Chapter in Opportunistic Mobile Social Networks,
Taylor and Francis, CRC Press, 2014, pp. 339-376.

Chapter 1

Incentivizing Participatory Sensing via Auction Mechanisms

Buster O. Holzbauer

Rensselaer Polytechnic Institute

Boleslaw K. Szymanski

Rensselaer Polytechnic Institute

Eyuphan Bulut

Cisco Systems

CONTENTS

With the increasing availability of smart phones and other mobile devices with sensing capabilities, a class of problems known as participatory sensing is becoming increasingly popular. Participatory sensing refers to a sensing system in which humans voluntarily participate in the system, actively contributing to sensing, either by passively carrying devices or actively engaging in sensing. With respect to opportunistic networks, humans provide mobility and decision-making difficult to implement in dedicated sensor networks. Participatory sensing is also a more scalable solution since the costs are far lower than the costs associated with deploying a dedicated larger or computationally intensive sensor network. However, by involving humans, many issues must be addressed in designing a participatory sensing system. In this chapter we discuss several issues in participatory sensing system design, then introduce several papers on existing systems and study lessons that can be applied towards designing a new system. Finally, we conclude by presenting our own participatory sensing system and results from its simulations.[1]

1.1 Introduction

Over the past decade, our ability to gather information about our world has drastically improved. Technology has allowed for cheaper sensors, better communication, hardware that can be powered longer, and increasingly mobile sensor networks [1]. This has led to a type of sensor network applications often referred to as participatory sensing. Specifically, participatory sensing refers to a sensing system in which humans are carriers for sensing platforms known as "nodes" and voluntarily participate in the system. Various definitions with varying specifics exist in literature, but the general consensus is that the problem involves humans directly contributing to sensing, either by passively carrying devices or actively engaging in sensing.

In this chapter we discuss participatory sensing by first introducing its definitions in Section 1.2, then identifying some of the challenges that designers of such systems face in Section 1.3. Since the primary concern in running such systems is maintaining participation, we introduce economics concepts to help formalize the idea of incentives for rewarding long-term participation in Section 1.4. Having introduced foundations and challenges, we shift our focus in Section 1.5 to privacy, which is an important issue in participatory sensing. We then describe two more participatory sensing systems, one of which is our own system, in Section 1.6. Finally we end the chapter with concluding remarks in Section 1.7.

1.2 Problem Definition

In order to monitor the environment, sensors have been used in a variety of situations ranging from static deployments such as personal weather monitoring to mobile swarms of nodes designed to locate and track phenomena [2, 3, 4]. In sensor networks, range and lifetime of systems are limited by available power. In traditional mobile networks, sensors must dedicate some of their limited energy to movement, which detracts from the amount of energy they can use to sense, process, and communicate. By using mobility of living organisms, such as animals in ZebraNet [5], device energy does not have to be used for movement. Carriers will always go to areas of their interests, whether or not the application is suited to their lives. In ZebraNet, where the goal is to track a zebra population, the application is inherent to mobility of the carriers, which are the zebras themselves.

As mentioned previously, participatory sensing can be viewed as the problem of using voluntarily contributions from humans and their devices to collect measurements about a particular phenomenon. More discussion on how a system can leverage human involvement follows in Section 1.3.

1.3 Issues In Participatory Sensing

Participatory sensing is not without its challenges. Since it is still a type of sensor network, problems relating to hardware, communication, and application design found in traditional sensor networks still apply. Mobility is achieved through human movement [6], but integrating humans into the system introduces several new types of challenges as well. Before examining a paper on one of these challenges, we outline some of the key issues and their impact on designing participatory sensing systems.

1.3.1 Data

The purpose of a participatory sensing system is to perform some variety of sensing task. One of the most obvious design decisions is determining what measurements are required to achieve the system's goal, and what hardware is required to support these measurements. Beyond the basic type of measurements, additional requirements may be needed such as resolution of data, how often samples are provided, geographic coverage, etc.

Once data has been acquired it must be put to use, or stored so it can be later used. Ignoring privacy for the moment, a design decision still needs to be made about if the system will have a central repository, if data only exists on nodes, or if a third-party entity owns the data. Additionally, designers must anticipate what kinds of queries and reporting should be run on data, and ensure that these extractions are facilitated by the system's information flow and processing capabilities.

As an example, a myopic design for wildlife detection might be a system that runs a vision algorithm on collected images, and reports only if a particular species is in the image or not. Later, a query of interest might be "What is the population distribution of a particular species across the area of interest?" This question cannot be answered with the data stored (a binary 'yes' or 'no'), but could have been answered if the reports extracted from the vision algorithm had included a count of features corresponding to wildlife. Once a system is deployed it can be difficult to change design details. It may be difficult to communicate with users, access user-owned nodes and push updates. Additionally, doing so may incur costs in the form of inconvenience to users.

