The Journal of Systems and Software 84 (2011) 1928–1946

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

A survey on privacy in mobile participatory sensing applications

Delphine Christin[a,*], Andreas Reinhardt[b], Salil S. Kanhere[c], Matthias Hollick[a]

[a] Secure Mobile Networking Lab, Technische Universität Darmstadt, Mornewegstr. 32, 64293 Darmstadt, Germany
[b] Multimedia Communications Lab, Technische Universität Darmstadt, Rundeturmstr. 10, 64283 Darmstadt, Germany
[c] School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 16 February 2011
Received in revised form 29 June 2011
Accepted 29 June 2011
Available online 6 July 2011

Keywords:
Mobile sensing
Participatory sensing
Privacy

ABSTRACT

The presence of multimodal sensors on current mobile phones enables a broad range of novel mobile applications. Environmental and user-centric sensor data of unprecedented quantity and quality can be captured and reported by a possible user base of billions of mobile phone subscribers worldwide. The strong focus on the collection of detailed sensor data may however compromise user privacy in various regards, e.g., by tracking a user's current location. In this survey, we identify the sensing modalities used in current participatory sensing applications, and assess the threats to user privacy when personal information is sensed and disclosed. We outline how privacy aspects are addressed in existing sensing applications, and determine the adequacy of the solutions under real-world conditions. Finally, we present countermeasures from related research fields, and discuss their applicability in participatory sensing scenarios. Based on our findings, we identify open issues and outline possible solutions to guarantee user privacy in participatory sensing.

1. Introduction

In recent times, mobile phones have been riding the wave of Moore's Law with rapid improvements in processing power, embedded sensors, storage capacities and network data rates. The mobile phones of today have evolved from merely being phones to full-fledged computing, sensing, and communication devices. It is thus hardly surprising that over 5 billion people globally have access to mobile phones. These advances in mobile phone technology coupled with their ubiquity have paved the way for an exciting new paradigm for accomplishing large-scale sensing, known in literature as *participatory sensing* (Burke et al., 2006; Campbell et al., 2006). The key idea behind participatory sensing is to empower ordinary citizens to collect and share sensed data from their surrounding environments using their mobile phones.

Mobile phones, though not built specifically for sensing, can in fact readily function as sophisticated sensors. The cameras on mobile phones can be used as video and image sensors. The microphone on the mobile phone, when it is not used for voice conversations, can double up as an acoustic sensor. The embedded GPS receivers on the phone can provide location information. Other embedded sensors such as gyroscopes, accelerometers, and proximity sensors can collectively be used to estimate useful contextual information (e.g., if the is user walking or traveling on a bicycle). Further, additional sensors can be easily interfaced with the phone via Bluetooth or wired connections, e.g., air pollution or biometric sensors.

Participatory sensing[1] offers a number of advantages over traditional sensor networks which entails deploying a large number of static wireless sensor devices, particularly in urban areas. First, since participatory sensing leverages existing sensing (mobile phones) and communication (cellular or WiFi) infrastructure, the deployment costs are virtually zero. Second, the inherent mobility of the phone carriers provides unprecedented spatiotemporal coverage and also makes it possible to observe unpredictable events (which may be excluded by static deployments). Third, using mobile phones as sensors intrinsically affords economies of scale. Fourth, the widespread availability of software development tools for mobile phone platforms and established distribution

* Corresponding author. Tel.: + 49 6151 16 70923; fax: +49 6151 16 70921.
E-mail addresses: delphine.christin@seemoo.tu-darmstadt.de (D. Christin), andreas.reinhardt@kom.tu-darmstadt.de (A. Reinhardt), salilk@cse.unsw.edu.au (S.S. Kanhere), matthias.hollick@seemoo.tu-darmstadt.de (M. Hollick).

[1] Without loss of generality, we use the generic term *participatory sensing* to designate applications using mobile phones as sensors (or as data sink for interfaced sensors) where participants voluntarily contribute sensor data for their own benefit and/or the benefit of the community. The notion of participatory sensing therefore includes *mobiscopes* (Abdelzaher et al., 2007) and *opportunistic sensing* (Campbell et al., 2006), and also covers specific terminologies focusing on particular monitoring subjects, such as *urban sensing* (Campbell et al., 2006), *participatory urbanism* (Paulos et al., 2007), *citizen sensing* (Burke et al., 2006), *people-centric sensing* (Campbell et al., 2006, 2008), and *community sensing* (Krause et al., 2008).

0164-1212/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2011.06.073

channels in the form of App stores makes application development and deployment relatively easy. Finally, by including people in the sensing loop, it is now possible to design applications that can dramatically improve the day-to-day lives of individuals and communities.

A plethora of novel and exciting participatory sensing applications have emerged in recent years. CarTel (Hull et al., 2006) is a system that uses mobile phones carried in vehicles to collect information about traffic, quality of en-route WiFi access points, and potholes on the road. Micro-Blog (Gaonkar et al., 2008) is an architecture which allows users to share multimedia blogs enhanced with inputs from other physical sensors of the mobile phone. Other applications of participatory sensing include the collection and sharing of information about urban air (Paulos et al., 2007) and noise pollution (Rana et al., 2010), cyclist experiences (Eisenman et al., 2009), diets (Reddy et al., 2007), or consumer pricing information in offline markets (Dong et al., 2008).

A typical participatory sensing application operates in a centralized fashion, i.e., the sensor data collected by the phones of volunteers are reported (using wireless data communications) to a central server for processing, as illustrated in Fig. 1. The sensing tasks on the phones can be triggered manually, automatically, or based on the current context. On the server, the data are analyzed and made available in various forms, such as graphical representations or maps showing the sensing results at individual and/or community scale. Simultaneously, the results may be displayed locally on the carriers' mobile phones or accessed by the larger public through web-portals depending on the application needs.

Current participatory sensing applications are primarily focused on the collection of data on a large scale. Without any suitable protection mechanism however, the mobile phones are transformed into miniature spies, possibly revealing private information about their owners. Possible intrusions into a user's privacy include the recording of intimate discussions, taking photographs of private scenes, or tracing a user's path and monitoring the locations he has visited. Users are reluctant to contribute to the sensing campaigns, once they are aware of possible consequences. Since participatory sensing exclusively depends on user-provided data, a high number of participants is required. The users' reluctance to contribute would diminish the impact and relevance of sensing campaigns deployed at large scale, as well as limiting the benefits to the users. To encounter the risk that a user's privacy might be compromised, mechanisms to preserve user privacy are mandatory.

Within the scope of this manuscript, we analyze the current state-of-the-art in privacy-preserving mechanisms applied in participatory sensing campaigns. Besides describing the solutions currently applied to address user privacy, we also highlight and discuss open issues and their impact on privacy. Our contributions can be summarized as follows:

- We analyze existing participatory sensing applications to identify the different modalities of sensor data contributed by users. We also investigate the extent of personal information that can be inferred by examining the uploaded data either in isolation, or by combining different sensor modalities.
- We define the notion of privacy in participatory sensing and we highlight threats to privacy resulting from the disclosure of this data to untrusted parties.
- We examine how the current state-of-the-art protects the privacy of the participants by conducting a cross-analysis of the existing countermeasures and the architectural elements present in typical sensing applications.
- Based on this analysis, we identify and discuss future research directions that need to be addressed to prevent privacy from being the limiting factor for user participation in sensing campaigns.

We present these contributions as follows: In Section 2, we provide an overview of various participatory sensing applications with a particular focus on the different kinds of sensor data that are being collected. We examine the resulting implications on revealing personal information in Section 3. Subsequently, we analyze existing privacy-preserving countermeasures in Section 4. In Section 5, we highlight and discuss future research directions before concluding the paper in Section 6.

2. Participatory sensing applications and system model

The emergence of the participatory sensing paradigm has resulted in a broad range of novel sensing applications, which can be categorized as either *people-centric* or *environment-centric* sensing. People-centric applications mainly focus on documenting activities (e.g., sport experiences) and understanding the behavior (e.g., eating disorders) of individuals. In contrast, environment-centric sensing applications collect environmental parameters (e.g., air quality or noise pollution). As many of the applications make use of the same sensing modalities, we confine our discussion to a selection of representative applications and illustrate the varied usage models of participatory sensing in this article. After presenting an overview of more than 30 illustrative applications, we derive a general system model and examine the different sensing modalities of each application.

2.1. *People-centric sensing applications*

People-centric sensing uses the sensor devices integrated in mobile phones to collect data about the user. We discuss a representative selection of existing people-centric participatory sensing applications and analyze the sensing modalities used.

2.1.1. *Personal health monitoring*

In personal health monitoring, mobile phones are used to monitor the physiological state and health of patients/participants using embedded or external sensors (e.g., wearable accelerometers, or air pollution sensors). For example, *DietSense* (Reddy et al., 2007) assists participants who want to lose weight by documenting their dietary choices through images and sound samples. The mobile phones are worn on necklaces and automatically take images of the dishes in front of the users. The images document the participants' food selection and allow for an estimation of the food weight and waste on the plates. Moreover, the mobile phones capture the participants' context during their meals by recording time of day, location, and sound samples to infer potential relationships between the participants' behavior and their context (e.g., having lunch in a restaurant or eating chips late at night on the sofa). All captured data are uploaded to a personal repository, where the participants can review them to select/discard the information to be shared with their doctors and nutritionists.

A system to monitor pediatric obesity through multimodal activity detection is presented in Annavaram et al. (2008). The system is based on a heterogeneous wireless body-area network which employs sensors for heart frequency, acceleration, electrocardiography, blood oxygen saturation, and the user's galvanic skin response. All sensor data are tagged with location information, and optionally with audio and video tags. Adult obesity often results from an imbalance between calorie intake and calorie expenditure. The *BALANCE* system (Denning et al., 2009) combines an intuitive entry form for calorie intake with a body-area sensor which caters for activity classification. The system relies on the analysis of acceleration patterns to classify the participants' activities, e.g., sitting, running, walking, or bicycling. By correlating activity types and their corresponding durations, the users' calorie expenditures are estimated. Activity and calorie monitoring is also combined in the

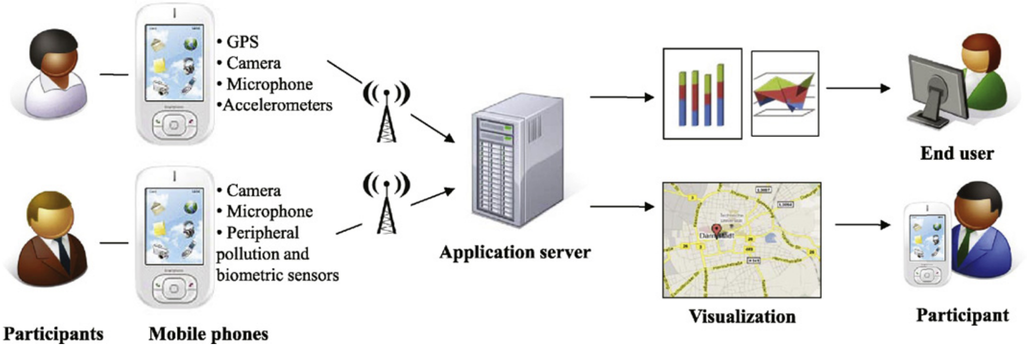

Fig. 1. Architectural overview of a typical participatory sensing application.

Jog Falls project (Nachman et al., 2010), which is based on the combination of body-area sensors (acceleration and heart rate) with a simple interface for entering calorie intake. Both functionalities are integrated within a mobile phone application. Complemented by separate blood pressure and weight measurements, participants and their nutritionists are notified of the overall achievements with regard to their diet and targeted weight loss.

The *HealthSense* project (Stuntebeck et al., 2008) targets the automated detection of health-related events that cannot be directly observed by current sensor technology, like tow conditions, pain, or depression. Acceleration data is being used in conjunction with machine learning approaches to detect correlations, e.g., diagnose itching as the source for a user's scratching motions. *MobAsthma* (Kanjo et al., 2009) monitors the asthma condition of the patients and their exposure to pollution. A peak flow meter and a pollution sensor are interfaced to the mobile phone via a Bluetooth connection, and measure both the volume of air inhaled and expired by the patients as well as the airborne particle concentration. These measurements are coupled with the patients' location and made available to allergists and asthma specialists to investigate the personal relationships between asthma attacks and exposure to air pollution. In addition to investigation purposes, MobAsthma can detect asthma attacks in early stages and autonomously alert remote medical staff.

Mobile phones can also be applied to remotely monitor the activity and posture of patients (e.g., elderly people living alone) using peripheral or the embedded accelerometers (Györbíró et al., 2009). For example, medical staff can gather the physical condition of elderly people by analyzing the temporal repartition of their postures among, e.g., sitting, standing, or lying. The granularity and accuracy of the activity recognition depend on the amount and position of the accelerometers (worn, e.g., on the hip, or in front/back pockets). Similarly to MobAsthma, medical staff may be directly alerted via the mobile phones in case of abnormal behavior or when users fall with *SenSay* (Siewiorek et al., 2003).

2.1.2. Calculating environmental impact and exposure to particles

PEIR (Personal Environmental Impact Report) is a system that allows users to use their mobile phone to determine their exposure to environmental pollutants (Mun et al., 2009). A sensing module installed on the phone determines the current location of the user as well as information about the currently used mode of transportation (e.g., bus vs. car), and transfers this information to a central server. In return, the server provides the users with information about the environmental impact of their traveling in terms of carbon and particle emissions. Additionally, the server estimates the participants' exposure to particle emissions generated by other vehicles and to fast food restaurants while commuting. The latter may be useful for health conscious users who may want to avoid the temptation of stopping by such restaurants. The mode of transport is inferred using accelerometer readings, while the route travelled is extracted from the captured location traces. Additional input parameters and models are considered for determining the environmental factors, such as weather conditions collected by weather stations, road traffic flow models, and vehicle emission models.

