectories from both sets and distribute the exchanges over time between different participants.

4.4.2. Pseudonymity

To protect the participants’ privacy, various applications replace the participants’ real identity by a unique pseudonym, including [134, 29, 23, 97]. However, it has already been shown in [62] that the provided protection is insufficient, as the real identifies may be inferred based on the reported location information. In [97], the participants have the possibility to mitigate this threat by dropping the collected sensor readings, if they estimate that they could endanger their anonymity. This however requires a manual intervention of the participants before each reporting. Moreover, it is assumed that the participants can correctly identify privacy threats by, e.g., taking into account their past contributions, which may not always be the case.

An alternative to unique pseudonyms is proposed in the IncogniSense framework [98, 99], which relies on the utilization of periodic pseudonyms and a transfer of reputation between these pseudonyms. By doing so, the framework combines to conflicting aspects namely the participants’ privacy and reputation. Indeed, most reputation and incentive schemes require a link between the participants’ contributions over time to, e.g., identify participants contributing incorrect measurements or reward active participants. In contrast, protecting the
participants’ privacy often requires to break the link between the participants’ contributions. Moreover, the chosen approach is independent of any trusted third parties because of the utilization of blind signatures. If blind signatures theoretically ensure that successive pseudonyms cannot be linked, the evolution of the pseudonyms’ reputation may reveal their chronology. To prevent such linking attacks, the reputation transferred between pseudonyms is cloaked to build groups sharing similar reputation values and render the pseudonyms indistinguishable.

Building upon periodic pseudonyms, different alternatives have been introduced. For example, a trusted third party is responsible for building the pseudonyms’ groups [100]. In comparison to IncogniSense, this approach enables a dynamic adaptation of the cloaked values, but the trusted third party knows the link between the participants’ identity and their pseudonyms in addition to their original and cloaked reputations values. Another alternative also based on blind signatures or a trusted third party adds an incentive mechanism on top of the reputation framework to reward contributing participants [101]. The idea of rewarding participants using pseudonyms is shared with [102] and [103]. In the former approach, participants use a unique pseudonym, while they use multiple pseudonyms in combination with a group signature scheme in the latter called SPPEAR. Instead of only supporting positive reputation updates as in the aforementioned approaches, the solution called ARTSense proposed in [104, 105] also allows negative reputation updates. This aims at further penalizing malicious participants by increasing the impact of incorrect measurements on their reputation. When adopting the LotS framework [106], the identity of misbehaving participants may even be revealed.

4.4.3. Spatial Cloaking

As we have seen in Sec. 4.4.1 and 4.4.2, the above solutions protect the participants’ privacy by breaking the link between their identity and contributions. By mixing the contributions between participants or utilizing different pseudonyms, most presented solutions preserve the original spatiotemporal in-
formation. Other methods alter the location information by providing it at a coarser degree of granularity and/or building groups of $k$ participants sharing the same location. Note that a comparison of both methods [72] shows that the degree of granularity has additional influence on the participants’ location privacy than the number of participants sharing the same location.

Instead of selecting a predefined degree of granularity before the measurements (cf. Sec. 4.2), dynamic solutions, such as ipShield [107] and that presented in [108], can be applied. These solutions running on the participants’ mobile phone consider the participants’ current context and past contributions to compute the appropriate degree of granularity at which the location information can be released to the application server. As a result, end users will have access to this data with either the same (if the campaign administrators release the data as such) or a coarser degree of granularity (if the campaign administrators apply further processing).

However, providing both coarse-grained information to the application server and fine-grained information to specific end users can be useful in certain application scenarios. For example, participants may be willing to share detailed information, such as the time of their asthma crisis, with their doctor to analyze its potential environmental causes but not with other end users. This is possible with the scheme proposed in [109]. It releases the location information at two different degrees of granularity: the coarsest one is anonymized and publicly available, whereas the finest one is encrypted using attribute based encryption. This means that only end users showing the attributes defined by the participants (e.g., age, location, role...) will be able to decrypt and have access to the participants’ fine-granular locations.

Instead of adapting the degree of granularity to potential data recipients, the approach introduced in [110] applies two different techniques based on the size of the area of interest, which is assumed to be centered on a predefined point of interest. In the case of large areas, the participants replace their precise location by the nearest point of interest and report the corresponding sensor readings to a broker. The broker aggregates them and transmits the results to
the application server. By applying this technique, the location granularity is hence degraded. In contrast, the participants apply double encryption on the sensor readings including their precise location in the case of small areas. They use first the application server’s public key, before using the broker’s public key. Again, the encrypted sensor readings are sent to the broker, which waits for several participants to do the same before forwarding them to the application server. By doing so, the location granularity is preserved but the link between the participants’ identity and contributions is broken by the intermediary of the broker. Therefore, this hybrid approach dynamically leverages the concepts of either spatial cloaking or aggregation (cf. Sec. 4.4.5) depending on the size of the monitored area.