1.3.2 Coordination

An important decision, independent of the types of measurements taken, is who takes samples and how often. Assuming that privacy is not an issue, there are still several factors to consider. A simple sensor network deployment design is to have static sensors that sense and report on a fixed schedule. Without incurring communication overhead or adding a central controller, such a system cannot react to dynamic changes or events that cannot be perceived without a global view. If the system is synchronous, and run by one or more central controllers which dictate which nodes take measurements and which types of samples to take, these issues can be avoided. However, this “client-server” model adds communication overhead, and depending on the network conditions, can result in significant lag. Either the controller risks running on an outdated view of the sensors, or sensors may spend time idling while states are updated and while waiting for instructions from the controller.

Alternatively, decisions could be made by the human operators. To continue with the aforementioned wildlife detection example, in a synchronous system the controller might decide to request samples when nodes were known to be close to local areas of interest, or areas where no information was recorded. Without such a controller, the nodes might resort to simply taking an image periodically and submitting a measurement. When considering resource usage, it might be better if users took a picture when they spotted wildlife or were at a location known to require addition sampling. Such a user-driven approach is one alternative to having a central coordinator. Upon examination, the idea of users knowing which locations need additional data came up. This ties into the challenge in Section 1.3.1, since seeing something like a report of number of measurements by location is yet another query that might not be anticipated by a designer who only considered the system’s end-goal. Such reporting also brings up software engineering challenges in APIs and usability.

A third option is to use a peer-to-peer, or ad hoc, network. In this case there is little to no structure, and events or queries drive the behavior of the network. A discussion of ad-hoc techniques is outside the scope of this chapter. While in traditional sensor networks work has been done on peer-to-peer setups [7], applying the paradigm of participatory sensing to such networks is an open problem [8].

1.3.3 Privacy And Security

By involving humans in the sensing process, the data that comes out of the system may reflect information about the participants. In general, this information is not part of the goals of the sensing task, but is an issue nonetheless. For example, many sensing tasks involve spatio-temporal context (a time and a place). If the identity of the user is not adequately protected, then access to reports could reveal exactly where someone was at a particular time. This constitutes a breach of privacy, and is undesirable [8].

To solve this issue, a designer might implement security measures to provide authentication and encryption so only authorized users could see records. This could be combined with an anonymization technique to let participants be credited for their contributions, but not directly exposed [6]. Having a secure repository or identity authority adds a point of failure to the system, and depending on the sensitivity of user data may not be sufficient. For example, consider a participatory sensing system in which users detect chemicals known to be byproducts of improvised explosive devices (IEDs). Such devices have been used in areas of unrest. Privacy violations could result in an individual being at risk of harm or suffering social losses.

A problem with anonymizing identifiers is that they are not sufficient to protect against cases where a potential adversary has prior knowledge about its victims. In the above example, if access to the reports was gained, the adversary could simply target locations that corresponded to residences to discourage participation. Here the identity of the individual is not revealed directly through the report's identifier, but the location information is sufficient to cause privacy violation. This example illustrates that in order to provide privacy, attributes other than identification need to be considered. This challenge is addressed in several of the papers that we discuss throughout the chapter. Furthermore, this example illustrates that a solution in which nodes handle real-time queries with no central data repository is still at risk of privacy violations.

One solution to the problem of particular locations being compromising is to allow participants to disable participation in specific locations. However, if participants opt to send in very few measurements which are at locations and times where other participants also report, then they are less unique and thus harder to identify based on attribute analysis. This suggests a trade-off between exposure and privacy, as well as privacy and coverage of a system. Regulating the diversity of measurements ties into the previous issue of coordination since a controller can leverage its knowledge about the current data repository and participant privacy demands to minimize privacy loss when selecting which participants should sense. Note that not every application is well-suited to using a controller to make decisions about measurements.

Privacy is also a concern in anticipating data uses [9, 1]. If bandwidth and storage are not issues, one approach to prevent the system from being too short-sighted about data collection would be to collect as much data on as many types of measurements as possible, and append the richest metadata to all reports. However, the more information provided in reports, the easier it is to leverage prior knowledge and violate user privacy. Thus a balance must be struck between how much information needs to be stored to enable using data, and how limited the information should be to provide privacy protection. An example where privacy invasion may be overlooked is in the case of browser-based e-mail, which may collect a user's location data, but the primary functionality may blind users to the privacy risks [10].

