6

Network Services for Mobile Participatory Sensing

Sasank Reddy, Deborah Estrin, and Mani Srivastava

Abstract

The rapid explosion of mobile phones over the last decade has enabled a new sensing paradigm – participatory sensing – where individuals act as sensors by using their mobile phones for data collection. Participatory sensing relies on the sensing capabilities of mobile phones, many of which have the ability to detect location, capture images and audio, the networking support provided by cellular and WiFi infrastructure, and the spatial and temporal coverage along with interpretive abilities provided by the individuals that carry and operate mobile phones. If successfully coordinated, participants involved in data collection using their mobile phones can open up new possibilities uniquely relevant to the interests of individuals, groups, and communities as they seek to understand the social and physical processes of the world around them. Responsibly realizing a vision of sensing that is widespread and participatory poses critical technology challenges. To support mobile participatory sensing applications, the future Internet architecture must provide network services that enable applications to select, task, and coordinate mobile users based on measures of coverage, capabilities, and participation and performance patterns; attestation mechanisms that enable sensor data consumers to assess trustworthiness of the data they access; and privacy and auditing mechanisms that enable sensor sources to control sharing and disclosure of data.

6.1 Mobile Participatory Sensing Vision

6.1.1 Individuals Carrying Mobile Phones as Sensors

Embedded wireless sensing provides scientists and engineers unique insights into the physical and biological processes of the natural and "built"

https://doi.org/10.1017/CBO9780511921117.007 Published online by Cambridge University Press

environments. Here we consider a shift into the public realm, a move that anticipates sensing's use by the general public and suggests new possibilities for understanding social, political, or, more generally, "urban" processes. In this expanded view, sensing can serve as a technological platform for advocacy – "making a case" through distributed documentation of some need. Or it can be a tool for introspection into the habits and situations of individuals and communities – self-discovery through private or social data analysis. A key distinction between this use of sensing and traditional embedded scientific applications is its reliance on individuals' participation in data collection and analysis. Perhaps more important, however, is an accompanying proliferation of purpose; that is, the applications of publicly deployed sensing actively emerge from the interests of the public. Traditional approaches to networked sensing cannot achieve this because embedding the necessary sensors in real-world environments is too costly, requires broad deployments that are likely to be either aesthetically or politically unacceptable, and ultimately proves to be inflexible in the face of diverse users' needs. As an alternative to this sensing of the public, we consider sensing by the public (Burke et al. 2006; Eisenman et al. 2006; Paulos et al. 2007). We take as our starting point the cellular and WiFi networks that currently support billions of mobile phone users. Most phones are already equipped with acoustic, image, and location sensors – in the form of microphones, cameras, and GPS, WiFi, or cellular positioning – and a Bluetooth interface that can be used to connect external sensors. They also provide text and graphics entry for the manual description of events. These devices can be tools for sensing, and we focus on what it would take to establish their role in a participatory sensing network. In this context, mobile devices are the sources of digital content, network services provide higher-level understanding and organization of personally contributed data streams, and mobile users are active in defining, participating in, and analyzing data from coordinated observing "campaigns." The resulting platform is parsimonious, introducing very little new equipment into the environment and requiring from participants only the data necessary to achieve the impacts they desire, and yet is uniquely able respond to diverse need and interest.

6.1.2 Types of Participatory Sensing

6.1.2.1 Authored versus Ad Hoc Data Collection

In the simplest model of participatory sensing, individual participants gather sensor data about social and environmental processes, publish, and share it in an ad hoc fashion. "On the scene" citizen reporting, like CNN's I-Report, where an individual's serendipity is an asset, is an example that has emerged in

popular culture. The relatively uncoordinated nature of this approach as a sensing paradigm, however, limits its utility for campaigns that have stricter requirements in what, where, and how data should be collected. To address this deficiency, we introduce the notion of "authored" campaigns. In this model, mobile phone-based data gathering is coordinated across a potentially large number of participants over large spans of space and time. Such coordinated sensing could be initiated by individuals, groups, or institutions and might involve dynamic decisions about the data being gathered, the spatial extent and temporal frequency of sampling, and the overall level and character of the participation needed. Network services are necessary to support the critical element of human participation.

6.1.2.2 Opportunistic versus Guided Sensing

In authored data collections, the level of coordination can range from being opportunistic to being guided. In the opportunistic case, participants are involved in an autonomous manner in which the sensing on the mobile phone occurs without the participant's direct involvement. The main goal is to obtain necessary sensor values without putting a burden on the participant, and thus the system infers situations when sensing should occur and activates the appropriate sensing on behalf of the participant. Examples of opportunistic involvement include taking pictures from the camera automatically every twenty seconds while the phone is exposed externally or sampling the microphone when the phone is held. At the opposite end of the spectrum is guided sensing where network services work in coordination with the participant to inform them of specific campaign needs, such as where spatial or temporal gaps exist in the data collection. The system can provide suggestions to participants of sensing needs in the field, as well and incentivize them to fill sensing gaps. For instance, a service can provide a route plan that maximizes sensing utility or inform the user of nearby sensing opportunities as they walk through an area of interest.

6.1.3 Application Space

Participatory sensing enables data-collection campaigns that can make an impact in a wide variety of application spaces including urban planning, environmental monitoring, and cultural exploration. Here we give scenarios of how mobile phone sensing can be used for such "make a case" sensing deployments. These campaigns show the need for network services that account for individuals' geospatial coverage, availability, and reputation for delivering useful campaign data while respecting participant privacy concerns.

6.1.3.1 Truck Traffic Assessment

Our first example is inspired by T. S. Lena and colleagues' work with community documentation of diesel truck traffic in the Hunts Point peninsula of New York's South Bronx, home to a primarily low-income population and a hub in the tri-state freight transportation system (Lena et al. 2002). Consider a community with higher-than-average asthma hospitalizations and deaths, which is concerned about diesel truck exhaust in their neighborhoods, a primary source of airborne particulate matter. Documenting average diesel truck traffic counts through many streets in the neighborhood would create a valuable resources for community members to (1) assess the amount of traffic relative to zoning and regulatory requirements; (2) find unexpected "hot spots" of traffic; (3) coordinate with a university or public health organization to supplement more specialized monitoring; all to (4) generate material necessary to advocate for further study, legislation, or research. In this application, there could be many willing community participants, but with minimal free time and a need to obtain high data credibility. Coordinated, participatory sensing campaigns could be employed by the community and university to best organize their willing and intelligent human resources to capture truck traffic counts through both directed sensing ("please go to this corner and make some recordings") and participatory interaction ("you're already near a place we need data, please take a few photos or enter the number of trucks you see"). This coordination is made more challenging because the time and spatial variations in truck traffic could be initially informed, perhaps under the guidance of the participating university based on existing environmental models, traffic studies, legislation, or other data.