2.1.3. Monitoring and documenting sport experiences

The *BikeNet* (Eisenman et al., 2009, 2007), *Biketastic* (Shilton, 2009), and *SkiScape* (Eisenman and Campbell, 2006; Eisenman et al., 2006) projects monitor the sport experiences of the participants. Both BikeNet and Biketastic document the bicycling experiences of the participants. BikeNet draws a fine-grained portrait of the cyclist by measuring his current location, speed, burnt calories, and galvanic skin response. Multiple peripheral sensors are used to obtain this information: microphone, magnetometer, pedal speed sensor, inclinometer, lateral tilt, stress monitor, speedometer/odometer, and a sensor for CO_2 concentration. The peripheral sensors form a body area network and interact with the mobile phone over a wireless connection. In comparison, Biketastic concentrates on the road conditions, including the roughness of the road and the noise level along the road captured by on-board accelerometers and microphone, respectively. The captured data can be reviewed by the cyclists themselves, but can also be merged with other participants' data or combined with additional parameters, such as air quality and traffic properties, in order to construct complete maps for the cycling community.

In contrast, SkiScape focuses on winter sports and is deployed in ski resorts. Peripheral sensors (temperature sensor, accelerometers, and microphone) are attached either on skis or personal equipment to measure the body temperature, the maximum acceleration, and the travelled distance. Besides the sensors bound to the user, several static nodes are located along the trail and cater for the localization of the user within the ski resort. Using SkiScape, the participants can document their traces, locate their friends, and select lifts depending on the queue length. Simultaneously, the manager can optimize the maintenance operations, and emergency staff can easily localize participants in case of accidents.

2.1.4. Enhancing social media

A large pool of applications utilizes data captured by sensors to enrich the contents shared in social media, such as blogs, social networks, or virtual worlds. *Micro-Blog* (Gaonkar et al., 2008) proposes to build a "virtual information telescope", which provides a high-resolution view of the world by leveraging the mobile phones serving as lenses. The participants can create geotagged blog entries and enhance them with multimedia information (e.g., audio records, pictures, accelerometer data, or WiFi coverage) captured via their mobile phones. The created entries are then uploaded to a server, which may position them at their capture location on a global map accessible by the public. Users can browse the entries displayed on the map to find information about particular points of interests (e.g., audio reviews about restaurants, pictures of the nearest beaches). If the information required is not contained in existing entries, the users can send queries (e.g., "how is the WiFi coverage near this beach?") to the server, which relays them to mobile phones currently located in this area.

Similarly, *CenceMe* (Miluzzo et al., 2008; Musolesi et al., 2008) integrates virtual representations of the participants' current state and context in social networks and virtual worlds. Based on multimodal information (acceleration, audio samples, pictures, neighboring devices, and location) captured by the mobile phone, context information is inferred in various dimensions, including the user's mood, location, and habits, as well as information about the currently performed activity and the environment. The inferred information is then posted as status message in social networks or translated into the virtual representation of participants in virtual worlds.

2.1.5. Price auditing

LiveCompare (Deng and Cox, 2009) and *PetrolWatch* (Dong et al., 2008) facilitate price comparisons of grocery products and fuel at different locations. Instead of manually reporting the prices, the participants use their mobile phones to take pictures of the displayed prices. In LiveCompare, the participants only need to take pictures of a product's price tag and its barcode. The barcode is decoded into a textual representation on the mobile phone, and transferred to the server along with capture time, location information, and the picture displaying the current price. To compare prices, participants can search for products in the application, which then retrieves the corresponding price reports, selects the stores in proximity of the participant's current location and displays the pictures of the corresponding price tags.

In comparison, the price collection process is not only simplified in PetrolWatch, but also automated. Each mobile phone is mounted on the passenger seat of a car and faces the road to automatically photograph fuel price boards (using GPS and GIS) when the vehicle approaches service stations. The pictures are then uploaded to a central entity, which is responsible for image processing and price extraction. The brand of the service station is first inferred from the capture location in order to reduce the image processing complexity, as price boards of different brands differ in colors and dimensions. Assisted by this information, computer vision algorithms extract the fuel prices, and uploads them to the database. Users can query the system to determine the cheapest fuel that is available in their area of interest.

2.2. Environment-centric sensing applications

In environment-centric scenarios, the mobile phones capture information via their embedded sensors and additional peripheral sensors about the surroundings of the participants. In contrast to most people-centric sensing scenarios, the captured data are mainly exploited at a community scale, e.g., to monitor the evolution of environmental parameters like air quality, thermal columns, noise, road and traffic conditions in cities, or to detect socially interesting events.

2.2.1. Air quality monitoring

In *Haze Watch* (Carrapetta et al., 2010), mobile phones were interfaced to external pollution sensors, in order to measure the concentration of carbon monoxide, ozone, sulphur dioxide, and nitrogen dioxide concentration in the air. In comparison to meteorological stations, the mobile phones may collect less accurate measurements. However, their inherent mobility allows them to observe unpredictable events (e.g., accidental pollution), which can seldom be detected by static stations and provide large spatial coverage. The mobile phones can thus complement static high-fidelity data captured by traditional meteorological stations by providing finer-grained readings. In addition to pollution measurements, the mobile phones can capture temperature and wind speed, such as shown in *PollutionSpy* (Kanjo et al., 2009) and by Paulos et al. (2007). The timestamped and geotagged measurements are then uploaded to a server to build maps, which aggregate the readings of all participants and are accessible by the public. Individual measurements may also be displayed on the participant's mobile phone. Despite a common interest in air pollution, this application scenario and the related applications differ from the PEIR project presented in Section 2.1.2, as the pollutant concentrations are actually sensed with real sensors carried by the participants and not inferred based on weather conditions, traffic condition, and emission models.

2.2.2. Monitoring thermal columns

Ikarus is a participatory sensing application for paraglider pilots (von Kaenel et al., 2011). It collects information about thermal columns, which are used by pilots to gain in altitude during their flights. Ikarus is based on the collection of barometric pressure information, annotated with the time and location of sampling. Thermal maps are then extracted from the contributed sensor data, and processed for visualization on web-based maps and distribution on other pilots' navigation units.

2.2.3. Monitoring noise and ambiance

Microphones in mobile phones can be configured to measure the surrounding noise level and give insights about the nature of contextual events. In *NoiseTube* (Maisonneuve et al., 2009), *Ear-Phone* (Rana et al., 2010), and *NoiseSpy* (Kanjo et al., 2009), noise levels are used to monitor noise pollution, which can, e.g., affect human hearing and behavior. The data are then used to build representative pollution maps to enable specialists to understand the relationships between noise exposition and behavioral problems.

In addition to noise level, the sound samples can be further analyzed to determine, e.g., whether human voices were recorded in order to recognize the sound context in *SoundSense* (Lu et al., 2009). Depending on the recognized context, the corresponding sound samples may be used to document audio diaries or indicate places where music is currently played to other participants in online social networks.

Furthermore, in *MoVi* (Bao and Roy Choudhury, 2010), the mobile phones collectively sense the surrounding ambiance to detect precursor signs (e.g., outburst of laughter, moves in the same directions) of relevant social events (e.g., speeches) and trigger a video recording of the upcoming events in case of positive detection. The recordings collected by different mobile phones can then be timely and automatically assembled in a common video clip, which may potentially reduce cumbersome manual video edition. As the video recordings are automatically triggered, the participants can focus on the events instead of having to focus on taking recordings.

MetroTrack (Ahn et al., 2010) is used to track mobile noise sources in outdoor environments. Similar to the aforementioned approaches, mobile phones are carried by participants are being used as mobile noise sensors. Collaboration between neighboring mobile phones is used to estimate the future trajectory of a noise source through the application of distributed Kalman filtering. The tracking task is automatically forwarded to nodes in proximity to the estimated trajectory to ensure accurate detection of the mobile noise source.

2.2.4. Monitoring road and traffic conditions

The mobile phones can be exploited to document road and traffic conditions. In *Nericell* (Mohan et al., 2008), the embedded accelerometer, microphone, and positioning system (GPS or GSM radio) are used to detect and localize traffic conditions and road conditions, e.g., potholes, bumps, or braking and honking (which are both implicit indicators of traffic congestion). The application integrates the provided information about the surface roughness of the roads, the surrounding noise, and the traffic conditions into traffic maps, which are available to the public. In addition to parameters related to a cyclist's activity and his physical condition, the BikeNet project (cf. Section 2.1.3) measures environmental parameters such as pollution, noise levels and irregularity of the roads. Hence, this project can be regarded as a hybrid application, combining components from both people-centric and environment-centric sensing.

Current solutions for automotive traffic monitoring, such as inductive loops, are usually expensive in deployment and maintenance costs, and often prone to errors. The concept of *Virtual Trip Lines* (Hoh et al., 2008) targets to replace these conventional monitoring solutions by smartphones mounted within vehicles. Instead of deploying costly monitoring infrastructure, the locations of street segments of interest are modeled as virtual trip lines and forwarded to the participants' smartphones. A phone application constantly monitors its current location, and transmits its position and travelling speed to the traffic monitoring infrastructure whenever a virtual trip line has been crossed. A similar approach for traffic monitoring is based on the use of multiple positioning sensors (GPS, WiFi, and cellular radios) in Thiagarajan et al. (2009). The *VTrack* system relies on mobile devices which deliver timestamped location estimates to a server, in order to estimate driving times on different road segments at a fine spatio-temporal granularity. Based on the availability of traffic information, roads with unusually high travel times are identified, and real-time information about these traffic hotspots is then used for route planning. An extension to the approach is presented in Thiagarajan et al. (2010), where the concept of *cooperative transit tracking* is presented. The lack of public information about the timeliness of public transportation is encountered by a participatory approach to report the current location of transit vehicles by the users on board. A background task on smartphones automatically detects if the user has entered a transit vehicle, both above ground and underground, and uploads its current location coordinates to the tracking server.

Although the *CarTel* (Hull et al., 2006) and *GreenGPS* (Ganti et al., 2010) projects utilize dedicated sensing devices with more resources than current mobile phones, both rely on contributions by participants and can thus be considered as implementations of participatory sensing. GreenGPS targets to provide a map of the least fuel-consuming routes to drivers. The fuel consumption is measured via specific sensors accessing the gauges and instrumentation of the vehicles, and is correlated with location information from an external positioning system. The sensor readings are stored on a memory card and are manually uploaded to the application by the participants themselves. The obtained results have shown that high speed and long distance routes do not necessarily reduce the fuel consumption. Using an embedded computer coupled with sensors, CarTel analyzes the time it takes a participant to commute to work, determines traffic congestion, and represents jammed roads on a map. Additionally, driving patterns and readings from automotive on-board diagnosis systems can be taken into account.

2.3. *System model*

From the analysis of the previous people-centric and environment-centric sensing applications, we derive the following general system model including stakeholders and architectural components. Fig. 2 summarizes the resulting model and the interactions between its elements.

2.3.1. *Stakeholders*

In the above applications, we identify the following stakeholders:

- *Campaign administrators*: They are members of organization, research groups or individuals who initiate the participatory sensing campaigns. They design, implement, and deploy the system architecture and are responsible for the maintenance and the management of the infrastructures. For example, this includes making available the application for download on the campaigns' websites or in App Stores and setting up the application server to collect and process the data.
- *Participants*: They install the sensing application on their mobile devices and voluntarily contribute to the participatory sensing campaigns by gathering sensor readings using the mobile phones they own and carry. Besides being driven by the motivation to benefit from the data provided by themselves and other participants, their contributions to the campaigns can be motivated by different factors primarily influenced by the nature of the campaign. At an individual scale, they may be willing to, e.g., improve their health conditions, monitor their impact on the environment, or document their sport experiences. While at a community scale, they may aim at monitoring pollution and thermal columns to help scientists to understand the monitored phenomena, or helping other users by providing information about road and traffic conditions. Consequently, their motivations can be either self-centered (e.g., in Jog Falls), altruistic (such as in PollutionSpy), or a combination of both (e.g., in Ikarus). Note that the degree of involvement of the participants in the sensing process depends on the application characteristics, as discussed in detail in Section 2.3.2.
- *End users*: They access and consult the data gathered by the participants according to their interests and preferences. End users include, e.g., contributing participants willing to consult their own collected data, campaign administrators verifying the actual contributions and results, specialized scientists attempting to gain insights about the monitored phenomena, participants' relatives consulting the last reports to encourage the concerned participants, health professionals checking patient data, or the general public.

2.3.2. *Architecture*

In Fig. 2, we identify typical architectural components common to the existing participatory sensing architectures and determine their function in the general system model. The components are generally organized into a client–server architecture and interact from the sensing process to the presentation of the results to the end users.

- *Sensing component*: It is located on the participants' mobile phones and captures different kinds of sensor data, prevalently time, location, pictures, sound samples, accelerometer data, pollution data, biometric data, and barometric pressure. We summarize the sensing modalities used in the presented applications in Table 1.