1.3.4 Human Concerns

In participatory sensing, the device which humans use to sense is often their cellular device [8, 11]. This may be a feature phone or a smartphone, possibly with additional sensors interfacing through technologies such as Bluetooth. In all of these cases, the user has everyday uses for their devices, such as phone calls, messaging, and a wide variety of apps. These uses require energy, as does running a participatory sensing system [12]. Thus, supporting the participatory sensing task has a tangible cost. Furthermore, this cost may not be considered a renewable resource depending on the timescale, since there is no guarantee about when a user can charge their device next.

Communication also imposes upon the user. Bandwidth used for participatory sensing may cause performance degradation in other activities the user would normally do on their phones. In addition, the users may have a limited amount of data they are allowed to transmit and receive based on service provider restrictions.

Convincing users to go "off the beaten path" and perform sensing tasks or go places that are outside of their normal behavior is something the system should be designed to do if necessary. As a simple example, a sensing campaign to assess the levels of background noise in a city may require users to go to areas that are considered less desirable for reasons such as safety. While some users may normally traverse these areas, in order to get sufficient sampling the system may need more users to travel through those regions. While human mobility is convenient to use when it coincidences with the needs of a sensing system, designers must be aware of its limitations.

If the sensing task is not passive, then users spend time and effort to contribute to the it. Users may have a feeling that their time has worth, and justifying the use of their time towards participatory sensing should be something designers are prepared to do. In the case of subjective and qualitative metadata added by users, additional resources may have to be spent to verify user input and manage users.

1.3.5 Participants

The salient feature of participatory sensing is the voluntary participation of users. Without users, the system cannot survive. Furthermore, humans can act in a variety of ways that do not benefit the system, such as never contributing, falsifying results, or providing information that does not satisfy requirements. While we do not discuss issues in reviewing data or recruitment of users in-depth, research exists in literature on the topic [13].

Another issue is that since user participation is voluntary, they can stop participating at any time for any reason. Whether motivation is intrinsic (such as users participating in a sensing campaign that will better their community) or extrinsic (such as the system providing compensation for user participation), the system designers should be aware of what is needed to provide and reinforce motivation [11]. Additionally, system designers must be aware that participants can start or stop participating at any time. In Section 1.4 we cover terms and develop ideas about incentive to support the idea of extrinsic motivation to encourage participation.

1.4 Applying Market Mechanisms

Earlier in the chapter, it was established that persistent participation is essential to participatory sensing, and that this required potential users to be motivated to use the system. In Section 1.5, a scheme for participatory privacy regulation will be discussed. One of the advantages of the approach described in that paper is that by having users be involved in design, research, and regulation, they are intrinsically motivated to continue participating. Unfortunately, this kind of involvement is not always a realizable option. Even in cases where users are involved, incentive that is extrinsic to the sensing application can help reinforce participation. When users do not have any personal reason to join a system, or when they avoid a system because of perceived inconvenience incurred through participation, incentive is a useful tool. To discuss incentives, we refer to economic and market theory, which is a well studied subject [14].

In economics, a market is a system with one or more goods. These goods, which are produced by sellers, have a price associated with them. Buyers purchase goods from seller in exchange for currency. In this chapter we will consider the currency to be "incentive" and do not specify whether it is a monetary or otherwise tangible reward, or some sort of intangible reward such as points for ranking on a virtual leaderboard. One way to model participatory sensing with a central controller is to view it as a buyer of goods which are produced when participants perform sensing tasks. This makes participants the seller, meaning they place a price on the imaginary good produced by performing a sensing task. This price may be affected by factors about the human behind it, such as perceived sensitivity or valuation of their time and resources. Much like in markets with tangible goods, multiple producers, and supply exceeding demand, sellers (participants) are forced to compete with each other to try and sell their goods to a buyer. If participation is sufficient, this supply and demand relationship can be met and competition affects prices. Otherwise, the only limiting factor is budget of buyers (sensing applications).

Deciding how to set prices under competition is not a trivial problem, however before addressing this issue we examine another problem. How sellers and buyers interact must be defined to have any idea what sorts of strategies a buyer or seller might take to try and maximize their utility. Utility for a seller is the price paid less the cost of producing the good. For a buyer, the utility is their valuation of the good they wish to buy less the price they pay. Since buyers and sellers in a participatory sensing campaign have an interest in the same types of goods, we can apply the concept of an auction. In an auction, an auctioneer requests bids, and provides a good in exchange for payment from a winner selected by the auctioneer. In a reverse auction, the auctioneer still collects bids and selects a winner. However, the auctioneer gives the winner a payment and receives a good in exchange. While reverse auctions are not the only way to leverage incentive, they are the way our approach (discussed at the end of Section 1.6) manages incentive to address participation, inspired by Lee and Hoh's work described next. The auction model considers buyers and sellers interacting directly, however other successful incentive mechanisms have been used, such as recursive incentive [15].