4.4.4. Data Perturbation

The key principle of data perturbation is to hide the individual participants’ contributions while allowing the application server to computes statistical trends over the whole participants’ set. To hide the individual contributions, the first method consists in adding noise on the participants’ sensor readings. As a result, the noise selection determines the participants’ privacy protection.

Assuming that all participants share the same noise distribution. There is a risk that malicious participants may be able to reconstruct it based on their own data and hence breach the privacy of other participants. To mitigate this threat, the authors propose the PESP scheme [111], which distributes different noise distributions to the participants and adapt them to the sensor readings already reported to the application server.

Instead of automatically adapting the individual noise distributions, ALPS [112] adjusts the perturbation according to the participants’ preferences. As a result, a tailored Gaussian perturbation is first applied followed by a smoothing function to remove potentially remaining insights about the participants and hence protect their privacy.

An alternative to using noise is to leverage the concept of negative surveys as introduced in [113]. In this case, the collected sensor readings are divided
into different complementary categories. Instead of reporting their own sensor readings, the participants choose sensor readings from another category and report those to the application server. Using a perturbation matrix that maps the probability of perturbing a category to another, the application server is able to reconstruct the probability density functions of the original sensor readings without having access to them.

4.4.5. Data Aggregation

The idea behind data aggregation is also to break the link between the participants’ identity and their contribution. In contrast to the mechanisms detailed in Sec 4.4.1, sensor readings from different participants are however merged together to build aggregates. The application server receiving the aggregates is hence unable to isolate individual sensor readings and link them to the collecting participants. Different methods can be applied to aggregate the data, ranging from centralized to distributed solutions. For example, NoiseTubePrime [114, 115] relies on a network of trusted brokers. The participants report their results on a map common to all participants and encrypted using the campaign administrators’ public key to their respective broker. The brokers organized in a ring topology successively add the contribution of their participant until the aggregate map is completed.

By applying the scheme proposed in [116], participants however not need to trust one or several aggregators (such as the brokers in NoiseTubePrime). The authors make use of additive homomorphic encryption combined with a new key management scheme to reduce both the communication and encryption overhead while still supporting sum and min aggregates. To better support the participants’ dynamic (i.e., new participants joining or leaving), the original scheme has been extended in [117] by a novel ring-based interleaved grouping technique that diminishes the number of participants that need to renew their cryptographic keys.

Also not trusting the aggregator, VPA+ [118] adopts a hybrid solution, in which the participants first register their sensor readings to the aggregator
before contributing them to the aggregate computed in a distributed fashion. Using a homomorphic MAC of the participants’ sensor readings, the registration does however not reveal the original data but allows the aggregator to later verify which participants have contributed and hence guarantee the aggregate’s integrity. Similarly, the participants collaborate to compute the sum aggregate without revealing the individual sensor readings.

4.5. Hiding Selective Locations

The solution proposed in [119] is devised for application scenarios, in which the temporal annotations of the sensor readings are irrelevant. In such cases, the participants explicitly choose the sensor readings corresponding to the locations they want to share with the application server. Their mobile phone then mixes them to modify their chronology and hence break the link between both spatial and temporal information. As a result, the application server can compute the application results based on the data voluntarily shared by the participants.

4.6. Storage and Access Control

The sensor readings reported to the application server can be either individually stored or directly processed on the application server to build, e.g., statistics or maps. The participants may thus only maintain control over their data in the former case. For this purpose, they may use dedicated access control and data sharing solutions, such as SensorSafe [120, 121, 122, 123] or PDVLoc [124]. Using SensorSafe, the participants can manage several individual repositories using a broker and tailor the granularity at which the collected data are shared based on, e.g., its nature and context, the resolution required by potential end users, their identity or attributes, as well as the degree of trust of the participants in these end users. Following a similar model, PDVLoc allows participants to select potential data recipients as well as the corresponding degree of granularity at which the data is shared. It further supports the participants in configuring their settings and shows them the consequences of their decisions. An additional function notifies the participants when their current settings seem to diverge from their personal conception.
4.7. Presentation

Participatory sensing results can be made available to end users in different forms. For example, end users may access a map aggregating all participants’ results or be able to query and filter individual results based on, e.g., a region, sensor modality, or operating system of interest. As soon as end users are able to query participatory sensing results, the privacy of both participants and end users may be put at stake.