  The sensor data can be captured according to one of the following sensing modes: manual, automatic, and context-aware (Estrin, 2010). In the manual mode, the participants trigger the collection of sensor readings themselves when they detect relevant events, such as noise pollution or traffic congestion. This mode is also referred as *participatory sensing* in the literature (Burke et al., 2006), as the participants directly participate in the sensing process. On the contrary, the participants are not directly involved in the automatic and context-aware modes. In the automatic mode (also known as continuous mode (Eisenman et al.,

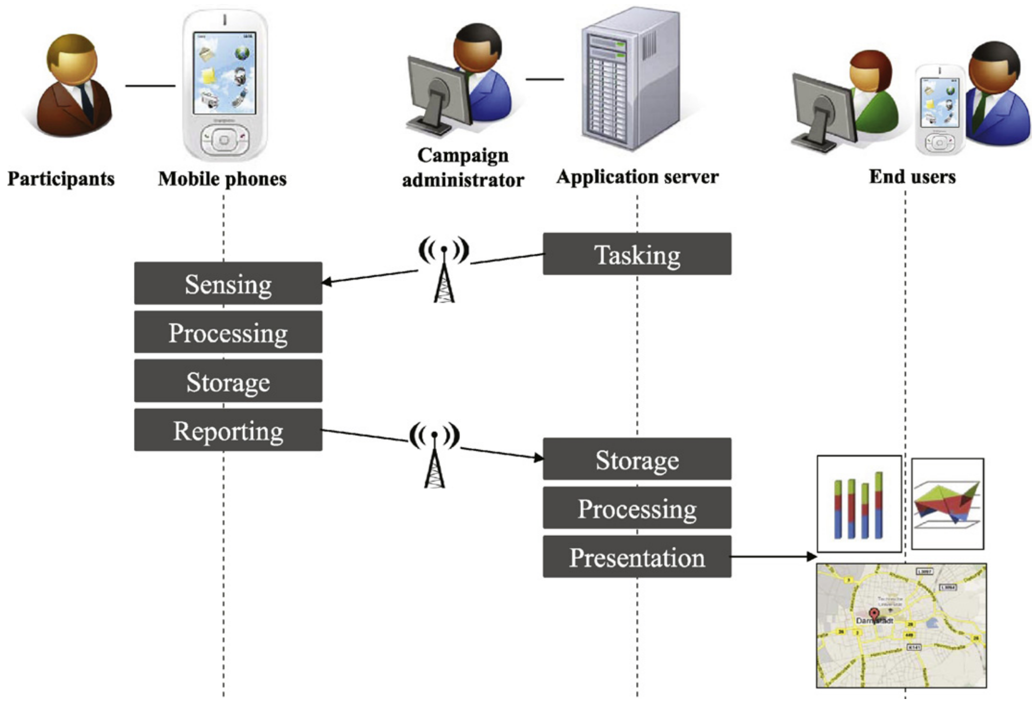

Fig. 2. Stakeholders and architectural components of participatory sensing applications.

2009)), the sensor readings are collected at a constant sampling frequency, while their collection depends on the surrounding environment in the context-aware mode (also known as *opportunistic sensing* (Campbell et al., 2006)), where the embedded sensors monitor their environment and activate the sensing function when previously set thresholds are exceeded.

- The *tasking component* supports the above sensing component by distributing the sensing tasks to the mobile phones. These tasks specify the sensing modalities based on the application requirements including criteria to fulfill to start the capture, sensors used and sampling frequency. One example of such requirements could be phones equipped with GPS and with embedded cameras that can capture 3 megapixel images. The tasks also contain information about location and/or time frame of interest.
- The *reporting component* ensures the transmission of the sensor readings collected by the sensing component to the application server. The data transfers mostly make use of communication infrastructure available to the mobile phone, such as Wireless LAN, or GSM/GPRS/3G connectivity. For example, the sensor readings can be transmitted to the server using SMS, TCP connections (Gaonkar et al., 2008), or remote procedure calls (Miluzzo et al., 2008).
- The *storage component* ensures the storage of the collected data on the mobile phone and the reported data on the server. While the server manages long-term storage of the reported data, the mobile phone ensures short-term storage of the data to be processed or transmitted to the server. On the server side, the data are commonly stored in relational databases (Gaonkar et al., 2008; Hull et al., 2006; Kansal et al., 2007) or databases specially adapted to the management of sensor readings, e.g., sensedDB (Grosky et al., 2007) and SensorBase (Chang et al., 2006).
- The *processing component* extracts features of interest from the sensor readings either directly on the mobile phones at individual scale or on the server side at larger scale. The component analyzes the data reported to the server and prepares them for the presentation component.
- The *presentation component* presents the results obtained by the processing components to the end users. The results are either locally display on the mobile phones for the participants only or presented through web-portals to a larger public. The results are presented in forms of raw data to allow the end users to analyze them themselves, or in forms of graphs, maps, and geographic overlays (Hull et al., 2006).

3. Privacy and threats in participatory sensing

As we have illustrated in Table 1, virtually all participatory sensing systems collect sensor readings related to the participants and/or their environments. Obviously, the collected data may be used to extract or infer sensitive information about a user's "private life, habits, act and relations"—the basic definition of *privacy* by Brandeis and Warren (1890). Simultaneously, contributed sensor data are vital to any participatory sensing application, and their deficiency endangers the success of participatory sensing systems. Campaign administrators therefore need to increase the user awareness of the consequences of the disclosure of sensor data as well as provide solutions to maintain user privacy in order to ensure the durability of the campaign and prevent participants from opting out.

In this section, we thus discuss the notion of privacy in participatory sensing systems. We first address the lack of a common understanding of privacy (Newell, 1995) by introducing our definition of privacy, specially tailored to the specifics of participatory sensing systems. We then conduct a privacy analysis to determine actors and processes that represent threats to the privacy of the participants, before highlighting possible consequences resulting from the disclosure of sensitive information.

3.1. Privacy definition

With the rise of communication equipment and computing systems, the notion of *information privacy* has emerged, which has been initially defined as "the claim of individuals [. . .] to determine

Table 1
Comparison of applications and sensing modalities.

Type	Monitored subject	Application	Time	Location	Pictures	Sound	Acceleration	Pollution	Biometric data	Barometric data
People-centric sensing	User health	DietSense	x	x	x	x				
		Pediatric obesity	x	x	x		x		x	
		BALANCE	x				x		x	x
		Jog Falls	x		x		x		x	
		HealthSense	x				x			
		MobAsthma	x	x				x	x	
		SenSay	x	x			x			
	Personal impact	PEIR	x	x						
	Sport experiences	BikeNet	x	x		x	x	x	x	
		Biketastic	x	x		x	x			
		SkiScape	x	x		x	x			
	Social media	Micro-Blog	x	x	x	x	x		x	
		CenceMe	x	x	x	x	x			
	Price auditing	LiveCompare	x	x	x					
		PetrolWatch	x	x						
Environment-centric sensing	Air pollution	Haze Watch	x	x				x		
		PollutionSpy	x	x				x		
		Paulos et al.	x	x				x		
	Thermal columns	Ikarus	x	x		x				x
	Noise and ambiance	NoiseTube	x	x		x				
		Ear-Phone	x	x		x				
		NoiseSpy	x	x		x				
		SoundSense	x	x		x				
		MoVi	x	x		x	x			
		MetroTrack	x	x		x				
	Road conditions	Nericell	x	x		x	x			
		Virtual Trip Lines	x	x						
		VTrack	x	x						
		Transit tracking	x	x			x			
		CarTel	x	x	x		x	x		
		GreenGPS	x	x						

for themselves when, how and to what extent information about them is communicated to others" by Westin (1967). Although the field of privacy is multifaceted and comprises several other dimensions (Renaud and Gálvez-Cruz, 2010), user-level control over sensitive sensor data represents the major concern in participatory sensing systems. In consequence, we refer to information privacy whenever the term privacy is being used in this article.

In order to cater for the specific characteristics of participatory sensing systems, we propose an adapted version of the above definition of information privacy, which we apply throughout the remainder of this article:

> Privacy in participatory sensing is the guarantee that participants maintain control over the release of their sensitive information. This includes the protection of information that can be inferred from both the sensor readings themselves as well as from the interaction of the users with the participatory sensing system.

Our definition implies that participants continuously need to control the release of their sensor readings to third parties (including the participatory sensing application) and have full control over the provided type of sensor readings, the degree of granularity, the spatiotemporal context, and the data recipients. Participants are therefore actively involved in pursuing the protection of their privacy. However, their actions should ideally be supported by the system with usable and understandable mechanisms (Christin, 2010). In addition to the direct protection of the sensor readings, the proposed definition includes the protection of the participants against the inference of sensitive information resulting from their interactions with the participatory sensing system. This implies that potential adversaries are unable to determine the spatiotemporal information about the participants, when they download tasks or report data to the application, or link their real identity with the pseudonyms they use. From this perspective, the participants are not actively involved in the privacy decisions, but the participatory sensing system is responsible for their privacy protection.

3.2. Privacy analysis

To further understand the privacy concerns in participatory sensing, we analyze participatory systems from a privacy perspective. We base our analysis on the theory of *contextual integrity* (Nissenbaum, 2004), which comprises the dimensions of *appropriateness* and *distribution*. Appropriateness defines if the revelation of a particular piece of information is appropriate in a given context, while distribution focuses on the occurrence of an information transfer from one party to another. The concept of *contextual integrity* defines breaches of appropriateness or distribution as violations to a user's privacy.

Socio-cultural and contextual differences have a strong impact on the individual perception of data sensitivity. For example, users make different privacy decisions depending on the number of recipients of their data (Tang et al., 2010). While it has been shown that users pragmatically determine whether they share their locations with a single person, additional parameters like their willingness to attract attention or boost their self-presentation are taken into account when they decide about sharing their whereabouts with a larger group of people. In consequence, the appropriateness entirely depends on the individual privacy conception of each participant. We thus discuss this dimension at a high level and confine our privacy analysis to a thorough examination of distribution aspects for the stakeholders present in typical participatory sensing applications. In all cases, we assume that the analyzed participatory sensing systems collect sensor readings including pictures, sound samples, acceleration, pollution data, biometric data, and barometric pressure (see Table 1), which are tagged with spatiotemporal information.

- *Who gathers the sensor readings?* The sensor readings initially collected by the participants are reported to the application server, which is run by the campaign administrators. The administrators therefore have a direct access and control over the collected sensor readings. While the distribution of the sensor readings to the campaign administrators is part of participatory sensing systems and the participants are aware of this fact, the reported sensor readings may reveal sensitive information about the participants, if no privacy-preserving processing is locally applied on the mobile phones. The participants are the only ones able to judge the appropriateness of the revealed information, as it depends on their personal privacy conception as well as the nature and the context of the revealed information. However, the participants may encounter difficulty in translating their own conception of privacy into, e.g. privacy settings, and to understand the implications of their settings and actions on their privacy (Debatin et al., 2009). For example, users are often unaware of the technical details of the underlying architecture. They may underestimate the risks related to their privacy (Acquisti and Grossklags, 2005), and pay little attention to policies and end user licensing agreement, both of which are often laid out in technical terms and thus hard to understand (Good et al., 2005). In summary, a combination of the inconsistent behavior of the participants and the lack of existing solutions to clarify privacy implications may more often than not lead to sub-optimal privacy protection. To improve on the current state-of-the-art, we propose further research in this direction in Section 5.1.
- *Who analyzes the sensor readings?* The sensor readings are analyzed after having been reported to the servers. The campaign administrators determine and implement the processing to apply on the sensor readings to prepare them for analysis. Particular privacy-aware functions can be applied to anonymize the participants and/or remove sensitive information about them before their release to the analysts. However, the sensor readings can still contain private information about the participants. After this preparation processing, the sensor readings can be analyzed by different groups of people depending on the application scenario. These include the participants themselves, the campaign administrators, doctors, researchers in the field, etc. The participants maintain different relationships with the groups of analysts that directly impact the appropriateness. For example, it is considered as appropriate for participants to share personal information related to their health conditions with their doctors (Nissenbaum, 2004), whereas sharing the same information with a different analyst, such as a campaign administrator, may be inappropriate. By reporting their sensor readings to the application, the participants are aware that they can be analyzed by the application (i.e., the campaign administrators). However, they may ignore that the distribution of their sensor readings can be extended to external analysts potentially unknown to them, endangering the appropriateness.
- *Who accesses the analyzed sensor readings?* The analyzed sensor readings are released by the campaign administrators to the end users. The access to the sensor readings is determined by access control rules defined by either the participants themselves or the campaign type. The authorized persons can be restricted to only the participants themselves or extended to their relatives, friends, or a larger public, depending on the application. In the worst case, the circle of authorized persons is thus enlarged from the campaign administrators and related analysts to the general public, raising the issue of appropriateness to its climax. The level of appropriateness can be moderated by proposing different degree of granularities at which the data are released depending on the nature of the relationships between the participants and the end users. However, the distribution of the sensor readings is ensured by the campaign administrators. The participants must thus trust them not to disclose sensitive information about themselves to untrusted parties.

In summary, the respect of the privacy of the participants therefore primarily depends on the campaign administrators who have direct access to the reported sensor readings and ensure their distribution to potential analysts and end users. Malicious campaign administrators or inefficient mechanisms to both remove private information from the sensor readings and control their access can thus contribute to the violation of the privacy of the participants.