6.1.3.2 Citywide Resource Survey

The second campaign example is inspired by the Getty Conservation Institute's Historic Resource Survey Project [GCI08] and the USC GeoDec (Shahabi et al. 2006) group's "social image mapping" project. The Getty is collaborating with the City of Los Angeles to develop professional survey methodology to document the city's historic resources (primarily buildings). Once this is done, the city will face the challenge of actually implementing the data collection. Even though professional survey is not possible using mobile phones with untrained participants, a citywide participatory project that involves everyone in deciding what is historic and why, and building up a secondary library of media documentation, is very exciting and can be implemented. A coordinated participatory sensing campaign can contribute geo-tagged historic resource images, audio, and other data to augment the Getty's data collection. This enables never-before-possible documentation of our built environment, which involves people

in decision making about what to document, when, and with what thematic keywording, but coordinated by intelligent network services.

6.1.4 Network Services for Coordination, Feedback, and Privacy

Overall, participatory sensing's challenges and promises come from the same characteristics. First, the approach's tremendous potential emerges from leveraging existing technology and infrastructure to actively consider human concerns. It is precisely this reliance on systems designed for other purposes that challenges us to create more human-aware network services to coordinate participation. Second, as coordinated and model-assisted sensing scales up, it requires network management of credibility and reputation on behalf of participants and their self-expressed goals. Finally, given a network with coordination functions, people will participate in sensing things that matter to them, but it is their intimate involvement in the sensing process that must be respected and responsibly designed for, in a secure, flexible, and transparent approach to participation, data control, and privacy regulation.

6.2 Context Inference and Coordination

For participatory sensing campaigns to succeed, network services need to exist that operate continuously on behalf of all involved in a sensing campaign to (1) identify the potential participants who are suited to the goals of the campaign; (2) negotiate with the participants the constraints on their involvement, and incentivize them to participate; (3) opportunistically exploit sensing and data-sharing opportunities that present themselves as people move around; and (4) optimize the sampling coverage while assuring credibility of sensor data and conforming to constraints negotiated with the participants. The network service architecture, shown in Figure 6.1, will embody these functions through the Recruiter, Coordinator, and Guardian modules whose designs will be impacted by various human factors. As their names suggest, the three modules represent network services that respectively select participants for a campaign, coordinate them to perform the sensing task, and monitor their performance throughout the data-collection effort. Their roles in the architecture are described in detail in the next subsection. The system has to select from and manage a diverse population of potential but uncommitted participants with different availability, mobility and activity patterns, history of participation, diligence, predisposition, skills, timeliness, phone capability, and privacy constraints. Further complexity arises because humans are self-willed, intelligent, and creative. These human factors will be captured for each participant through models of context-annotated mobility profiles, reputation, and privacy constraints.

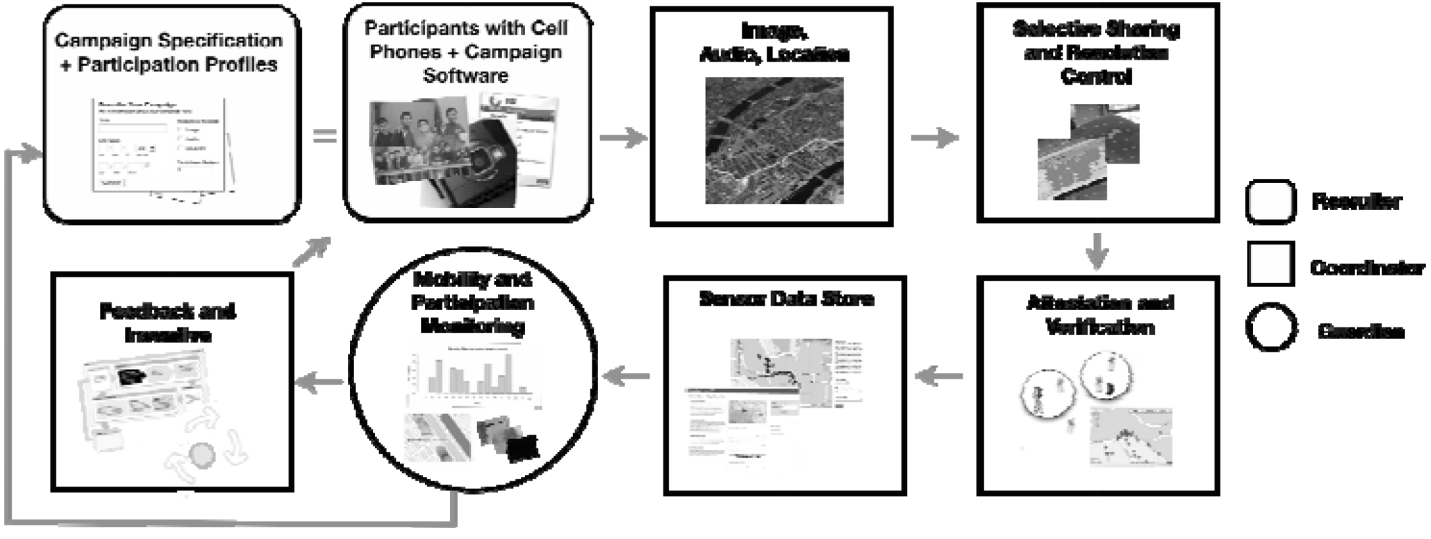

Figure 6.1. Architecture for coordinating participatory sensing data collections.

6.2.1 Architecture Components

6.2.1.1 Recruiter

Like many Web 2.0 applications that rely on contributions from individuals, a campaign seeks concerned participants willing to volunteer their time to help with data collection and analysis. We envision a Recruiter that takes campaign specifications and participant profiles as input and recommends potential participants, much like the friend-finding features of social networking sites. This engine should be useful for the group initiating a campaign as well as those who would like to volunteer, allowing the latter group to judge the feasibility of their participation. The campaign specifications might include the sensing modalities needed, the regions of space and time over which to conduct the campaign, the overall campaign budget (which may consider all human and material resources needed to run the campaign, not just cases where participants are compensated for their time), the demographic diversity of participants, and so on.

Multipart profile information for each potential participant is the other crucial input to the Recruiter. The profile contains information in regards to the incentives needed by the participant, along with interest vectors. Furthermore, participants are evaluated in terms of capabilities, availability, and performance. Capability information captures relevant characteristics of the cell phone carried by that individual, such as the set of sensing modalities and their quality. Availability information would indicate when and where the participant is likely able to gather and contribute data. This would include models of the participant's mobility and activity in space and time, as well as constraints due to privacy rules. Performance information would indicate how this individual performed on previous sensing campaigns in terms of metrics such as quality and timeliness of contributed data, consistency relative to their commitments, and responsiveness to data collection requests. The campaign designer could compare released performance information with what they consider minimum qualifications for their campaign.