Among the existing solutions, the one proposed in [125] concentrates on protecting the participants’ privacy. For this purpose, the participants report only partial trajectories to the application server, which then optimizes the query answers by merging partial data collected by several participants. Instead of submitting only partial results, the participants encrypt their complete sensor readings and distribute multiple duplicates to other participants at regular and common time intervals in [70, 71]. The encrypted sensor readings are annotated with a tag corresponding to the time interval at which they have been collected and exchanged. End users can then send a query indicating the sensor readings and interval of interest to a gateway, which will forward it to a set of participants. Participants matching the query provide the corresponding encrypted sensor readings to the end users through the gateway. As a result, neither the gateway nor the end users can infer who has initially collected the sensor readings, as the link between the participants’ identity and their contribution are broken during the exchanges.

Besides, PEPPeR [126] only focuses on end users’ privacy and relies on tokens distributed by the application server to authorized end users. By using them, end users can directly query participants, who provide access to their sensor readings once they have verified the corresponding token with the application server. The verification process however does not disclose the identity of the end users to the application server and preserves hence their privacy.

In comparison, PEPSI [73, 74, 75] simultaneously protects the privacy of both participants and end users. It relies on a trusted third party responsible for their registration and authorization. During the campaign, the registered par-
Participants report their sensor readings encrypted to the application server, while the registered end users send their queries to the application server. Leveraging authorizations and tokens delivered by the trusted third party during the registration process, the application server blindly matches both sensor readings and queries. This means that both collected data and query content remain hidden from the application server.

For the purpose of data aggregation, we construct and analyze an additively homomorphic IBE scheme as a variant of the Boneh-Franklin IBE scheme and prove its security under the Decisional Bilinear Diffie-Hellman assumption in the random oracle model. This IBE scheme can be directly used within our generic collusion-resistant participatory sensing protocols to achieve data aggregation. We note that our additively homomorphic IBE scheme is only suitable for calculations with small sensor readings, which however is sufficient for most of the above mentioned applications of participatory sensing. For all our constructions, we analyze the performance and other comparisons to prior work.

5. Past and Future Challenges

We have highlighted the progress of privacy-preserving schemes for participatory sensing applications in the past years. While this domain still attracts many novel solutions, some privacy challenges still remain. In what follows, we successively consider the research challenges identified in our previous survey and discuss how these have been addressed, before highlighting novel research directions.

5.1. Including Participants in the Privacy Equation

Among the previously identified challenges, this is probably the one which has been the most addressed in the last years. The participants were initially only able to choose the granularity at which sensor readings are collected, manage individual storages, and control the data stored and disclosed to end users. We now observe an increasing inclusion of the participants in the control of
novel privacy-preserving mechanisms. For example, their privacy preferences is taken into account during both the tasking and the reporting processes (see Sec. 4.1 and Sec. 4.4, respectively). To cater for this control, specific interfaces have been designed and evaluated based on user studies. In addition to server in the evaluation of technical approaches, user studies, such as [135] and [136], have been conducted to analyze potential users’ privacy expectations and how these are influenced by contextual factors (e.g., sensing modality, duration, or purpose of the data collection). While the first stones have recently been laid, additional efforts are needed to cater for additional privacy control and awareness. This however remains challenging as it requires to finely balance the trade-off between control and the resulting overhead for the participants. Moreover, finding representative and large user sets makes the evaluation of future approaches demanding.

5.2. Providing Composable Privacy Solutions

All aforementioned approaches still focus on providing a particular mechanism to address one specific threat to privacy at a time. This means that none of them addresses the full spectrum of possible threats to privacy. Application developers still need to select each mechanism individually. Instead, an integrated holistic solution is required to help them in selecting appropriate mechanisms depending on the specifications of their application. Besides, this solution should be easily and intuitively configurable.

5.3. Trade-offs between Privacy, Performance, and Data Fidelity

An increasing number of mechanisms consider potential trade-offs linked to the respect of the privacy of the participants and end users. For example, solutions have been proposed to (1) optimize the sensing coverage during the task distribution process (cf. Sec. 4.1), (2) optimize the incentives given to the participants depending on the expected quality of information [137], and (3) run reputation algorithms (cf. Sec. 4.4.2). All of them simultaneously respect the participants’ privacy. Pursuing in this direction is necessary as both application
utility and participants’ privacy are two tightly-coupled key components in the acceptance of both participants and end users. If the application utility is limited, end users may not be interested in the results. The same is valid for the participants: if their privacy is not respected, they may not be willing to contribute to the application and hence limit its utility.

5.4. Making Privacy Measurable

Most solutions are evaluated using different privacy metrics. Depending on the scenario, the authors either introduce a new metric or choose an existing one. Existing metrics include k-anonymity, l-diversity, entropy-based, error-based, or probabilistic-based [138]. As before, this makes a comparison of the performance of the different schemes difficult. However, more and more solutions adopt a decentralized or distributed setting. By doing so, they reduce the trust that the participants need to have in a unique entity to guarantee the protection of their privacy.