3.3. Privacy threats

Let us assume that a worst case scenario where campaign administrators break the trusted relationship to the participants and reveal sensitive information about them. We examine the potential social consequences of such disclosure by successively considering the sensing modalities listed in Table 1:

- *Time and location*: It is evident by examining the table that virtually all applications (except for Jog Falls, HealthSense, and BALANCE) collect time and location information independently of their people-centric or environmental-centric nature, thus underpinning the importance of these two contextual factors. GPS receivers embedded in most current smartphones provide very accurate location coordinates. However, in the absence of GPS (due to lack of coverage or if the user does not want to reveal fine-grained location information), WiFi or cellular network based triangulation can be used to obtain coarse-grained location information (LaMarca et al., 2005). Contextual information collected from other embedded sensors (such as points of interest Azizyan et al., 2009, light, and noise) can also be used to identify a person's location.

  Given their importance, the disclosure of data from these two modalities has been shown to leak privacy-sensitive information about the participants, including their home and workplace locations, as well as their routines and habits (Shilton, 2009). For example, frequent visits to hospitals may allow employers to infer the medical condition of their employees, and similarly, attendance at political events may provide information about the political views of users (Liu, 2007). In summary, without any protection mechanism, the disclosure of location information may lead to severe consequences ranging from social to safety and security threats (Shilton, 2009). Additionally, the threats resulting from location/time traces are not confined to applications where authentication is required. Even in the case of anonymous contributions, location traces may be analyzed to infer the identity of the participants based on their residence location and reverse white page lookups (Mun et al., 2009).
- *Sound samples*: Besides inferring identities and preferences from spatiotemporal data only, the portrait of the user can be refined by complementing this data by samples of other sensing modalities. In several of the aforementioned applications, sound samples are either recorded intentionally by the participants, or captured automatically by the mobile phones. While participants can easily preserve their privacy by only recording non-sensitive events in the former case, mobile phones effectively behave as smart spies

in the case of automated recordings. Dedicated user interaction is required to prevent the applications from recording private conversations about intimate or confidential subjects. Even in public locations, the recognition of characteristic sound patterns that are unique to certain events and locations may allow adversaries to determine a participant's current context.

- *Pictures and videos*: The content of contributed pictures and recorded videos is also likely to reveal personal information about the participants and their environment. Although DietSense (Reddy et al., 2007) targets to take photos of consumed meals, no countermeasures are taken to conceal the faces of persons sharing their meal with the participants. In all scenarios, in which the camera is oriented away from the participant, faces of other people in the vicinity are possibly captured in the images, and thus conclusions about the number and identity of the participant's social relations can be drawn. The publication of captured pictures may lead to similar consequences as in online social networks, such as Facebook, where a teacher was suspended due to a picture showing her holding glasses filled with alcohol (CBS News, 2011), or a depressed woman who lost benefits from her health insurance for pictures showing her attending parties and relaxing on the beach (CBC News, 2009). Similar to sound recordings, the current user context and the surrounding environment may also be extracted from sensor data. For example, pictures showing points of interest may easily establish the participant's presence at those locations.
- *Acceleration*: Raw accelerometer readings may appear less threatening in revealing private information about the participants. However, this hypothesis not always true and may often only serve as a false sense of security. For example, if the mobile phone is carried on the hip, information about the gait, and thus possible indications about a user's identity, may be inferred (Derawi et al., 2010). Additionally, the research field of activity recognition also makes extensive use of accelerometer readings (Györbíró et al., 2009). The exploitation of these data by malicious users may have negative consequences. For example, employers may want to verify that their employees are actually working during their working hours. If the employers detect anomalies, they might suspend the respective employees.
- *Environmental data*: Recording particles and gas concentrations or barometric pressure may not directly threaten the privacy of the participants by themselves. However, particular air compositions combined with secondary information, such as precise air temperature, might identify the location of the participants at a level of granularity as fine as room levels within buildings, where location information can be inaccurate due to non-availability of GPS or other location services.
- *Biometric data*: Biometric sensor data can be used for a diagnosis of a user's current physiological state. Similarly to medical staff, adversaries may identify health anomalies or diseases based on the captured sensor data. Leaked medical information may then be used by health insurance companies or employers to revoke contracts, if an impairment of the physiological conditions of the participants is identified.

Privacy threats represent an inherent problem of any participatory sensing application. Although the subjects of interests of environment-centric applications are not the participants themselves, all considered applications monitor the spatiotemporal context of the participants and therefore represent a danger to their privacy. Furthermore, additional captured sensing modalities may provide further insights about the participants. As a result, environment-centric applications can similarly endanger the privacy of the participants, even if the threats are less perceptible at first sight than in the case of people-centric applications.

Location privacy has been predominately addressed in the literature in comparison with the other sensing modalities. The particular interest for location privacy may be due to the sharing of similar concerns with orthogonal domains addressing this issue, such as vehicular networks (Raya and Hubaux, 2007), location-based services (Myles et al., 2003), pervasive computing (Beresford and Stajano, 2003), ubiquitous computing (Beresford, 2005), etc. In contrast, the remaining sensing modalities are less represented in the literature, as they are one of the particularities of participatory sensing applications. However, we have shown in this section that their combinations may leak sensitive information about the participants and people in their vicinity, or provide information about their locations, even if those are protected by privacy-preserving mechanisms. In comparison with other application domains, addressing the privacy threats in participatory sensing requires thus to solve a multi-dimensional problem, as opposed to location privacy only. In the current applications, these threats are partially addressed by means of access control mechanisms. For example, the DietSense system ensures that pictures tagged with location and time are only shared with trusted parties, specifically the nutritionist (Reddy et al., 2007). Although the identities of user and nutritionist are mutually known, possible threats to the user privacy still comprise the recognition of faces or items in the pictures. A generic and sufficiently fine-grained solution to balance the extent and detail of data revealed to the target audience still remains to be found.

4. Countermeasures to privacy threats

After having outlined the privacy threats and the corresponding need for privacy-protecting mechanisms to encourage user participation, we examine the current state-of-the-art solutions which attempt to address these threats. We conduct a cross-analysis of the architectural elements present in typical sensing applications and present existing countermeasures, as illustrated in Fig. 3.

The sensed data passes through several stages between its collection and the consumption by the target audience. By tracing the distribution path of sensor data, we outline how privacy can be maintained and improved at each step. First we address the implications on privacy as outlined in the preceding section by discussing tailored sensing and anonymous task distribution. We then consider different schemes to anonymize and protect the privacy of the users while the data are reported to the application server. Subsequently, we detail solutions for privacy-aware data processing and storage including mechanisms to review, delete, and control the retention of sensed data. We conclude the discussion of countermeasures by presenting current solutions to control and audit their access.

4.1. Tailored sensing and user preferences

A first measure to encounter the privacy threats discussed in Section 3 is to control the data collection process at the user level and allow the participants to express their privacy preferences. In the presented scenarios, this control is applied to different extents. Although some solutions allow the users to fully disable the sensing function (Miluzzo et al., 2008; Shilton et al., 2008), doing so is of little use for participatory sensing, since the user would be unable to contribute any data. As proposed in Das et al. (2010), this binary scheme (full access to sensor data, or none at all) can be extended by introducing additional intermediate levels. For example, the participants may decide to selectively enable sensor measurements depending on a variety of factors, such as presence in sensitive locations (home or office), or their current social surroundings (presence of friends or family members). The selection

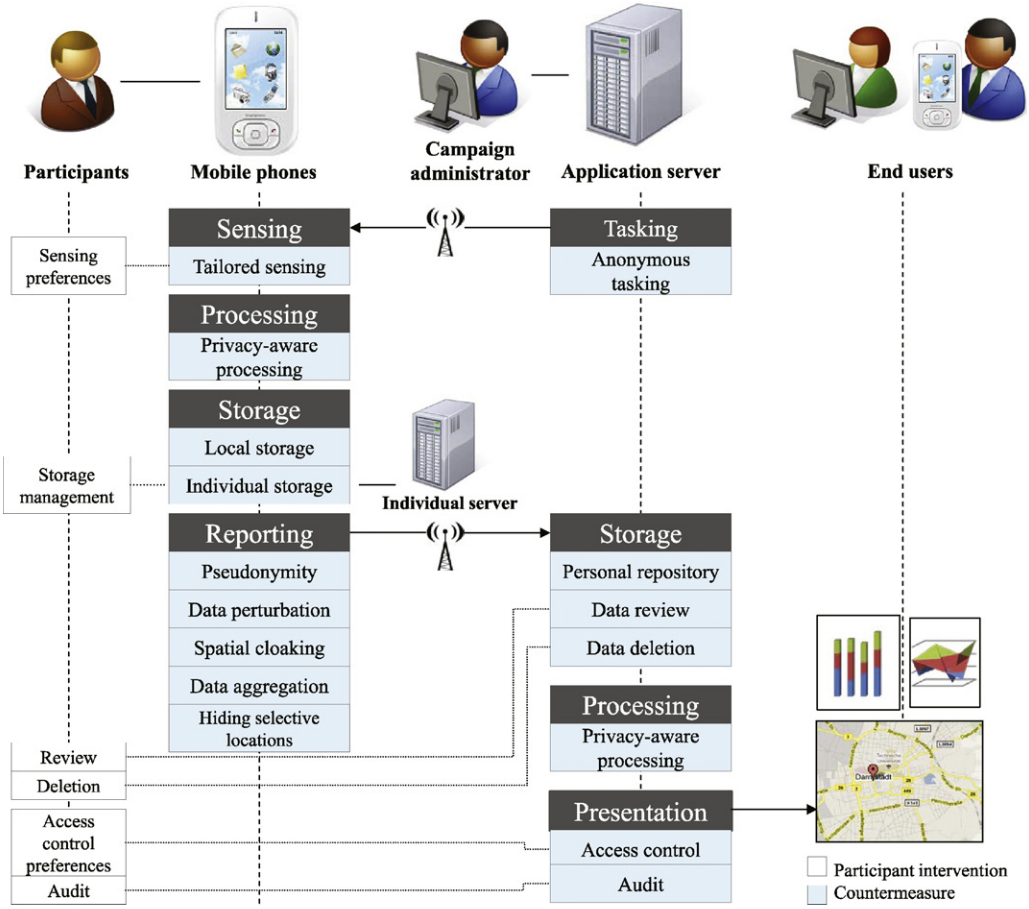

Fig. 3. Countermeasures and their relationships to the architectural components.

of theses factors allows the participants to explicitly indicate the type of information that they are happy with being collected in different contexts. They thus define their personal conception of appropriateness as defined in Section 3.2.

To reflect the user's privacy preferences whilst optimizing data fidelity, a finer-granular scheme allows users to adjust the sensing granularity and the time resolution. For example, samples may be collected every hour instead of every 15 s, or location information be captured at different degrees of granularity (Schilit et al., 2003). Table 2 illustrates possible degrees of granularity and related examples which may be applied in the context of some of the applications discussed in Section 2. Starting with the finest granularity on top, i.e., the unaltered original/raw data, the data resolution decreases towards the bottom of the table until reaching the coarsest meaningful degree of granularity. Depending on the application characteristics, the granularity of the different sensing modalities can be tuned in different ways, with the ones presented in the table only serving as illustrative examples.

Although this scheme does not provide a generic trade-off between privacy and disclosed data, it improves the overall acceptance of the solution by offering additional resolutions of granularity to the users. From the perspective of the application, coarse-grained data is still better than no data. For example, the location of air pollution measurements can be released at the granularity of a district. It should also be noted that sensing granularity has a direct impact on the energy consumption of the mobile phone. Whenever a sensor is activated for collecting data, it consumes the phone battery. As such, sampling at a coarser granularity implies that the sensors may be turned off for longer periods of time, thus resulting in energy savings. Simultaneously, both the size and quantity of the data to be transmitted are reduced. Transmitting data also consumes energy for the radio (3G or WiFi) transmission, thus less data volume also results in additional energy savings. However, additional processing may be required to filter the data and report them with a coarser granularity (e.g., recognizing the presence of faces on the images to eliminate them) leading to supplementary energy consumption, which needs to be regarded specifically.

4.2. *Anonymous task distribution*

Sensor data collection is generally triggered through *tasks*, which specify the sensing modalities (e.g., regions of interest, criteria to fulfill to start the capture, sensors used and sampling frequency) based on the application requirements. The tasks are distributed to the mobile phones that satisfy the tasking requirements.

The tasks are either statically deployed on the phones or initially downloaded from the application server. A central *tasking component* located on the application server or a dedicated *tasking server* (Kansal et al., 2007) selects appropriate devices based on predetermined criteria to optimize the sensing process, such as the current location of the mobile phone or its available resources (embedded sensors, battery lifetime, currently executed tasks, or processing capabilities) (Reddy et al., 2009). Once the devices are selected in a centralized task distribution model, the application server deploys the sensing tasks on the devices, e.g., via a push model based on executable binaries (Das et al., 2010).

Table 2
Example of granularity degrees for different sensing modalities.

Granularity	Location	Sound	Photo	Acceleration
Fine-grained	Precise position	Original sample	Original image	Raw data
	Street name	Voices removed	Faces blurred	Activity type
	District	Spectral properties	Number of people	Activity category
Coarse	City	Loudness level	Environment (indoor/outdoor)	Motion (yes/no)

In contrast to centralized tasking, mobile phones autonomously manage and distribute tasks in case of decentralized task distribution, such as proposed in Eisenman et al. (2008). The distributed character of decentralized task allocation allows it to transfer the sensing responsibilities to devices in its proximity. For example, a mobile phone *A*, which is not equipped with the sensors required by the tasks, transfers its tasks to other mobile phones embedding such sensors in the vicinity. *A* initiates the task transfer by broadcasting requests including the task to execute and the corresponding sensors to mobile phones *B* and *C* in its proximity. At the reception of these requests, *B* and *C* verify if they can fulfill the included task conditions. Assuming that only *B* possesses the required sensors, *A* transfers its sensing task to it. *B* can fulfill the transferred task either instantly if both mobile phones are located in the task's region of interest or remotely when *B* enters the region of interest.