The designer would then use participant recommendations to selectively recruit a participant list that achieves the highest utility while adhering to the campaign resource budget. In keeping with a participatory approach, at the time of recruitment, the system could negotiate with the participant a level of commitment to the campaign that would be part of the basis for valuation, incentives, and reputation in the system. At a technical level, the recruitment problem is similar to that of sensor selection in traditional embedded applications like those studied in Ganesan et al. (2004), Krause et al. (2006), and Krause et al. (2008), with the obvious distinction that these papers do not consider direct human involvement at all or only consider certain aspects of it in the measurement

process. In the case of campaign recruitment, recommending participants requires the more difficult task of modeling factors tied to human behavior.

6.2.1.2 Coordinator

The Coordinator orchestrates data collection by remotely tasking and configuring participants' cell phones, keeping owners in the loop of the coordinated sensing process, and attesting the authenticity of collected data through verification techniques that check that samples represent the phenomenon that occurred at that time and space and were taken by a human (as opposed to an automated bot). Furthermore, the verification methods could be used to assign reputation scores to participants involved in the campaign. The Coordinator is supported by software that runs on the mobile phone that facilitates in-situ data collection, remote configuration, and interactive feedback (Burke et al. 2006; Froehlich et al. 2007).

One of the main objectives of the Coordinator is to promote participation. It can do this by a feedback system that is informed by persuasive computing – prompts, an incentive system, and social validation are employed (Fogg 1998). Prompts can be visual and auditory aids that remind participants to take samples. The objective is to deliver these prompts in a simple, clear, and nonobtrusive fashion based on the participants' current location and the sensing uncertainty at that location relative to campaign requirements (Intille 2004). An incentive system based on "credits" that can be redeemed for monetary rewards or for additional capabilities in the campaign system will be used to encourage high-quality participation (Pryor 2002). Credits can be removed from participants as well if they deliberately deliver wrong information (for the purpose of collusion or some other type of self-gain). Finally, the Coordinator may use social validation – the concept that people determine what is correct based on what other people consider is correct – by showing a particular participants contribution level compared to other individuals involved in the campaign system (Cialdini 2001). By providing this relative comparison, we hope to encourage individuals to compete to achieve or keep a high participation level.

The Coordinator is also involved in the verifying whether data contributed by a participant actually took place at a particular time and location and that it was contributed by a human as opposed to an automated program (bot). More details about these verification and attestation mechanisms are given in the next section.

6.2.1.3 Guardian

Working in close connection with the Coordinator module in our architecture is the campaign Guardian module. The Guardian observes overall campaign

performance and how each participant is doing relative to their negotiated commitment and the campaign's needs. It must assess participant activity/mobility patterns, and their sensing performance in terms of quality and utility of their contributions. In soft real time, the Guardian provides campaign status to the Coordinator in case participants join or leave, must be added or removed, or need incentives to collect better data or data at desired locations. Over longer time scales, the Guardian updates the profiles associated with each participant based on data about their performance relative to their commitments and mobility/activity pattern during the campaign. Although performance tracking during the execution of a system is not new, monitoring and assessing data collection by humans, especially for coordination and execution purposes, is something we consider novel and challenging. This updated information provided by the Guardian can then be used by the Recruiter to adapt the participant list based on the current behavior of participants in the campaign.

6.2.2 Multipart Profiles

Multipart profile information for each potential participant is crucial to both recruitment and execution. Capability information captures relevant characteristics of the cell phone carried by that individual, such as the set of sensing modalities and their quality. Availability information would indicate when and where the user is likely able to gather and contribute data. This would include models of user's mobility and activity in space and time, as well as constraints due to privacy rules. Performance information would indicate how this individual performed on previous sensing campaigns in terms of metrics such as quality and timeliness of contributed data, consistency relative to their commitments, and responsiveness to data collection requests. Commitment weights each of these relative to the negotiation made with the participant for each campaign.

The annotated mobility profiles used to model participants' behavior over time and geography could be quite fine-grained, such as whether one is outside, walking, eating, or on the phone with a colleague. In general, annotated mobility profiles are quite difficult and inconvenient to sense, and may even require additional sensor hardware. However, even the coarse-grained macro notions of context that can be gathered with existing infrastructure are of great utility in coordinating participating sensing. Good examples are whether the participant is indoor or outdoors, and mode of transportation, whether one is stationary, walking, running, biking, or traveling in a vehicle. Figure 6.2 shows an example of such traces for an individual over a period of eight days. Different portions of each daily trace are coded in different shades to indicate the inferred activity state of the individual at various locations and times: still, walking, or in a vehicle. The activity state was inferred in real time by software running

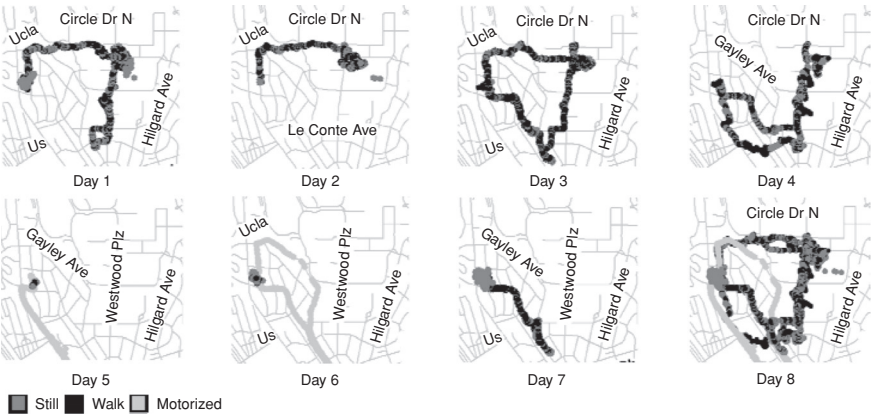

Figure 6.2. Context annotated mobility data for an individual over several days.

the GPS-equipped mobile phone and employing machine-learning algorithms. Moreover, although the traces are quite diverse, they do show several locations and paths that occur frequently, corresponding to repetitive mobility patterns that may be useful in deciding the suitability of a participant for a sensing task.

6.2.2.1 Inferring Context Using Mobile Phones

We envision a system that gathers <location, time, context> traces for participants according to their privacy constraints and then represents it as a compact evolving model used to assess how an individual could contribute to a specific campaign's needs. Location and time can be obtained via GPS embedded in many cell phones or from network infrastructure. Capturing mobility or activity context is more difficult, especially given our goal not to rely on hardware beyond the mobile phone. The challenge lies in identifying the context using sensors likely to be embedded on the cell phones, without requiring its owner to significantly change their habits of use.