5.5. Defining Standards for Privacy Research

While some studied approaches share the same dataset for their evaluation, there is still no consensus on a common open data set, which would be adopted by all to ease the comparison of the proposed schemes. As compared to other research fields, only few schemes are thoroughly compared against each other especially in terms of performance using, e.g., simulations. Not having a benchmarking corpus in terms of dataset and open privacy solutions hence prevent application developers from easily identifying and selecting the best solution to address privacy threats.

5.6. Holistic Architecture Blueprints

In comparison to our previous survey, we have observed an increased diversity among the sources and targets of privacy threats addressed in the studied approaches. Instead of only focusing on threats coming from the campaign administrators and targeting the participants, recent solutions highlighted in
Tab. 2 address the privacy protection of both participants and end users. This fact might be linked with the evolution of the underlying participatory sensing applications. In the first proposed applications, end users are mostly only able to access already collected information displayed, e.g., in forms of maps or statistics. In contrast, end users can also trigger the collection of sensor readings depending on their own interests in recently developed applications. Having this capability, end users have hence recently moved from a passive to a more active role in the system. Such evolution leads to an increase of the threats to their privacy. Tailored privacy-preserving mechanisms to these specific threats are therefore needed. However, only few solutions address threats coming or targeting different groups of stakeholders. While the existing solutions can still be composed to provide a complete solution, it may be useful for application developers to have a solution covering most to all threats to privacy.

In summary, most of the challenges identified in our previous survey have still not been fully tackled and require additional attention. The following challenges complete our list of previously identified challenges based on the solutions surveyed in this manuscript.

5.6.1. Bystander Privacy

Most of the studied schemes consider each participant as an individual entity and omit their relationships with others (i.e., participants or non-participants). By doing so, this poses the risks for participants to unwillingly reveal information about others. For example, collected sound samples can reveal private conversations of bystanders or similar sensor readings can disclose the proximity of two or more participants as shown in [68]. To prevent this scenario, the participants can leverage NotiSense [86], which notifies the participants about currently running sensing tasks and potential risks for their privacy. While this approach offers a first solution to this issue, we believe that additional alternatives can be devised by, e.g., integrating this aspect in the privacy-preserving solutions to be developed.
5.6.2. Multidimensional Privacy

The aforementioned approaches mainly focus on the annotation of the collected sensor readings. Different solutions are proposed to hide, cloak, perturb, or aggregate them. Only little attention is however given to the collected sensor readings themselves. While the collection of sensor readings is the key characteristic of participatory sensing and solutions tailored to this scenario might hence only be transferable to other research domains to a limited extend, the sensor readings can also endanger the participants’ privacy as highlighted in Sec. 3.1. Additional emphasis should therefore be put on the sensor readings themselves and more especially, on the impact of their combination and correlation on the participants’ privacy.

5.6.3. Internet of Things and Smart Cities

Up to now, the developed mechanisms have been tailored to participatory sensing and only consider data collected using the participants’ mobile phones. However, participatory sensing may not evolve to a stand-alone paradigm. In the future, it may be integrated into the vision of the Internet of Things, where all participants’ devices and appliances are foreseen to be connected and collaborate to fulfill different tasks. In this scenario, the sensor readings collected by the mobile phones might complete the information provided by other devices. By extending the scale of this vision, mobile phones can also be part of the next concept of Smart Cities. Again, bystander privacy (or “passive sensing”) plays a major role here because building owners can collect data about people without their consent. Their integration will however pose new privacy threats and solutions to cover the collection of multidimensional information about the citizens are still to be designed.

6. Conclusions

We have reported on the latest trends in the area of privacy in participatory sensing. For this purpose, we have classified and analyzed recently proposed privacy-preserving approaches. Based on this analysis, we have refined
the threat model that apply in this field and studied how solutions recently emerged address the research challenges identified in [61]. While some challenges have been tackled, there is a strong need to future research, especially when considering the integration of participatory sensing in larger visions, such as the Internet of Things or Smart Cities. By introducing more and more information about the physical world into the digital space, the privacy of the citizens’ will be increasingly endangered and novel approaches tailored to such scenarios are needed.

Acknowledgment

The author would like to thank A. Reinhardt as well as the anonymous reviewers for their valuable comments and suggestions.

References


of the 9th IEEE International Conference on Pervasive Computing and Communications (PERCOM Workshops), 2011, pp. 344–347.


[46] P. Zhou, Y. Zheng, M. Li, How Long to Wait?: Predicting Bus Arrival Time with Mobile Phone Based Participatory Sensing, in: Proceedings of


[89] D. Christin, J. Guillemet, A. Reinhardt, M. Hollick, S. S. Kanhere, Privacy-preserving Collaborative Path Hiding for Participatory Sensing


T. Dimitriou, I. Krontiris, A. Sabouri, PEPPeR: A Querier’s Privacy Enhancing Protocol for PaRticipatory Sensing, in: A. Schmidt, G. Russello,