The Bubble-sensing model (Lu et al., 2008, 2010) is a hybrid alternative to the purely centralized and decentralized schemes. Although mainly based on decentralized distribution of tasks, the concept also requires the presence of a central entity to maintain the sensing tasks. Participants, designated as *bubble creators*, can create persistent sensing areas at defined places of interests. They initiate a sensing request including the geographical region, duration, and sensing modality corresponding to each bubble and broadcast it to potential *bubble carriers* (i.e., other participants) in a distributed fashion. Bubble carriers may then move to the specified location, perform the sensing, and report the collected data to the central entity, the *bubble server*, from where the data can be retrieved by the bubble creator. The persistence of bubbles during the sensing period is ensured by designating bubble carriers with low expected node mobility as *bubble anchors*, which maintain bubbles on behalf of their creators, should they become disconnected from the bubble. Nevertheless, bubbles may disappear in absence of anchors. To encounter this issue and restore orphaned bubbles, the participants additionally contact the bubble server in regular time intervals to search for bubbles near their current location, and join them if the required sensing modalities match.

In all three presented approaches for task distribution, the tasking and downloading processes may endanger the privacy of the participants in several ways. First, downloading tasks provides information to the tasking server about the location of the participants at precise timestamps, while the nature of the tasks provides hints about the devices used. Additionally, even when pseudonyms are used, the tasking server may infer the identity of the participants by tracking their locations over multiple downloads, as it may expose the location of their workplaces and homes (Krumm, 2007). To protect the participants, the following mechanisms for ensuring anonymity and location-privacy have been proposed:

- *Using tasking beacons* (Kapadia et al., 2009). The participating devices receive the broadcast beacons including the sensing tasks without having to register/authenticate themselves to a central entity.
- *Downloading the tasks in densely populated locations* (Shin et al., 2010). The high density of people present at such locations makes the identification of the participants by the server difficult, and hence conceals their identities.
- *Using attribute-based authentication* (Kapadia et al., 2009). Instead of using their precise identity, the participants can use cryptographic-based credentials showing their memberships to a particular group (e.g., students registered in the cycling club of the university), as realized in Camenisch and Stadler (1997). Again, the identities are hidden within the group of participants and the level of protection depends on the size of this group.
- *Using location privacy-preserving routing schemes* (Kapadia et al., 2009). Although not directly anonymizing the participants, these schemes hide their location using specific router/relay organization (Al-Muhtadi et al., 2002; Dingledine et al., 2004). For example, the TOR-based routing scheme used in Shin et al. (2010), anonymizes the connections to the tasking server using multiple relays and onion routing to hide the IP address, and thus information about the current location, of the participants.

Moreover, a malicious tasking entity may submit tasks with restrictive acceptance conditions including, e.g., a rare sensor type or specific mobile phone brands. Known as *narrow tasking* (Shin et al., 2010), this attack may allow the attacker to de-anonymize the mobile phones accepting these highly device-specific tasks, as only one or few mobile phones share these restrictive conditions. A countermeasure to the narrow tasking attack consists of introducing a trusted third party storing the attributes of all mobile phones and verifying that a sufficient number of mobile phones (above a pre-determined threshold) are able to fulfill the acceptance conditions in order to protect the anonymity of the mobile phones accepting the tasks.

Furthermore, a malicious tasking entity may attempt to differentiate and identify anonymous participants by launching *selective tasking* attacks (Shin et al., 2010), where the tasking entity distributes a task to only a restricted pool of mobile phones (greater than the aforementioned threshold). The selective tasking attack differs from the narrow tasking attack, as the selective tasking attack aims at linking the anonymous mobile phones uploading the tasking reports to the reports themselves, while the narrow tasking attack straightforwardly infers the identity of the devices/participants based on the task acceptance. As the amount of tasked devices is restricted, the adversary can easily link each anonymous mobile phone to the reports it has uploaded. A further analysis of the uploaded reports may breach the anonymity of mobile phones. To prevent this attack, the responsibility of selecting the tasks can be transferred from the tasking server to the participants themselves, who select a random amount of available tasks to execute.

While protecting the anonymity and privacy of the participants, these mechanisms simultaneously impact the performance of the sensing applications in terms of data integrity and the associated overhead. Since the identity of the participants is not revealed to the application, the anonymous devices may report falsified or faulty data, and the application will not be able to identify them and eliminate them from the tasking process in the future. Moreover, without knowing the participant's identity and location as well as the device specifications, the application may need to task a larger pool of devices to obtain similar results when compared to non-anonymous tasking. This may cause additional resource consumption and delays, potentially affecting the results of time-constrained applications.

4.3. Anonymous and privacy-preserving data reporting

In most applications, captured sensor data are reported to the central server directly after they have been recorded. Almost exclusively all participatory sensing applications record location/time traces (cf. Section 3). A prominent attack is thus the inference of location traces, as all presented participatory sensing campaigns share the common attribute of collecting location information along with the sensor data. The collection of location traces over several reports may allow to identity frequently visited locations, and the disclosure of the raw location data is likely to reveal the identity of participants and may thus endanger their privacy.

Besides the actual data contained in the reports, metadata collected from submissions may also threaten the privacy of the participants. As data transfers mostly make use of communication infrastructure available to the mobile phone, such as Wireless LAN, or GSM/GPRS/3G connectivity (Dong et al., 2008; Eisenman et al., 2009), location information can be extracted from the IP address assigned at the time of submission, or the upload intervals and schedules at which data are being transferred to the server. Based on the upload schedule, additional conclusions about the whereabouts of the participants can be drawn, even without considering the primary sensor readings. Similarly to an in-depth inspection of reports, the participants may be identified and de-anonymized by analyzing the metadata collected across multiple reports (Shin et al., 2010).

In this section, we consider the mechanisms used to protect the privacy of the participants against the analysis of reporting patterns and report contents.

4.3.1. Pseudonymity

A common mechanism to protect the anonymity and privacy of the participants is the use of pseudonyms (cf. Section 4.2). Instead of transmitting names in plain text, all interaction with the application is performed under an alias. Pseudonymity is currently used in various applications, including Shilton et al. (2008), Shilton (2009), and Deng and Cox (2009). When used in conjunction with authentication mechanisms, pseudonym-based solutions suggest anonymity and confidentiality to the user (Shilton et al., 2008). The participants tend to share their sensor readings without apprehension as they feel more protected behind pseudonyms. This subjective feeling however leads to a false perception of security, as the use of pseudonyms does not necessarily guarantee privacy in location-based applications. As demonstrated in Huang et al. (2005), an analysis of the reported data in conjunction with the reporting patterns may allow identifying the participants' residences among other significant places, such as workplaces and favorite entertainment centers, based on their location traces. The residence addresses may then be exploited to find the corresponding participants' real names using reverse address lookups. Pseudonyms must therefore be complemented by additional mechanisms to protect the participants' locations (during both the sensing and reporting processes) to efficiently provide privacy guarantees. The anonymity-based TOR network (Dingledine et al., 2004) is used in Shin et al. (2010) to hide the origins of reports and prevent an identification of the real identities of the participants from their locations. Before transmitting reports, the mobile phones select random relays along the path to the application server, instead of a direct route. The selected routes are then appended to the reports using a layered scheme, similar to onion layers. At each relay on the selected routes, a layer is removed using a symmetric key shared between this relay and the mobile phone. As a result, no relay knows the complete path from the report's source to the application server, but only the identities of preceding and following hops/relays.

4.3.2. Spatial cloaking

In addition to privacy-aware routing (e.g., TOR-based networks), mechanisms based on k-anonymity (Sweeney, 2002) can be applied to protect the location privacy of the participants who upload reports. The key idea behind k-anonymity is to build groups of k participants or reports such that they share a common attribute (e.g., k participants located in the same district), rendering them indistinguishable from each other. Different methods can be used to find an appropriate and common attribute in order to construct groups of k users. These methods can be classified into the two main categories of *generalization* and *perturbation* (Huang et al., 2010).

In the former, the original value of the attribute is generalized by a value with less degree of detail. For example, the exact coordinates of the k participants are replaced by the name of the district of their current location. In contrast, perturbation is based on replacing the original sensor data by a new value resulting from a function applied to the k sensor readings of the group members. For example, the location of each group member can be replaced by the average location of all members.

Tessellation is a form of generalization based on the division of a geographic area into multiple tiles. It is applied in Shin et al. (2010), where tiles are established by applying Voronoi partitioning to a map of Wireless LAN access points. The access point in the center of each tile keeps track of the average number of connected devices, which is equal to the maximum value of k_t that can be achieved within the tile. To guarantee k-anonymity throughout the network, neighboring tiles with $k_t < k$ are combined into cells with an effective value equal to the sum of each individual tile. Once the cells have been defined, the participants tag their sensor data with the geographical boundary of their current cell instead of supplying their exact coordinates. Alternatively, instead of transmitting the dimensions of the cell, the participants may report the geographical center of the cell, as proposed in Huang et al. (2010).

In contrast, microaggregation (Domingo-Ferrer and Mateo-Sanz, 2002) does not generalize the participant location to a cell, but replaces the real location by the averaged location of the k nearest participants, the so called *equivalence class*. Setting up equivalence classes is known to be NP-hard (Brucker, 1978), and among the proposed heuristics, the Maximum Distance to Average Vector (MDAV) algorithm has been shown to be very efficient in setting up equivalence classes. However, as user mobility is inherent in participatory sensing applications, equivalence classes may need to be changed dynamically and thus protect new users less efficiently. Frequent recomputations in turn may negate the efficiency advantage of MDAV. The V-MDAV (Solanas et al., 2006) algorithm offers variable class sizes and improves the performance of its predecessor in dynamic environments. The recursive algorithm is composed of two main steps. In the first step, k participants are clustered based on their locations and the relative Euclidean distances. Once the clusters are formed, additional participants can join the clusters in the second step, even if each cluster already contains k members and thereby fulfills the k-anonymity requirement.

A risk to k-anonymity is the possibility of homogeneity attacks, as outlined in Machanavajjhala et al. (2007). Such attacks exploit the monotony of certain attributes to identify individuals from the set of k participants. The authors thus present an extension to k-anonymity, termed l-diversity, which additionally requires the group members to provide at least l different values for the sensed attribute of interest. As a result, at least l distinct values for the sensitive attributes are present within each user group, which represents an effective countermeasure to homogeneity attacks. The principle of l-diversity is employed in Huang et al. (2010), where an l-diverse extension to the aforementioned V-MDAV (LD-VMDAV) algorithm is proposed and evaluated. The authors also extend the V-MDAV scheme further by introducing Hybrid V-MDAV, which combines V-MDAV with tessellation (with tile center reporting). The hybrid scheme thus benefits from the advantages of both approaches. Tessellation is applied if a user can construct a tile within his own cell, meaning that at least k other users share the same cell. Otherwise, the V-MDAV scheme is used, as it performs better in case of sparse distribution of users across multiple cells.

Nevertheless, these approaches rely on a trusted third-party managing the generalization or perturbation of the locations for all participants. To generate the cloaked values, the participants need to report their exact locations to the third-party entity. A further improvement in privacy is achieved by adding Gaussian noise on the location information, as proposed in Huang et al. (2010), to perturb and blur the locations of the participants before sending location information to the third-party, as detailed in the following section.

4.3.3. Data perturbation

Data perturbation intentionally perturbs the sensor samples by adding artificial noise to the data at the mobile phone side, such as Gaussian noise. The overall intention of data perturbation is to determine community trends and distributions without revealing individual data. The characteristics of the applied noise must thus be chosen carefully, as it needs to perturb individual sensor readings sufficiently while still ensuring that the statistical trend remains unaffected. For example, independent random noise has been demonstrated to be insufficient to prevent adversaries from reconstructing the original data in Krumm (2007).

A data perturbation scheme particularly tailored to the requirements of participatory sensing applications was proposed in Ganti et al. (2008). Its principle is as follows. First, a noise model with characteristics similar to a realistic data set is generated using an approximate model of the phenomenon monitored by the application. Note that preliminary knowledge about the data distribution is required, which may not be available in all participatory sensing applications. The designed model, which is composed of a structural description as well as the probability distribution of the parameters, is then distributed to the community. The participants use the distributed model to locally generate noise and superimpose it on their sensor readings. To complicate the reconstruction of each individual data, the proposed approach allows the participants to change the values of the noise generation parameters regularly. The data perturbed by the participants are then reported to the application. As the statistical characteristics of the noise model are known, the sum, average and distribution of the added noise over the data of all participants can be approximated. The community results (including trends and distribution) can be estimated by subtracting the average noise time series from the sum of all individual perturbed data, the precision of the estimation increasing with the community size. Other examples that make use of data perturbation, though not in the context of participatory sensing, can be found in Agrawal and Srikant (2000), Agrawal and Aggarwal (2001), and Evfimievski et al. (2003).

4.3.4. Hiding sensitive locations

Sensitive locations can be selected by the participants and protected using location selective hiding (Mun et al., 2009), which was introduced in the PEIR campaign presented in Section 2.1.2 and represents an alternative to data perturbation. When the user approaches a location which has been priorly defined as sensitive, the application generates fictitious location traces which intentionally avoid the selected location. However, the generated traces remain realistic, i.e., following exiting roads and streets. The algorithm selects the closest routes first, then refines the selection by taking the history of the participants' results into account (e.g., their physical capacities based on their preceding experiences). Furthermore, the algorithm even shifts the activities timely and modifies their duration to maintain the consistency of the application results.