Recent work has had good success in identifying significant locations, indoor versus outdoor status, and mode of transportation. For instance, density-based clustering of GPS location points has been employed with distance and time boundaries to divide mobility traces of individuals into locations and routes (Kang et al. 2005; Reddy et al. 2008a; Zhou et al. 2004). Likewise, GPS on the cell phone has been used as an indoor versus outdoor sensor by using a vector of features including the number of satellites available for GPS, geometric dilution of precision, accuracy, and speed variance. Finally, transportation mode of an individual has been inferred using both GPS and accelerometer features. Specifically, coarse-grained transportation mode classification, such as whether

https://doi.org/10.1017/CBO9780511921117.007 Published online by Cambridge University Press

a user is stationary, walking, or in motorized travel, has been inferred by using only GPS data by dividing routes based on changes points (speed close to zero, loss of GPS signal) and then calculating features based on speeds for these segments, such as average, maximum, and minimum speeds, along with distance information (Zheng et al. 2008). Also, fine-grained transportation mode inference, which includes differentiating between running and biking along with other modes, is possible with high accuracy based on analyzing the GPS speed along with accelerometer features such as mean, variance, and certain frequencies of the magnitude force vector (Reddy et al. 2008b).

6.2.2.2 Mobility Profiles for Coverage Assessment

To assess a participant's suitability for the space-time coverage needs of a campaign, the context-annotated mobility information must be represented in a participant's profile in a compact manner, adaptable to variations in their behavior, and efficiently query-able by the Recruiter. Mobility modeling for coordinating participatory sensing requires predicting the statistics of movement patterns at fine granularity over longer period of times than is currently done in cellular networks. It differs from prior work on mobility models for network simulation, which focuses on generating mobility traces (Bai et al. 2003; Hong et al. 2001; Jardosh et al. 2003; Tian et al. 2002). Also, it differs from traffic aids and resource allocation for wireless hand-off systems, which focus on short-term prediction of location (Bhattacharya and Das 1999; Choi and Shin 1998; Hariharan and Toyama 2004; Krumm and Horvitz 2006; Simmons et al. 2006).

Based on both anecdotal evidence and our exploratory data gathering, we know that human mobility has common patterns (Eagle and Pentland 2006), such as repeating routes and frequent locations. But human mobility also exhibits significant temporal jitter on a day-to-day basis. For example, one can imagine taking the same route from home to work but departing at different times in the morning or going to the grocery store every week but during different days of the week. Also, schedules of individuals might change over time, and the mobility model needs to be able to adapt. College students often have dramatic schedule shifts from one semester to another, so the mobility model will need to adapt to these changes quickly. But the updating scheme should also be aware of outliers (vacations, conference visits, etc.). Figure 6.3 shows an example of mobility profiles as they change over time for an individual where both a natural variation and dramatic shift is shown. Natural variations occur as the individual visits different stores, restaurants, and so forth, whereas the dramatic shift occurs due to an event such as change in place of work or residence. The algorithms and models used for mobility profiling must capture the natural variations in mobility patterns and also be able to detect dramatic shifts to allow trigger retraining.

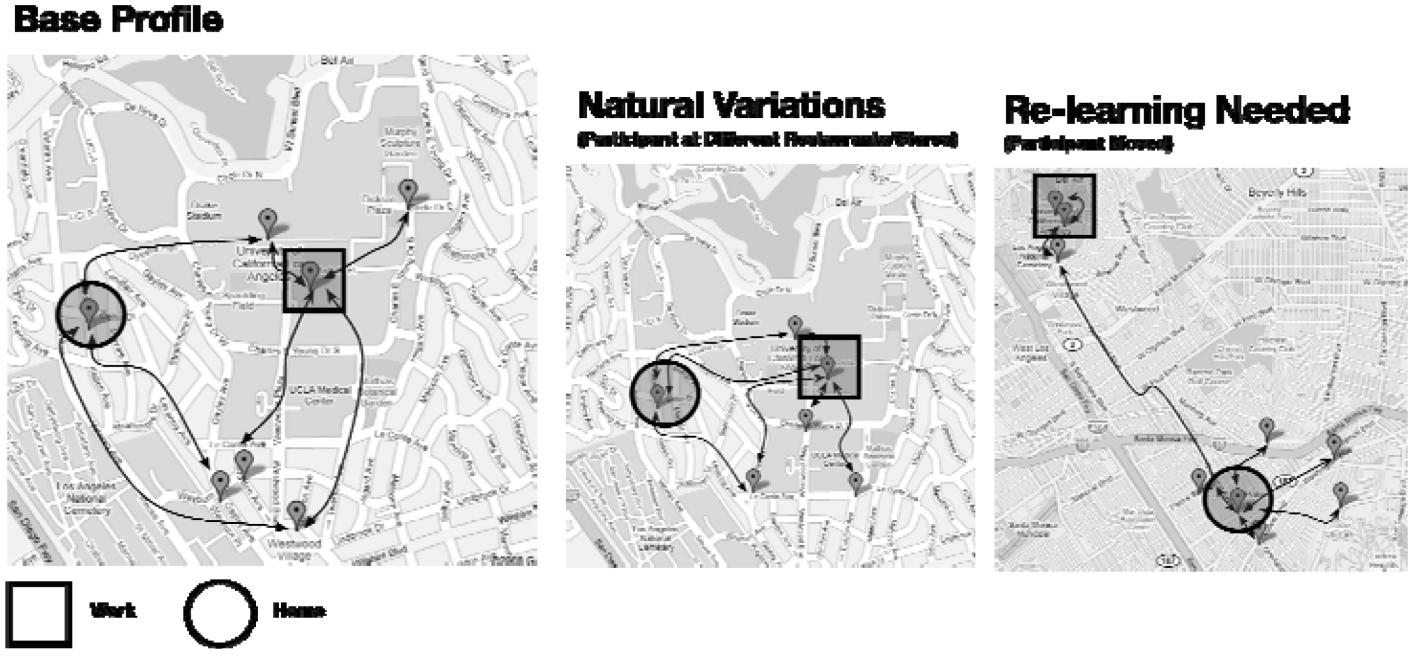

Figure 6.3. Changes in mobility profile for several weeks for an individual.

Overall, the models will require a mix of data mining to identify and cluster patterns and statistical representation to identify the patterns and use them as higher-level building blocks for expressing the overall mobility behavior. One such technique organizes mobility information into a profile that consists of an "association matrix" that captures the amount of time spent in a particular context during a time period. This association matrix is used to infer which individuals would be the "best fit" for coverage. Individuals' profiles can be compared over a time period for consistency by performing Singular Value Decomposition to obtain the eigenbehaviors (main column signatures in the association matrix) and then comparing consecutive time periods of eigenbehaviors by calculating similarity (Eagle and Pentland 2006; Gonzalez et al. 2008; Reddy et al. 2008a).