In comparison with data perturbation (Section 4.3.3) and spatial cloaking (Section 4.3.2), the location selective hiding scheme improves the location privacy without impacting the application results. In fact, data perturbation and spatial cloaking only modify the location information without considering the consistency of the results and visits to the sensitive locations may still be identifiable after the application of these schemes depending on the granularity and noise model selected.

4.3.5. Data aggregation

In comparison to the previous schemes, the privacy-preserving data aggregation approach proposed in Shi et al. (2010) does not rely on a central entity to protect data privacy, but on a mutual protection within participants. Before transmitting data to the server, the mobile phones partially distribute their data among their neighbors. The mobile phones then upload the sensed data coming from their neighbors and the remaining of their own data. This distribution diminishes the probability to successfully attribute each sensor reading to the mobile phone which actually captured it. For example, if two mobile phones A and B exchange half of their data, the probability that the data reported by A was actually captured by itself is 50% (in absence of any additional and prior knowledge), and the same for B.

Moreover, this approach does not require any preliminary knowledge about the data distribution which is necessary in data perturbation schemes (cf. Section 4.3.3). Depending on the nature of the aggregation functions, two distinct schemes can be applied. For additive functions, each mobile phone/node partitions its data into $n+1$ slices and sends one slice to each of n selected nodes. There are three ways in which nodes can be selected. In the first model, the nodes are selected randomly regardless of their location leading to an additional energy consumption overhead if multi-hop communication is supported. In the second model, the one-hop neighbors are selected via a single broadcast. The efficiency of the privacy protection directly depends on the density of neighbors located in the broadcast regions, as n is lower in sparsely populated regions than in dense ones. An h-hop version is proposed in the third model, where h is a system parameter. Once each node has distributed its slices to his neighbors, the exchanged slices and the node's own slice are combined and sent to the aggregation server which is then able to compute the aggregation result. For non-additive aggregation functions, such as percentiles and histograms, a method combining slicing, count query, and binary search can be applied. Nevertheless, this approach only ensures data privacy protection if the nodes and the server do not conspire to breach the privacy of potential targets. We present a simple example to illustrate this. Assume that a node A is surrounded by two malicious nodes B and C. A partitions its data into three slices and distributes two of them to B and C, which do the same. Then the three nodes report the mixed slices to the server. As B and C can recognize their own slices and those distributed by A, the server can easily reconstruct the complete data set from the three uploaded slices and associate it to A.

4.4. Privacy-aware data processing

In typical participatory sensing applications, data processing is shared between the mobile phones and the application server. However, due to the resource constraints on mobile platforms, the distribution of the processing tasks between both parties is typically biased towards the server. While preprocessing is generally carried out on the phones to reduce the amount of data to transfer in order to save bandwidth and transmission energy, complex processing tasks may exceed the computational power of mobile phones, mandating the execution on the server.

In the applications discussed in Section 2, the data processing on the mobile phone mainly constitutes extracting features from the raw data to remove sensitive information (e.g., human voices recorded, or people photographed) endangering the privacy of the participants and for resource saving purposes. For example, an audio classifier can analyze the sound samples to determine whether human voices were recorded (Miluzzo et al., 2008). Further, the loudness level of the audio samples can be determined locally by running signal processing algorithms to minimize the data to transfer to the server (Maisonneuve et al., 2009). After processing, the raw data may be deleted from the local storage and the processed summaries are reported to the central server.

On the server side, the reported data may then be processed to eliminate privacy-sensitive information, such as the identity or data characteristics threatening the anonymity/privacy of the participants. For example, the captured data can be aggregated among several participants to render them indistinguishable or published in the form of statistics (Ganti et al., 2010) and maps (Dong et al., 2008). By doing so, sensitive data is not directly revealed to the end users, which avoids direct identification of the participants. However, the participants must rely on the application to: efficiently anonymize the data, sufficiently protect their privacy, and not disclose the privacy-sensitive information contained in their reported data to third parties.

4.5. Review, deletion, storage, and retention of data

After the collecting sensor data, the participants can review them to verify that they do not contain sensitive information (e.g., faces in pictures, or sensible locations) and judge of the appropriateness of the released information. If privacy-sensitive or inappropriate items are identified, the participants can discard and delete them before the data is being reported to the application server (Abdelzaher et al., 2007). Alternatively, the application can automatically discard the buffered sensor data unless the participants review them and indicate their willingness to share the same (Shilton et al., 2008).

Most of the applications discussed earlier rely on a centralized system storage managed by the application itself. The whole pool of sensor data is easily accessible for processing, which is advantageous for the application. However, this solution reduces the participants' control over their data. Once they have uploaded their data, the participants must trust the central entity not to disclose them to unauthorized third parties. Even if applications may authorize the users to delete the uploaded data or adopt retention policies favorable for the participants (e.g., deleting the location information every six months by default (Mun et al., 2009)), the participants do not typically receive any confirmation that data has been definitely removed from the server.

To address this loss of control, a short-term solution may be to locally store the sensitive data on the mobile phones to prevent third parties from accessing and potentially misusing them. However, the mobile phones often suffer from resource and energy constraints, which effectively limit storage and processing capability. Moreover, this storage modality may not be appropriate to community-oriented applications, where global processing may be necessary to highlight interesting features of the data.

A solution for the secure data storage is the use of so called *personal data vaults* (Mun et al., 2010), which uncouple the acquisition of sensor data and their secure storage. Personal data vaults are individually controlled secure data repositories, which may only be fully accessed by their owner. The owner may however choose to share information based on its time or location annotations, or provide post-processed data to external services. Personal virtual machines also represent an alternative to local storage, designated as *virtual individual servers* (Cáceres et al., 2009), where the participants can upload their raw data. The data is only uploaded once and may then be released to all applications the participants are involved in. Different applications can be authorized to access different sets of data according to their demands. Consequently, the participants maintain the control over their data and can dynamically determine potential recipients of selected sets of data. In comparison with a centralized scheme, this approach may generate additional management overhead for the participants, but they retain ownership of their data and can directly control their privacy.

4.6. Access control and audit

Depending on the application scenario, the sensing results may not only be of interest for the participants, but also different stakeholders, e.g., researchers, medical staff, friends, family members, city councils or a larger public. However, the participants may not be willing to share their data with all types of people within the stakeholder group and with the same granularity. In addition to addressing privacy concerns before the sensing process (cf. Section 4.1) and the data release to the application (cf. Sections 4.4 and 4.5), the participants can define the intended audience who are authorized to access their data from the user interface of many applications. They can define groups (Grosky et al., 2007; Gaonkar et al., 2008; Shilton, 2009), select persons individually (Reddy et al., 2007; Miluzzo et al., 2008; Gaonkar et al., 2008; Shilton, 2009) or authorize everyone (Gaonkar et al., 2008). The participants can refine their selection by specifying the nature of the data they share and may additionally define particular subsets of accessible data (Reddy et al., 2007; Miluzzo et al., 2008; Shilton, 2009). Furthermore, they can define precise release conditions (e.g., time, location, and data type) under which the data are made accessible (Shilton et al., 2008). Similarly to the selection of the sensing modalities presented in Section 4.1, these actions support the expression of the concept of appropriateness and its personalization by the participants.

To highlight the privacy implications of sharing their data, graphic tools including maps, charts, or pictures are used to visualized the data being released and increase the participants' awareness (Shilton et al., 2008). After data has been published, the participants can also monitor access to the data by consulting application log files. These logs records the nature of the accessed data, the frequency of these accesses, and the identity of the people accessing it (Shilton et al., 2008; Shilton, 2009; Mun et al., 2010). Based on the results of these audits, the participants control the distribution of their information and can judge of their appropriateness. If needed, they may update their access control policies to restrict or enlarge the authorization conditions in order to match their privacy preferences.

5. Future research directions

Above, we have discussed the privacy threats in participatory sensing applications and we have surveyed selected privacy solutions to mitigate these threats. We found that tailored and practical privacy solutions are scarce, and that fundamental research in the field of privacy for mobile participatory sensing is still in its infancy. Thus, a broad range of research challenges still remain unsolved. In the following, we highlight future research directions in the area of privacy in mobile participatory sensing applications, both from the perspective of the authors of the surveyed work as well as from our own perspective; note that our list of given challenges is by no means exhaustive, but contains our subjective impression of the most relevant challenges at the time of writing this article.

5.1. Challenge 1: including the participants in the privacy equation

We identified the three groups of stakeholders in Section 2.3.1 to be the campaign administrators (providing the application platform, maintaining the infrastructure, etc.), the participants (contributing the sensor data), and the end-users (consuming the obtained sensing results). A key challenge for the future is to better include the participants (or unsuspecting participants)[2] in the privacy equation. This includes aspects such as:

- *Tailored privacy interfaces*: The notion of privacy is highly individual and depends on the user's views and opinions. It is thus crucial to create awareness for privacy threats to the user, and assist him by making the complex configuration of participatory sensing applications easily understood. However, most of the discussed privacy-preserving countermeasures do not present a user interface that can raise awareness and facilitate the comprehension of the complex mechanisms in use. Tailored user interfaces, which consider unique features of mobile devices, are thus highly sought after. Moreover, the existence of such interfaces can encourage the acceptance and later, the adoption of the privacy-preserving mechanisms by the participants, as they will be able to better understand (through the interfaces) the ins and outs of the mechanisms.
- *Ease of use*: The usability of the application and its privacy settings needs to be taken into consideration. Extensive manual configuration often overstrains a participant's patience and leads him to either leave the default settings or not understand the implications of his choices, as demonstrated in the case of privacy settings in online social networks (Gross and Acquisti, 2005).
- *Transparency of privacy protection levels*: To judge whether the offered privacy protection is adequate, users need to be able to compare the offered level of protection against their individual protection requirements. Although most of the surveyed solutions base their evaluations on mathematical and verifiable metrics, the user perception and their level of satisfaction with existing solutions has not been explicitly regarded yet.
- *Incorporation of user feedback*: Besides providing user interfaces to configure privacy levels, insights about how the protection is perceived and to which extents users are involved in configuring their privacy settings are required to bring forward the usability.

  User studies are deemed an integral tool to assess the usability and utility of privacy solutions (Mun et al., 2010). Existing studies in the field already correlate privacy concerns with the used sensing modalities (Klasnja et al., 2009), or analyze how the participants understand, select, and feel comfortable with different obfuscation methods to achieve location privacy (Brush et al., 2010).

In the future, the integration of the participants in the privacy equation could be supported by concepts and methodologies issued from *participatory action research* (McNiff, 1988) and *community-based participatory research* (Altman, 1995). Even though, these ideas have been conceived for other application domains, the participatory sensing community could benefit from these tools, which promote tight cooperation between participants and application developers to build solutions addressing the needs of the participants based on their direct feedback.

5.2. Challenge 2: providing composable privacy solutions

We have witnessed a tremendous growth in mobile participatory sensing applications during the last years. Novel sensing modalities have been incorporated with the ongoing technological development of mobile phone platforms, leading to a growing family of applications with innumerous application-specific privacy challenges. To be able to deal with this broad range of scenarios, it is necessary to cater for composable and adaptable privacy solutions.

- *Application independence vs. tailored privacy solutions*: Some of the presented countermeasures are tailored to specific application scenarios. For example, hiding sensitive locations by creating fake location traces to avoid correlations between users and locations (Section 4.3.4) was only evaluated with the PEIR scenario (Section 2.1.2). Further scenarios need to be investigated to determine the potential limits and drawbacks of the proposed solutions depending on the application specifics. This investigation will highlight necessary changes to proceed from tailored privacy solutions to application-agnostic privacy concepts.
- *A system approach to preserving privacy*: In the various countermeasures that we have discussed (Section 4), only selected privacy aspects are addressed. However, to gain widespread acceptance among the participants, a flexible privacy architecture that addresses the problem from a system perspective is essential. By analyzing assets and drawbacks of centralized and distributed system components, a research hypothesis could be that the combination of both distributed and centralized approaches allows to better represent the multi-party nature of the privacy trade-off between all stakeholders.
- *Privacy for evolutionary sensing scenarios*: In scenarios, in which the characteristics of sensor data are known in advance, privacy solutions can be adapted accordingly, e.g., by adding noise with corresponding properties (cf. Section 4.3.3). However, in case of dynamic and/or unpredictable sensing scenarios, where the characteristics of sensor data cannot be determined in advance, novel privacy concepts need to be devised.

5.3. Challenge 3: trade-offs between privacy, performance, and data fidelity

Strong mechanisms (such as removal or obfuscation of sensor readings, as shown in Section 4.3.2) for privacy protection might influence the data fidelity, the sensing delay, or the integrity of the sensed data. Protecting the integrity of sensor data however counteracts mechanisms for privacy preservation. In consequence, a trade-off between privacy guarantees and sensing fidelity is necessary.

- *Anonymity vs. data quality/integrity*: In all participatory sensing applications, user participation needs to be encouraged by guaranteeing their privacy. At the same time, this makes the systems vulnerable to malicious users and faulty devices, which may contribute corrupted or erroneous data to the applications. To prevent this data from degrading the accuracy of the application results, the devices or the data in question need to be identified and discarded from the pool of tasked devices/sensed data. Research on reputation systems that cater for both anonymity

[2] In April 2011, the Apple iPhone was publicly discussed in mainstream media, since its operating system was tracking information about encountered wireless access points and GSM cells without prior notice to the (unsuspecting) users. The main purpose of collecting these data in a participatory manner was claimed to be the use of a "crowd-sourced Wi-Fi hotspot and cell tower database which is downloaded from Apple into the iPhone to assist the iPhone in rapidly and accurately calculating location" (Apple Inc., 2011).

and the requirements and specifics of the sensing scenarios is therefore required.

- *Multi-party privacy protection*: While it has been shown in Tang et al. (2010) that participants value the location privacy of their friends, most of the current privacy-preserving mechanisms focus on the protection of the participants themselves. But the privacy of people in their surroundings may also be threatened, as shown in DietSense, where faces of uninvolved people can appear in the contributed images. In current systems, it is mostly the user's responsibility to protect the privacy of bystanders. However, malicious participants may also deliberately distribute compromising information about others. Current solutions in related domains specifically deal with the conflict between the owner of a photo and the people tagged in it (Besmer and Richter Lipford, 2010). Automated solutions to minimize the captured data such that it does not violate the privacy of others is of high interest.
- *Overriding privacy*: Although the protection of sensitive data is highly valued, certain situations, like emergency scenarios, necessitate means to override the specified privacy settings. This can be compared to privacy issues encountered in health scenarios (discussed in Section 2.1.1), where physicians might be able to override the access control of body sensors to get access to critical health data. Approaches for controlled access in case of emergency, e.g., by explicitly defining exception conditions and controlling that data are only made available when emergency conditions are met, are thus of high relevance.

5.4. Challenge 4: making privacy measurable

Different methods, criteria, or metrics are currently being used to evaluate the performance of the proposed solutions in terms of privacy protection. While it might be difficult or even impossible to come up with universal metrics to quantify privacy, the need to define generalized metrics is widely acknowledged. Capturing the level of privacy protection independent from the particular application scenario can thus be seen as a long-term research goal. The definition of generalized privacy metrics, independent from the application scenarios, is mandatory to achieve a common basis for comparing privacy-preserving mechanisms.

- *Generalized privacy metrics*: Similarly to the approach of attaining a common understanding of anonymity (Hughes and Shmatikov, 2004) due to the incapacity of current anonymity metrics (Serjantov and Danezis, 2003), universal privacy metrics are required to quantify degrees of privacy. To obtain such metrics, currently employed privacy metrics need to be surveyed to determine what input parameters (e.g., amount of participants in the same region, actual and perturbed location traces) are considered necessary to calculate the degree of privacy and what is the nature of the output parameters (e.g., Euclidean distance between the actual and perturbed traces) depending on the application scenarios. Additionally, privacy metrics from other application domains should to be analyzed with respect to their applicability in participatory sensing. A comprehensive and generalized framework for privacy metrics could be the result of further research.
- *Provable guarantees for privacy*: Most of the countermeasures discussed in Section 4 rely on a central entity to protect user privacy and anonymity. However, in the existing solutions there are no guarantees or proofs that the promised degree of privacy is respected, implementation details are hardly available, and even the general approach towards protecting the privacy of users is typically unknown. As a result, current systems are mostly designed as a black box, where the participants must trust the application that the announced privacy-preserving mechanisms are applied and the contributed data are not disclosed to third parties. Research into provability of the privacy mechanisms and into proving the correctness of the implementation of these mechanisms still remains an open field of research.

5.5. Challenge 5: defining standards for privacy research

Due to their sensitive nature, public real-world data sets for participatory sensing applications are scarce. Hence, privacy research mostly operates on either private or synthetic data sets. As a result, no well accepted data basis for the evaluation of novel mechanisms exists, and mechanisms cannot easily be benchmarked against each other.

- *Open data sets*: To overcome the limitations of researching on private or non-reproducible data, the research community should provide for open data sets that can serve as a baseline for performance and security evaluations. This includes real-world data sets as well as representative synthetic data sets for various different sensing modalities.
- *Open privacy solutions*: Implementations of privacy mechanisms, as discussed in Section 4, are often unavailable to the general public, thus making it hard or even impossible to benchmark them against newly proposed mechanisms. Making either a detailed technical description of the implementation or the implementation itself available to the research community allows to cross-validate the results and to benchmark individual solutions.

5.6. Challenge 6: holistic architecture blueprints

Finally, to implement privacy-preserving participatory sensing applications, it is important to provide a solution covering the entire system and satisfying the requirements of all stakeholders. To this end, architectural blueprints for privacy-preserving participatory sensing systems are required. To allow for a system-wide support, the individual components as well as their interworking needs to be analyzed. Research on holistic system support for privacy-preserving participatory sensing could be one step towards this end that complements the platform support on the mobile phones as well as the application platforms.

6. Conclusions

Participatory sensing leverages the ubiquity of mobile phones to open new perspectives in terms of sensing. Within the scope of this survey, we have analyzed existing applications and mapped their components into a generic system model, and identified the different modalities of sensor data contributed by the participants. Our analysis has revealed that virtually all applications capture location and time information. This information is used either as self-contained data, or to geo-tag and timestamp other collected sensor readings including pictures, sound samples, acceleration, pollution, and biometric data. We have then examined the extent of personal information that can be inferred from the collected sensing modalities, both individually and in combination. After having provided a definition of privacy in participatory sensing, potential threats to user privacy resulting from the uncontrolled disclosure of personal information to untrusted people have been highlighted. As these threats may discourage participants from contributing sensor readings, we have surveyed current state-of-the-art privacy countermeasures, and analyzed their position within the architecture of participatory sensing systems. Based on the state-of-the-art, we have presented and discussed future research directions which need to be tackled to efficiently protect privacy and, consequently, encourage the contribution of the participants.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This work was supported by CASED (www.cased.de).

References

Abdelzaher, T., Anokwa, Y., Boda, P., Burke, J., Estrin, D., Guibas, L., Kansal, A., Madden, S., Reich, J., 2007. Mobiscopes for human spaces. IEEE Pervasive Computing 6, 20–29.

Acquisti, A., Grossklags, J., 2005. Privacy and rationality in individual decision making. IEEE Security and Privacy 3, 26–33.

Agrawal, D., Aggarwal, C., 2001. On the design and quantification of privacy preserving data mining algorithms. In: Proceedings of the 12th ACM Symposium on Principles of Database Systems (PODS), pp. 247–255.

Agrawal, R., Srikant, R., 2000. Privacy-preserving data mining. ACM Sigmod Record 29, 439–450.

Ahn, G.S., Musolesi, M., Lu, H., Olfati-Saber, R., Campbell, A., 2010. MetroTrack: predictive tracking of mobile events using mobile phones. Distributed Computing in Sensor Systems 6131, 230–243.

Al-Muhtadi, J., Campbell, R., Kapadia, A., Mickunas, M., Yi, S., 2002. Routing through the mist: privacy preserving communication in ubiquitous computing environments. In: Proceedings of 22nd IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 74–83.

Altman, D., 1995. Sustaining interventions in community systems: on the relationship between researchers and communities. Health Psychology 14, 526–536.

Annavaram, M., Medvidovic, N., Mitra, U., Narayanan, S., Sukhatme, G., Meng, Z., Qiu, S., Kumar, R., Thatte, G., Spruijt-Metz, D., 2008. Multimodal sensing for pediatric obesity applications. In: Proceedings of International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems (UrbanSense), pp. 21–25.

Apple Inc., 2011. Apple Q&A on location data. Online: http://www.apple.com/pr/library/2011/04/27location_qa.html (accessed in April 2011).

Azizyan, M., Constandache, I., Roy Choudhury, R., 2009. SurroundSense: mobile phone localization via ambience fingerprinting. In: Proceedings of the 15th ACM Annual International Conference on Mobile Computing and Networking (MobiCom), pp. 261–272.

Bao, X., Roy Choudhury, R., 2010. MoVi: mobile phone based video highlights via collaborative sensing. In: Proceedings of the 8th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 357–370.

Beresford, A., Stajano, F., 2003. Location privacy in pervasive computing. IEEE Pervasive Computing 2, 46–55.

Beresford, A.R., 2005. Location privacy in ubiquitous computing. Technical Report 612. University of Cambridge. http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-612.pdf.

Besmer, A., Richter Lipford, H., 2010. Moving beyond untagging: photo privacy in a tagged world. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI), pp. 1563–1572.

Brandeis, L., Warren, S., 1890. The right to privacy. Harvard Law Review 4, 193–220.

Brucker, P., 1978. On the complexity of clustering problems. System Modeling and Optimization 157, 45–54.

Brush, A.B., Krumm, J., Scott, J., 2010. Exploring end user preferences for location obfuscation, location-based services, and the value of location. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing (Ubicomp), pp. 95–104.

Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M., 2006. Participatory sensing. In: Proceedings of the 1st Workshop on World-Sensor-Web (WSW), pp. 1–5.

Cáceres, R., Cox, L., Lim, H., Shakimov, A., Varshavsky, A., 2009. Virtual individual servers as privacy-preserving proxies for mobile devices. In: Proceedings of the 1st ACM Workshop on Networking, Systems, and Applications for Mobile Handhelds (MobiHeld), pp. 37–42.

Camenisch, J., Stadler, M., 1997. Efficient group signature schemes for large groups. Advances in Cryptology (CRYPTO) 1294, 410–424.

Campbell, A., Eisenman, S., Lane, N., Miluzzo, E., Peterson, R., 2006. People-centric urban sensing. In: Proceedings of the 2nd Annual International Wireless Internet Conference (WICON), pp. 18–31.

Campbell, A., Eisenman, S., Lane, N., Miluzzo, E., Peterson, R., Lu, H., Zheng, X., Musolesi, M., Fodor, K., Eisenman, S., Ahn, G., 2008. The rise of people-centric sensing. IEEE Internet Computing 12, 12–21.

Carrapetta, J., Youdale, N., Chow, A., Sivaraman, V., 2010. Haze Watch project. Online: http://www.pollution.ee.unsw.edu.au (accessed in January 2011).

CBC News, 2009. Depressed woman loses benefits over Facebook photos. Online: http://www.cbc.ca/news/canada/montreal/story/2009/11/19/quebec-facebook-sick-leave-benefits.html (accessed in April 2011).

CBS News, 2011. Did the Internet kill privacy? Facebook photos lead to a teacher losing her job: what expectations of privacy exist in the digital era? Online: http://www.cbsnews.com/stories/2011/02/06/sunday/main7323148.shtml (accessed in April 2011).

Chang, K., Yau, N., Hansen, M., Estrin, D., 2006. SensorBase.org—a centralized repository to slog sensor network data. In: Proceedings of the Euro-American Workshop on Middleware for Sensor Networks (EAWMS), pp. 116–128.

Christin, D., 2010. Impenetrable obscurity vs. informed decisions: privacy solutions for participatory sensing. In: Proceedings of the 8th IEEE International Conference on Pervasive Computing and Communications (PerCom Workshop), pp. 847–848.

Das, T., Mohan, P., Padmanabhan, V.N., Ramjee, R., Sharma, A., 2010. PRISM: platform for remote sensing using smartphones. In: Proceedings of the 8th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 63–76.

Debatin, B., Lovejoy, J.P., Horn, A.K., Hughes, B.N., 2009. Facebook and online privacy: attitudes, behaviors, and unintended consequences. Journal of Computer-Mediated Communication 15, 83–108.

Deng, L., Cox, L., 2009. LiveCompare: grocery bargain hunting through participatory sensing. In: Proceedings of the 10th Workshop on Mobile Computing Systems and Applications (HotMobile), pp. 1–6.

Denning, T., Andrew, A., Chaudhri, R., Hartung, C., Lester, J., Borriello, G., Duncan, G., pp. 5:1–5:6 2009. BALANCE: towards a usable pervasive wellness application with accurate activity inference. In: Proceedings of the 10th workshop on Mobile Computing Systems and Applications (HotMobile).

Derawi, M., Nickel, C., Bours, P., Busch, C., 2010. Unobtrusive user-authentication on mobile phones using biometric gait. In: Proceeding of the 6th IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), pp. 306–311.

Dingledine, R., Mathewson, N., Syverson, P., 2004. Tor: the second-generation onion router. In: Proceedings of the 13th Conference on USENIX Security Symposium (USENIX Security), pp. 21–38.

Domingo-Ferrer, J., Mateo-Sanz, J., 2002. Practical data-oriented microaggregation for statistical disclosure control. IEEE Transactions on Knowledge and Data Engineering 14, 189–201.

Dong, Y., Kanhere, S., Chou, C., Bulusu, N., 2008. Automatic collection of fuel prices from a network of mobile cameras. In: Proceedings of the 4th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), pp. 140–156.

Eisenman, S., Campbell, A., 2006. SkiScape sensing. In: Proceedings of the 4th ACM International Conference on Embedded Networked Sensor Systems (SenSys), pp. 401–402.

Eisenman, S., Lane, N., Campbell, A.T., 2008. Techniques for improving opportunistic sensor networking performance. In: Proceedings of the 4th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS). Springer, pp. 157–175.

Eisenman, S., Lane, N., Miluzzo, E., Peterson, R., Ahn, G., Campbell, A., 2006. MetroSense project: people-centric sensing at scale. In: Proceedings of the 1st Workshop on World-Sensor-Web (WSW), pp. 6–11.

Eisenman, S., Miluzzo, E., Lane, N., Peterson, R., Ahn, G., Campbell, A., 2009. BikeNet: a mobile sensing system for cyclist experience mapping. ACM Transactions on Sensor Networks 6, 1–39.