6.3 Data Attestation and Credibility

Participatory sensing systems must establish credibility of collected information, considering that contributions come from participants with varying skill, intent, and understanding of the campaign's needs. They must do this while allowing participants to regulate their own privacy and participation. For example, in some cases, participants may be anonymous from either the perspective of the human campaign organizers or even the system itself. To increase confidence in contributed data, the system could verify samples as (1) taken at a particular location and time, (2) capturing the phenomenon of interest that occurred, and (3) contributed by an authorized participant and not an automated bot. Like other security measures, such verification results will not be absolute, and we thus consider them as input to a participant's reputation. Even with a tamper-proof trusted platform running trusted software (Aissi et al. 2004), there is sufficient variation in each measure that verification would not be binary. We focus on the compelling and more immediately scalable scenario, in which participants can use available mobile platforms. In the following section, we describe each verification task in more detail and suggest mechanisms to achieve them.

6.3.1 Verifying Participant Context (Location)

High confidence in where and when samples were taken increases their credibility. Providing direct assurance of location and time of capture is very difficult without having trusted platform components in the mobile device (Aissi et al. 2004). We approach this problem in a simpler but more immediately practical sense by verifying the location and time of the contributor. Thus, when a sample is uploaded, the system is at least able to verify that the sample was taken at a location and a time at or before the upload time. We propose creating a location and time attestation service that a participant's mobile device can query. A related approach has been presented in (Lenders et al. 2008). Location attestation can be implemented using trusted infrastructure or with location fingerprints. In

the first case, a trusted infrastructure exists, such as a WiFi network, which can be queried to attest location and time. Essentially, the participant's device sends its location and time according to its own measure, obtained either through GPS or the cellular network, to the verification service. The service compares the submitted location and time with its network-observed location using wireless signal strength and triangulation techniques (Letchner et al. 2005) and time, and sends back a certificate if it matches. This certificate can be then uploaded by the participant when data is submitted. The second method does not rely on infrastructure but instead works by having the participant's device advertise and listen for WiFi and Bluetooth beacons. It then uploads the identifiers of scanned beacons along with its location obtained through GPS or the cellular network to the verification service. As regulated by the personal privacy decisions of participants, the verification service maintains a database of all devices, their time, location, and fingerprints defined by the set of devices they see. It uses this database to compare and verify the locations of participants who enable this feature to increase the credibility of their data.

6.3.2 Verifying Validity of Sampled Information

The system could also assess whether contributed data represents an occurrence of the phenomenon of interest. For instance, in our second campaign example, if a participant contributes resource images from another city, creates a digitally altered image, or includes an image from the past as if it was the present – how can we identify such misrepresentations? To address this challenge, the system could coordinate other participants in the campaign to cross-check the validity of each other's contributions. For instance, if a person is located within a certain area near where a sample was previously contributed, the system would request them to also take a sample at that location as well. This same request could be issued to other participants involved in the campaign and used to validate the contribution. For cases where the phenomenon exhibits dynamic behavior over time, constructing such opportunistic verification requests becomes more challenging but still important given the value of additional samples in building a model that establishes the validity of the documentation.

6.3.3 Verifying Human Contributions

Verifying that a human is in the loop during data collection is crucial to the respectability of participatory systems and to the creation of an equitable and relevant reputation framework. When reputation is meant to reflect some measure of engagement, the system is open to attack by automated "spam" processes that simulate participation. To counter this, we propose an in situ challenge-response system (Naor 1997). While a participant is sampling and uploading data, the Coordinator provides a challenge that, with high probability, only a human, and

not a computer process, can respond to in a timely manner. Specifically, we propose the use of a sensing CAPTCHA, the Completely Automated Public Turing Test to Tell Computers and Humans Apart, introduced by Ahn (2003). In participatory sensing, we can request tagging or annotation of other participants' samples as a challenge, and thus have the added value of "crowd-sourced" data verification as well (Gentry et al. 2005). Similarly, the challenge could use a randomized sequence of data captures done under a small set of known conditions (e.g., camera orientations, newspaper inserted into the image, simultaneous playing of audio signals that are picked up during audio capture, etc.). If the tags or data series match, then the system confirms that the participant collected the sample at the claimed place and time. In addition, a participant can be asked to rank the quality of people's samples, and in this way the system is able to score data. Finally, by sending the same challenge to several participants, the system can more reliably score and verify contributions (Chew and Tygar 2005).

6.3.4 Reputation Measure for Contributors

Beyond providing a measure of confidence in a given sample, the verification mechanisms described previously could also contribute to a measure of overall reputation for campaign participants, some of which could leverage trusted information, such as that from augmented handset hardware, without directly revealing it to the campaign organizers. The system can keep a reputation score associated with a participant based on how that participant performs as a data contributor relative to their commitments and campaign needs. Tracking user reputation is not new – in fact, it has been widely employed on the Internet to track whether to trust a user for transactions (Ebay) or in providing good input to a system (Yahoo Answers, Amazon MTurk) (Jøsang et al. 2007; Resnick et al. 2000; Resnick and Zeckhauser 2002). In the default operation of our system, providing verified data will result in a high reputation, whereas contributing samples suspected as invalid or contrived will result in a lower reputation score. The reputation score of the individual can also be associated with their sample to give data users a sense of the system's assessment of sample credibility to consider when they employ shared data. To achieve this, we face challenges that include how necessary privacy mechanisms affect verification and reputation calculations. For example, reputation may not be a scalar value but instead be a vector of performance and participation assessment metrics. Whereas performance reputation may be affected by privacy mechanisms lowering the verifiability of samples (blurring or adding noise to data, say), participation reputation would not be affected by these controls. Additionally, reputation mechanisms should be customizable by the campaign creator, given their knowledge of what success, reliability, and credibility mean for their campaign and its participants.

An additional challenge for maintaining reputation scores comes from the idea of identity. In participatory sensing, we envision a range of identity options

for campaign designers and participants: For some, there is strong protection of anonymity even from the managing system; for others, pseudonyms are created for particular campaigns but a consistent identity is known by the system; and for still others, authentication is based on time spent in a location, or device identity is decoupled from user identity. Mapping reputation services to these identities is challenging. It is tempting to follow the lead of many Internet services and reward users for maintaining a single, trackable identity across campaigns. Certainly this makes reputation management more immediately effective and its corresponding credibility metrics easier to understand. But this may be counterproductive or unnecessary in many cases where other forces exist "offline" from the system to regulate use – for example, through preexisting social mechanisms for authenticating participants. In these cases, intracampaign reputation based on community norms will be sufficient. We believe that participatory sensing systems can be created with sufficiently configurability to address this variety, as long as the network services do not fix a single concept of identity.