Eisenman, S.B., Miluzzo, E., Lane, N.D., Peterson, R.A., Ahn, G.S., Campbell, A.T., 2007. The BikeNet mobile sensing system for cyclist experience mapping. In: Proceedings of the 5th ACM International Conference on Embedded Networked Sensor Systems (SenSys), pp. 87–101.

Estrin, D., 2010. Participatory sensing: applications and architecture. In: Proceedings of the 8th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 3–4.

Evfimievski, A., Gehrke, J., Srikant, R., 2003. Limiting privacy breaches in privacy preserving data mining. In: Proceedings of the 22nd ACM Symposium on Principles of Database Systems (PODS), pp. 211–222.

Ganti, R., Pham, N., Tsai, Y., Abdelzaher, T., 2008. PoolView: stream privacy for grassroots participatory sensing. In: Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems (SenSys), pp. 281–294.

Ganti, R.K., Pham, N., Ahmadi, H., Nangia, S., Abdelzaher, T.F., 2010. GreenGPS: a participatory sensing fuel-efficient maps application. In: Proceedings of the 8th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 151–164.

Gaonkar, S., Li, J., Choudhury, R., Cox, L., Schmidt, A., 2008. Micro-blog: sharing and querying content through mobile phones and social participation. In: Proceedings of the 6th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 174–186.

Good, N., Dhamija, R., Grossklags, J., Thaw, D., Aronowitz, S., Mulligan, D., Konstan, J., 2005. Stopping spyware at the gate: a user study of privacy, notice and spyware. In: Proceedings of the Symposium on Usable Privacy and Security (SOUPS), pp. 43–52.

Grosky, W., Kansal, A., Nath, S., Liu, J., Zhao, F., 2007. SenseWeb: an infrastructure for shared sensing. IEEE Multimedia 14, 8–13.

Gross, R., Acquisti, A., 2005. Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society (WPES), pp. 71–80.

Györbíró, N., Fábián, A., Hományi, G., 2009. An activity recognition system for mobile phones. Mobile Networks and Applications 14, 82–91.

Hoh, B., Gruteser, M., Herring, R., Ban, J., Work, D., Herrera, J.C., Bayen, A.M., Annavaram, M., Jacobson, Q., 2008. Virtual trip lines for distributed privacy-preserving traffic monitoring. In: Proceedings of the 6th ACM International Conference on Mobile systems, Applications, and Services (MobiSys), pp. 15–28.

Huang, K.L., Kanhere, S.S., Hu, W., 2010. Preserving privacy in participatory sensing systems. Computer Communications 33, 1266–1280.

Huang, Z., Du, W., Chen, B., 2005. Deriving private information from randomized data. In: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 37–48.

Hughes, D., Shmatikov, V., 2004. Information hiding, anonymity and privacy: a modular approach. Journal of Computer Security 12, 3–36.

Hull, B., Bychkovsky, V., Zhang, Y., Chen, K., Goraczko, M., Miu, A., Shih, E., Balakrishnan, H., Madden, S., 2006. CarTel: a distributed mobile sensor computing system. In: Proceedings of the 4th ACM International Conference on Embedded Networked Sensor Systems (SenSys), pp. 125–138.

Kanjo, E., Bacon, J., Roberts, D., Landshoff, P., 2009. MobSens: making smart phones smarter. IEEE Pervasive Computing 8, 50–57.

Kansal, A., Goraczko, M., Zhao, F., 2007. Building a sensor network of mobile phones. In: Proceedings of the 6th International Conference on Information Processing in Sensor Networks (IPSN), pp. 547–548.

Kapadia, A., Kotz, D., Triandopoulos, N., 2009. Opportunistic sensing: security challenges for the new paradigm. In: Proceedings of the 1st International Conference on Communication Systems and Networks (COMNETS), pp. 1–10.

Klasnja, P., Consolvo, S., Choudhury, T., Beckwith, R., Hightower, J., 2009. Exploring privacy concerns about personal sensing. In: Proceedings of the 7th International Conference on Pervasive Computing (Pervasive), pp. 176–183.

Krause, A., Horvitz, E., Kansal, A., Zhao, F., 2008. Toward community sensing. In: Proceedings of the 7th International Conference on Information Processing in Sensor Networks (IPSN), pp. 481–492.

Krumm, J., 2007. Inference attacks on location tracks. In: Proceedings of the 5th IEEE International Conference on Pervasive Computing (Pervasive), pp. 127–143.

LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., Schilit, B., 2005. Place lab: device positioning using radio beacons in the wild. Pervasive Computing 3468, 116–133.

Liu, L., 2007. From data privacy to location privacy: models and algorithms. In: Proceedings of the 33rd International Conference on Very Large Data Bases (VLBD), pp. 1429–1430.

Lu, H., Lane, N., Eisenman, S., Campbell, A., 2008. Bubble-sensing: a new paradigm for binding a sensing task to the physical world using mobile phones. In: Proceedings of the International Workshop on Mobile Devices and Urban Sensing (MODUS), pp. 58–71.

Lu, H., Lane, N.D., Eisenman, S.B., Campbell, A.T., 2010. Bubble-sensing: binding sensing tasks to the physical world. Pervasive and Mobile Computing 6, 58–71.

Lu, H., Pan, W., Lane, N., Choudhury, T., Campbell, A., 2009. SoundSense: scalable sound sensing for people-centric applications on mobile phones. In: Proceedings of the 7th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 165–178.

Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., 2007. L-diversity: privacy beyond K-anonymity. ACM Transactions on Knowledge Discovery from Data 1, 1–52.

Maisonneuve, N., Stevens, M., Niessen, M., Steels, L., 2009. NoiseTube: measuring and mapping noise pollution with mobile phones. In: Proceedings of the 4th International Symposium on Information Technologies in Environmental Engineering (ITEE), pp. 215–228.

McNiff, J., 1988. Action Research: principles and practice. MacMillan, London.

Miluzzo, E., Lane, N., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S., Zheng, X., Campbell, A., 2008. Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application. In: Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems (SenSys), pp. 337–350.

Mohan, P., Padmanabhan, V., Ramjee, R., 2008. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In: Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems (SenSys), pp. 323–336.

Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., Hansen, M., Govindan, R., pp. 17:1–17:12 2010. Personal data vaults: a locus of control for personal data streams. In: Proceedings of 6th International Conference on Emerging Networking Experiments and Technologies (CoNEXT).

Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., Hansen, M., Howard, E., West, R., Boda, P., 2009. PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In: Proceedings of the 7th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 55–68.

Musolesi, M., Miluzzo, E., Lane, N., Eisenman, S., Choudhury, T., Campbell, A., 2008. The second life of a sensor: integrating real-world experience in virtual worlds using mobile phones. In: Proceedings of the 5th Workshop on Embedded Networked Sensors (HotEmNets), pp. 12–16.

Myles, G., Friday, A., Davies, N., 2003. Preserving privacy in environments with location-based applications. IEEE Pervasive Computing 2, 56–64.

Nachman, L., Baxi, A., Bhattacharya, S., Darera, V., Deshpande, P., Kodalapura, N., Mageshkumar, V., Rath, S., Shahabdeen, J., Acharya, R., 2010. Jog Falls: a pervasive healthcare platform for diabetes management. Pervasive Computing 6030, 94–111.

Newell, P.B., 1995. Perspectives on privacy. Journal of Environmental Psychology 15, 87–104.

Nissenbaum, H., 2004. Privacy as contextual integrity. Washington Law Review 79, 101–139.

Paulos, E., Honicky, R., Goodman, E., 2007. Sensing atmosphere. In: Proceedings of the Workshop on Sensing on Everyday Mobile Phones in Support of Participatory Research (SenSys Workshop), pp. 15–16.

Rana, R., Chou, C., Kanhere, S., Bulusu, N., Hu, W., 2010. Ear-Phone: an end-to-end participatory urban noise mapping system. In: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp. 105–116.

Raya, M., Hubaux, J., 2007. Securing vehicular ad hoc networks. Journal of Computer Security 15, 39–68.

Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., Hansen, M., 2007. Image browsing, processing, and clustering for participatory sensing: lessons from a DietSense prototype. In: Proceedings of the 4th Workshop on Embedded Networked Sensors (EmNets), pp. 13–17.

Reddy, S., Samanta, V., Burke, J., Estrin, D., Hansen, M., Srivastava, M., 2009. MobiSense—mobile network services for coordinated participatory sensing. In: Proceedings of the International Symposium on Autonomous Decentralized Systems (ISADS), pp. 1–6.

Renaud, K., Gálvez-Cruz, D., 2010. Privacy: aspects, definitions and a multi-faceted privacy preservation approach. In: Proceedings of the 2010 Information Security for South Africa Conference (ISSA), pp. 1–8.

Schilit, B.N., LaMarca, A., Borriello, G., Griswold, W.G., McDonald, D., Lazowska, E., Balachandran, A., Hong, J., Iverson, V., 2003. Challenge: ubiquitous location-aware computing and the "Place Lab" initiative. In: Proceedings of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH), pp. 29–35.

Serjantov, A., Danezis, G., 2003. Towards an information theoretic metric for anonymity. Privacy Enhancing Technologies 2482, 259–263.

Shi, J., Zhang, R., Liu, Y., Zhang, Y., 2010. PriSense: privacy-preserving data aggregation in people-centric urban sensing systems. In: Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM), pp. 1–9.

Shilton, K., 2009. Four billion little brothers? Privacy, mobile phones, and ubiquitous data collection. Communications of the ACM 52, 48–53.

Shilton, K., Burke, J., Estrin, D., Hansen, M., Srivastava, M., 2008. Participatory privacy in urban sensing. In: Proceedings of the International Workshop on Mobile Devices and Urban Sensing (MODUS), pp. 1–7.

Shin, M., Cornelius, C., Peebles, D., Kapadia, A., Kotz, D., Triandopoulos, N., 2010. AnonySense: a system for anonymous opportunistic sensing. Journal of Pervasive and Mobile Computing 7, 16–30.

Siewiorek, D., Smailagic, A., Furukawa, J., Krause, A., Moraveji, N., Reiger, K., Shaffer, J., Wong, F.L., 2003. SenSay: a context-aware mobile phone. In: Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC), pp. 248–249.

Solanas, A., Martinez-Balleste, A., Domingo-Ferrer, J., 2006. V-MDAV: a multivariate microaggregation with variable group size. In: Proceedings of the 17th IASC Symposium on Computational Statistics (COMPSTAT), pp. 917–925.

Stuntebeck, E.P., Davis, J.S., Abowd, G.D., Blount II, M., 2008. HealthSense: classification of health-related sensor data through user-assisted machine learning. In: Proceedings of the 9th Workshop on Mobile Computing Systems and Applications (HotMobile), pp. 1–5.

Sweeney, L., 2002. K-anonymity: a model for protecting privacy. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems 10, 557–570.

Tang, K.P., Lin, J., Hong, J.I., Siewiorek, D.P., Sadeh, N., 2010. Rethinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing (Ubicomp), pp. 85–94.

Thiagarajan, A., Biagioni, J., Gerlich, T., Eriksson, J., 2010. Cooperative transit tracking using smart-phones. In: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys), pp. 85–98.

Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., Toledo, S., Eriksson, J., 2009. VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. In: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys), pp. 85–98.

von Kaenel, M., Sommer, P., Wattenhofer, R., 2011. Ikarus: large-scale participatory sensing at high altitudes. In: Proceedings of the 12th Workshop on Mobile Computing Systems and Applications (HotMobile), pp. 55–60.

Westin, A.F., 1967. Privacy and Freedom. Atheneum, New York.

Delphine Christin graduated in electrical engineering from Ecole Nationale Supérieure de l'Electronique et ses Applications, France, and Technische Universität Darmstadt, Germany, in 2009. She is a research assistant at the Center for Advanced Security Research Darmstadt (CASED) since April 2009 and a member of the Secure Mobile Networking Lab (SEEMOO) at Technische Universität Darmstadt since October 2009. Her research interests include privacy schemes for participatory sensing scenarios, privacy interfaces, and privacy metrics.

Andreas Reinhardt received the Dipl.-Ing. degree (M.Sc. equivalent) in electrical engineering and information technology from Technische Universität Darmstadt, Germany, in 2007. Since December 2007, he is a research assistant at the Multimedia Communications Lab (KOM) at Technische Universität Darmstadt, Germany, where he works towards his doctoral degree. His research interests are in networked and embedded systems, with specific focus on energy-efficient sensor networking.

Salil S. Kanhere obtained a B.E. in electrical engineering from VJTI, Bombay, India in 1998. Subsequently he joined the Department of Electrical and Computer Engineering at Drexel University in Philadelphia, USA as a post-graduate student. He received his M.S. and Ph.D., both in electrical engineering in 2001 and 2003, respectively. Since April 2004, he is with the School of Computer Science and Engineering at the University of New South Wales in Sydney, Australia. His current research interests are in the areas of sensor networks, mobile networking, vehicular communication, wireless mesh networks and network security.

Matthias Hollick is heading the Secure Mobile Networking Lab (SEEMOO) at the Computer Science Department of Technische Universität Darmstadt, Germany. He received his Ph.D. degree in 2004 from the TU Darmstadt. He has been researching and teaching at TU Darmstadt, Universidad Carlos III de Madrid (UC3M), and the University of Illinois at Urbana-Champaign (UIUC). In 2005, for his research, he has received the Adolf-Messer Foundation award. His research focus is on secure and quality-of-service-aware communication for mobile and wireless ad hoc, mesh, and sensor networks.