6.4 Privacy

Privacy is a long-standing topic in mobile computing, especially with respect to the delivery of location-based services. In mobile participatory sensing, privacy becomes a first-order challenge because the sensing is enabled as a fine-grained resolution and is directly associated with the individuals performing the sampling. For instance, the data collected can be used to quantify habits, routines, associations, and the data (especially location traces) is easy to mine to obtain this personnel information. Furthermore, there are a host of negative impacts (location-based discrimination, safety and security threats) that could result if privacy is not considered seriously. The approach that the research community has taken to tackle this challenging issue has focused on two fronts: creating design principles that network services must meet to help balance data collection and privacy concerns of participating individuals; and designing system architectures that support core data services for privacy, such as audit of sampled data and the ability to enable filtering and resampling of information for sharing purposes.

6.4.1 Privacy Principles

There are many software architectures emerging to support participatory sensing in a privacy-preserving fashion. These systems typically focus their design around three underlying principles, as outlined by Shilton et al. (2009): participant primacy, data legibility, and longitudinal engagement. Participant primacy is the concept that participants should own the data they collect and have the ultimate control on how the data is used, whom the data is shared, and how long the data is retained. Services need to exist to support these privacy-based

data-collection decisions. Data legibility encourages systems to create visualizations so that participants can make sound decisions about their privacy needs – specifically providing intuitive interfaces to export processing, sharing, and retention details by components interacting with the collected data. Longitudinal engagement is the ideal that systems should strive to keep participants involved in the complete data-collection cycle, from initial sampling to processing, usage, and deletion, encouraging them to be active privacy stewards of their data.

6.4.1.1 Participant Primacy

Participants should have the ultimate control over how their data is used. For this to occur, however, certain tools need to exist in regards to data transformation, storage, and access control to help participants make sound privacy-enhanced decisions (Caceres et al. 2009; Hong and Landay 2004; Shilton et al. 2009). Transformation deals with how the data should be presented to different data sinks. For instance, in the case of a location stream, the sampling frequency of provided data could be changed (instead of providing a service with a location update every second, the sampling rate provided can be changed to five minutes), or the resolution could be adjusted (the level of uncertainty could be adjusted from a few feet to miles, if necessary). Currently, many mobile systems involve data simply being sent to an end-point service and no intermediate storage. By having a tool that enables the backup of all information sent, both audits and future dissemination is possible. An end-point service might perform a certain type of inference on data, but unless the participant has access to the raw data sent initially, there is no way to check whether the inference is being performed correctly or even adheres to the statement of service. Furthermore, the data collected for one particular data collection campaign might be useful to another in the future. By having a backed-up copy, the data is available for future use. Finally, by incorporating access control as part of a toolset for data collectors, participants have fine-grained control over who gets to access their data, how long it should be retained, and whether access should be revoked. The usefulness of this access control mechanism can be seen especially when unwanted information that gets uploaded to a service. Without having access control services in place, there would be no way for the user to revoke rights to the mistakenly uploaded data.

6.4.1.2 Data Legibility

Privacy is a negotiation between participants and sensing organizers of what information to share or withhold. But in order for individuals to make sound decisions about their sharing policies, system legibility is key. They must

understand who is asking for the data (identity); what the data will reveal about them (granularity); what the organizer wants to use the data for (purpose); and how long data will be retained by a requesting organizer (retention) (Reddy et al. 2008a). The system must communicate to participants the nature of the processing that is occurring with their sampled data, along with whom the data is shared with and for how long. Communication between the system and the participant is essential to the system's legibility: the ways in which a system enables people of all technical backgrounds to make informed disclosure decisions. Concerns of visibility and accessibility lead to considerations of data visualization and interpretation in relation to privacy. Work has emerged that faces the challenges associated with visualizing information obtained through data collection. Tools exist to enable "mashups" of collected data on a geo-spatial frame. The idea or providing a platform to enable "social data analysis" and "casual visualization" is becoming important (Wattenberg et al. 2007). To this end, these same visualization techniques are being employed to enhance system legibility (Mun et al. 2009).

6.4.1.3 Longitudinal Engagement

A key ideal that must exist with privacy tools for participatory sensing data collectors is continued engagement. Specifically, privacy should not be a one-time operation that occurs when data collection is first initiated, but instead should be an ongoing engagement with the participant throughout the data-collection life-cycle. Systems should be designed so that participants are reminded of their privacy settings and actively confirm that their settings are still valid. This can come in the form of regularly scheduled reminders that require active acknowledgment. For instance, FireEagle, a location-sharing service, sends out an email alert every three months to encourage participants to check their privacy settings (FireEagle 2009). This feedback can be simply alerts, as FireEagle does currently, or more detailed summaries of privacy policies that are in place. Furthermore, if changes on how data are being used occur, feedback should be given to participants so that they can make a sound, timely decision on whether their data should continue to be exported or if changes need to be made to the resolution or retention policies associated with the data.

6.4.2 Personal Data Vault

Several research groups have been working on experimental architectures to enable the privacy principles to be enacted in implementation form. Most of these architectures revolve around the idea of a personal data vault (PDV) that acts as a proxy for the participant in interacting with various applications that can use the data being collected. This PDV is designed to have a number of

intelligent features including the ability to perform backup and republishing, enable access control and auditing of data usage, and perform adaptive filters on the actual data in terms of resolution control and sampling frequency.

6.4.2.1 Storage and Republishing

Similar to existing backup services that exist for personal computers, one essential task that a PDV can perform is the backup of all data sent to it. Furthermore, since backups exist, the PDV can also act as an intelligent republishing tool so that information can be disseminated to other applications both in real time and in a delayed fashion. The storage of data is important because the ecosystem for end-point services is changing rapidly, and a service that is popular today might not be the one used in the future (Caceres et al. 2009). Easily being able to take existing data that was backed up and sent to another service at a future time is invaluable. Furthermore, the republishing strategy helps with efficiency on the phone. If a participant is running multiple data-collection campaigns, then it is more energy-efficient to send the data to a PDV and have it republish to other services on behalf of the participant (Caceres et al. 2009).

6.4.2.2 Access Control of Data Usage and Audit Trails

Another important feature of the PDV is the ability to perform access control and audit the usage of data. Access control works similar to current systems that exist on the desktop for sharing files with various individuals and groups. But the PDV would also incorporate features specifically available on the mobile phone, such as context information. Participants might set access control policies for data based on location (data is shared only in certain zip codes or other spatial regions), time (information is collected during certain parts of the day only), and activity (data is collected only if the user is performing a certain transportation mode) (Shilton et al. 2009). In addition to this access control mechanism, the PDV would also incorporate a trace audit tool that records access, use, inference, and manipulation of data by corresponding end-point services. Having this ability to audit data usage would require external services to log transformations and sharing back to the PDV, and a signing system could exist to verify that certain services perform their advertised tasks (Shilton et al. 2009).

6.4.3 Resolution Control and Resampling

In accordance with the ideal of participant primacy, the PDV should provide tools to enable participants the ability to have fine-grained control of the resolution and sampling exposed to external services. Although this can apply to any modality, one in which this is especially important is location data. Instead of streaming

the raw location field at the finest-grained level to external applications, the participant can instruct the PDV to share certain resolution levels to particular end points. For instance, the resolution can range from having a resolution of a few meters to one that is at a zip code or city level (Parker et al. 2006). In addition to resolution control, the PDV can also be used to change the sampling rate associated with sensor readings. Even if the mobile phone is publishing information at a very high rate (i.e., location updates every second), the PDV can instead share this data at a lower sample rate (every five minutes, one hour, or more) (Caceres et al. 2009; Parker et al. 2006). Furthermore, the PDV can delay when the data is actually sent to external applications as well, enabling a "lag" between when data is collected and when it can be used for external inference. This lag can also be used as a buffer for participants to make sharing decisions.

6.5 Implications for the Future Internet

The creation over the past decade of unanticipated applications of the Internet, such as web services, peer-to-peer (P2P) file sharing, networked gaming, IP telephony, and mobile application, has motivated researchers to revisit the core Internet infrastructure and the original architecture choices. The emerging class of mobile participatory sensing applications described in this chapter carries similarly significant implications for the Internet. While many prototypes of mobile participatory sensing applications are being realized over the current Internet infrastructure, experience also suggests that for these applications to scale, certain essential services will need to be incorporated in the fabric of the Internet.

The primary impact of mobile participatory sensing applications on the Internet architecture is not at the lower-layer protocols for routing, transport, and so on. Rather, these applications motivate the need for the network to provide primitives for privacy-aware sharing of personal sensory data, and for handling of certain physical context as a first-class entity.

Sharing of personal sensory data poses conflicting demands from producers and consumers. To the former, the network has to provide control over the quality of information disclosed to different consumers. To the latter, the network has to provide information attributes permitting its quality, provenance, and overall trustworthiness to be assessed. Doing these would require automated and cryptographically secure components in the network.

The handling of physical context as a first-class entity by the network would be limited to contextual information that has universal use, namely location, direction, and speed. Beside the need to formalize representation and dissemination of such information, the challenge is in ensuring that the information is trustworthy and that the client is also provided with an assessment of its quality.

Most techniques to estimate physical context are prone to cheating and adversarial manipulation, and network can take proactive measures to verify context information.

We anticipate the emergence of specialized mediator entities providing these context-handling and data-sharing services as becoming an integral part of the network fabric.

6.6 Conclusions

The challenge and the promise of participatory sensing come from the same characteristics. First, the systems' tremendous potential emerges from the use of existing mobile phone technology and cellular wireless infrastructure. But it is precisely this reliance on systems designed for other purposes that challenges us to create network services to coordinate participation. Second, this coordinated sensing scales down as well as up. It can bring value to even a few people, but increases in accuracy, scope, and worth as more participate – as long as credibility and reputation can be managed. Finally, having technology and coordination, people will participate in top-down, bottom-up, and personally reflective sensing about things that matter to them, but it is their intimate involvement in the sensing process that must be respected and responsibly designed for in a secure, flexible, and transparent approach to participation, data control, and privacy regulation. The future Internet can support such applications at large scale by incorporating as an integral part of its fabric certain critical services, such as sharing of data streams while ensuring trustworthiness and respecting privacy, and first-class handling of verifiable contextual information.

6.7 Acknowledgments

The work described in this chapter is part of a collective effort by several members of the Center for Embedded Networked Sensing, including Jeff Burke, Mark Hansen, Min Mun, Vids Samanta, and Katie Shilton. This research is funded by the NSF under grant CNS-0627084 and by the Center for Embedded Networked Sensing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

References

Ahn, L. V., Blum, M., Hopper, N. J., and Langford, J. 2003. CAPTCHA: Using Hard AI Problems for Security. *Advances in Cryptology – Eurocrypt*, pages 294–311.

Aissi, S., Maruyama, H., Miura, F., Nakamura, T., Saito, D., Takeshita, A., Wheeler, D. and Yoshihama, S. 2004. Trusted Mobile Platform Protocol Specification. *OASIS*. http://xml.coverpages.org/TMP-ProtocolV10.pdf

Bai, F., Narayanan, S., and Helmy, A. 2003. IMPORTANT: A Framework to Systematically Analyze the Impact of Mobility on Performance of Routing Protocols for Ad hoc Networks. *INFOCOM*.

Bhattacharya, A., and Das, S. K. 1999. LeZi-Update: An Information-Theoretic Approach to Track Mobile Users in PCS Networks. *Mobile Computing and Networking (Mobicom)*.

Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., and Srivastava, M. B. 2006. Participatory Sensing. ACM SenSys Workshop on World-Sensor-Web (WSW'2006).

Caceres, R., Cox, L., Lim, H., Shakimov, A., and Varshavsky, A. 2009. Virtual Individual Servers as Privacy-Preserving Proxies for Mobile Devices. *ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld)*, pages 37–42.

Cialdini, R. 2001. *Influence: Science and Practice*. Allyn & Bacon.

Chew, M., and Tygar, J. 2005. Collaborative Filtering CAPTCHAs. *Proceedings of the Second International Workshop in Human Interactive Proofs*, pages 66–81.

Choi, S., and Shin, K. G. 1998. Predictive and Adaptive Bandwidth Reservation for Hand-Offs in QoS-Sensitive Cellular Networks. *SIGCOMM Comput. Commun. Rev.* 28(4).

Eagle, N., and Pentland, A. 2006. Reality Mining: Sensing Complex Social Systems. *Personal Ubiquitous Comput.*, pp. 255–268.

Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Ahn, G. S., and Campbell, A. T. 2006. MetroSense Project: People-Centric Sensing at Scale, *ACM SenSys Workshop on World-Sensor-Web (WSW'2006)*.

FireEagle. 2009. Yahoo, http://fireeagle.com.

Fogg, B. J. 1998. Persuasive Computer: Perspectives and Research Directions. *Conference on Human Factors in Computing Systems (SIGCHI)*.

Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., and Landay, J. A. 2007. MyExperience: A System for In Situ Tracing and Capturing of User Feedback on Mobile Phones. *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services (Mobisys)*.

Ganesan, D., Cristescu, R., and Beferull-Lozano, B. 2004. Power-Efficient Sensor Placement and Transmission Structure for Data Gathering under Distortion Constraints. *Information Processing in Sensor Networks, Third International Symposium*, pages 142–150.

Gentry, C., Ramzan, Z., and Stubblebine, S. 2005. Secure Distributed Human Computation. *Proceedings of the 6th ACM Conference on Electronic Commerce*, pages 155–164.

Getty Conservation Institute. 2008. *Los Angeles Historic Resource Survey*, http://www.getty.edu/conservation/field_projects/lasurvey/

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A. L. 2008. Understanding Individual Human Mobility Patterns. *Nature*, 453, 779–782.

Hariharan, R., and Toyama, K. 2004. Project Lachesis: Parsing and Modeling Location Histories. *Geographic Information Science*.

Hong, J. I., and Landay, J. A. 2004. An Architecture for Privacy-Sensitive Ubiquitous Computing. *Conference on Mobile Systems, Applications, and Services (Mobisys)*.

Hong, X., Kwon, T. J., Gerla, M., Gu, D. L., and Pei, G. 2001. A Mobility Framework for Ad Hoc Wireless Networks. *Proceedings of the Second International Conference on Mobile Data Management (MDM),* pages 185–196.

Intille, S. S. 2004. A New Research Challenge: Persuasive Technology to Motivate Healthy Aging. *Information Technology in Biomedicine, IEEE Transactions*, 8(3), 235–237.

Jardosh, A., Belding-Rover, E. M., Almeroth, K. C., Suri, S., 2003. Towards Realistic Mobility Models for Mobile Ad hoc Networks. *Proceedings of Mobile Computing and Networking (Mobicom),* pages 217–229.

Jøsang, A., Ismail, R., and Boyd, C. 2007. A Survey of Trust and Reputation Systems for Online Service Provision. *Decis. Support Systems*, 43(2), 618–644.

Kang, J., Welbourne, W., Stewart, B., and Borriello, G. 2005. Extracting Places from Traces of Locations. *Mobile Computing and Communications Review*.

Krause, A., Guestrin, C., Gupta, A., and Kleinberg, J. 2006. Near-Optimal Sensor Placements: Maximizing Information while Minimizing Communication Cost. *Information Processing in Sensor Networks*.

Krause, A., Horvitz, E., Kansal, A., and Zhao, F. 2008. Toward Community Sensing. *Proc. of Information Processing in Sensor Networks (IPSN)*.

Krumm, J., and Horvitz, E., 2006. Predestination: Inferring Destinations from Partial Trajectories. *Proceedings of Ubiquitous Computing (Ubicomp),* pages 243–260.

Lena, T. S., Ochieng, V., Carter, M., Holguin-Veras, J., and Kinney, P. L. 2002. Elemental Carbon and PM2.5 in an Urban Community Heavily Impacted by Diesel Truck Traffic. *Environmental Health Perspectives*, 110(10), 1009–1015.

Lenders, V., Koukoumidis, E., Zhang, P., and Martonosi, M. 2008. Location-Based Trust for Mobile User-Generated Contents: Applications, Challenges and Implementations. *Proceedings of the 9th IEEE Workshop on Mobile Computing Systems and Applications (HotMobile 2008)*, pages 60–64.

Letchner, J., Fox, D., and LaMarca, A. 2005. Large-Scale Localization from Wireless Signal Strength. *Proc. of the National Conference on Artificial Intelligence (AAAI)*, pages 15–20.

Mun, M., Reddy, S., Shilton, K., Yau, N., Boda, P., Burke, J., Estrin, D., Hansen, M, Howard, E., and West, R. 2009. PEIR, the Personal Environmental Impact Report, as a Platform for Participatory Sensing Systems Research. *Conference on Mobile Systems, Applications and Services (Mobisys)*, pages 55–68.

Naor, M. 1997. Verification of a Human in the Loop or Identification via the Turing Test. Unpublished Manuscript. http://www.wisdom.weizmann.ac.il/∼naor/PAPERS/human.ps

Parker, A., Reddy, S., Schmid, T., Chang, K., Saurabh, G., Srivastava, M., Hansen, M., Burke, J., Estrin, D., Allman, M., and Paxon, V. 2006. Network System Challenges in Selective Sharing and Verification for Personal, Social, and Urban-Scale Sensing Applications. *IEEE Workshop on Mobile Computing Systems and Applications (HotMobile)*, pages 37–42.

Paulos, E., Honicky, R., and Goodman, E. 2007. "Sensing Atmosphere," Workshop on Sensing on Everyday Mobile Phones. *ACM Conference on Embedded Networked Sensor Systems (SenSys 2007)*.

Pryor, K. 2002. Don't Shoot the Dog!: The New Art of Teaching and Training. *Interpet*.

Reddy S., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. 2008a. Determining Transportation Modes on Mobile Devices. *IEEE International Symposium on Wearable Computing (ISWC)*.

Reddy S., Shilton, K., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. 2008b. Using Context Annotated Mobility Profiles to Recruit Data Collectors in Participatory Sensing. *International Symposium on Location and Context Awareness (LoCA)*.

Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. 2000. Reputation Systems. *Communications of the ACM*, 43(12), 45–48.

Resnick, P., and Zeckhauser, R. 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay Reputation System. *The Economics of the Internet and E-Commerce*, 11, 127–157.

Shahabi, C., Yao-Yi Chiang, Chung, K., Kai-Chen Huang, Khoshgozaran-Haghighi, J., Knoblock, C., Sung Chun Lee, Neumann, U., Nevatia, R., Rihan, A., Thakkar, S., and You, S. 2006. Geodec: Enabling Geospatial Decision Making. *Multimedia and Expo, 2006 IEEE International Conference*, pages 93–96.

Shilton, K., Burke, J., Estrin, D., Hansen, M., Govindan, R., and Kang, J. 2009. Designing the Personal Data Stream: Enabling Participatory Privacy in Mobile Personal Sensing. *Conference on Communication, Information and Internet Policy (TPRC)*.

Simmons, R., Browning, B., Yilu Zhang, and Sadekar, V. 2006. Learning to Predict Driver Route and Destination Intent. *Intelligent Transportation Systems Conference (ITSC)*, pages 127–132.

Tian, J., Hahner, J., Becker, C., Stepanov, I., Rothermel, K. 2002. Graph-based Mobility Model for Mobile Ad Hoc Network Simulation. *Proceedings of the 35th Annual Simulation Symposium,* pages 337–344.

Wattenberg, M., J. Kriss, J., and McKeon, M. 2007. ManyEyes: A Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics*.

Zheng, Y., Liu, L., Wang, L., and Xie, X., 2008. Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web. *ACM WWW Conference*.

Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., and Terveen, L. 2004. Discovering Personal Gazetters: An Interactive Clustering Approach. *ACM International Conference on Advances in Geographic Information Systems (GIS)*.