An auction mechanism defines the rules which determine how bids are submitted and how winners are determined from the bids. A simple auction mechanism is the first price sealed bid auction. In this type of auction, bidders submit their prices to the auctioneer. All prices are secret, so participants do not learn each other's bids. The auctioneer then selects the winner with the best bid (highest if a forward auction, and lowest if a reverse auction). The advantages of this mechanism are that it is very easy for the auctioneer to run, and very easy for bidders to understand. However, the first price sealed bid auction does not lend itself well to the recurring reverse auction scenario of participatory sensing. Recurring auctions simply mean that there are multiple rounds. Each round is like a single auction, however bidders and the auctioneer can learn over time from repeated rounds.

To illustrate why first price is not a good choice, we provide a simple example. Suppose that there are N bidders and each round of the auction M winners will be selected by the auctioneer. Further suppose that for any bidder i, there is some true valuation v_i^t, which is the lowest price bidder i is willing to offer. The auctioneer will select the M lowest bidders each time, by the rules of the auction. If the bids of the participants are static and participants are sorted so that for bidder i and j, with bids b_i and b_j respectively, $\forall i, j : b_i < b_j$, then participants $1...M$ always win, and $M+1...N$ always lose. The auctioneer has no reason to change which participants it picks as winners, since despite the fact $N-M$ participants never win, its expense is minimized.

However, Lee and Szymanski discovered the so-called "bidder drop phenomenon" [16, 17], which results from participants motivated by the belief that they should receive incentive at least some of the time. When expectations are not met, users stop participating by dropping out of the auction. A decrease in competition does not seem inherently bad - the remaining M nodes can satisfy the system, at least for some time. Even ignoring the eventuality of battery depletion, there is a problem with this situation. The assumption was that bids remain the same over time. However, suppose one or more nodes occasionally probe the market by increasing their bid slightly to b_i' when they win, and reducing their bid back to their original b_i if they lose. To examine the impact of this with the least complication, consider what happens with this exploration of price when only M participants remain. Any participant can increase its bid by an arbitrary amount and still win the next auction. As a result, the only limiting factor on how high the bids can go is the system's budget. This is an undesirable scenario, and can be prevented by keeping competition alive, which is done by maintaining participation. Thus, participation is important and cannot be sustained by first price auctions, even assuming that participants are honest and do not collude.

Now that we have defined reverse auctions and suggested how they might be used in a participatory sensing campaign, we explore a paper that shares our beliefs and presents an auction mechanism [10]. In addition to presenting an auction mechanism, **R**everse **A**uction **D**ynamic **P**rice with **V**irtual **P**articipation **C**redit (RADP-VPC), and ways to measure its performance, the authors provide a formula for Return on Investment (ROI) which is a formal way to discuss participant tolerance to losses.

The authors postulate the use of reverse auction for distributing incentives to participants. A problem in modeling a reverse auction is that the true valuations of participants must be decided. The authors suggest that true valuation encapsulates all aspects of the "user's investment" - power consumption, resources, privacy, etc. - but this is a dynamic valuation. Depending on the location, time, campaign, resources available, etc. the true valuation of a particular participant may vary. This observation is in line with other research which suggests that certain locations may be sensitive and thus reflect a higher true valuation, or that changing social pressures might alter the risk of social costs associated with participating.

A fixed price mechanism is a simple solution, where all goods are viewed equal and the auctioneer pays the same amount for any given measurement. However, due to the heterogeneous nature of prices, as well as the dynamic nature of valuations, fixed price incentive is not an optimal solution. Either the mechanism risks dispensing far too much incentive to retain participation, or selects a fixed price that leads to significant numbers of participants dropping out due to their expectations about winning not being satisfied.

Expectations about winning are formally viewed as whether or not a participant's ROI value is above a threshold or not. If ROI for a participant falls below a threshold, which is 0.5 in the paper, they stop participating. The authors define ROI as follows:

Definition 1.1 Let us assume that participant i at round r with ROI S_r^i, has participated in p_i^r rounds prior to r, with true valuation t_i, and tolerance to loss β_i. Then

$$S_i^r = \frac{e_i^r + \beta_i}{p_i^r \cdot t_i + \beta_i}$$

□

As discussed in Section 1.4, first price auctions run the risk of prices growing out of control. The authors describe the same scenario, referring to the M nodes that always win as a "winning class", and the remaining nodes as a "losing class". They call the unchecked growth of prices "incentive cost explosion." To prevent this from happening, they add the concept of VPC - virtual credits that are rewarded to participants when they bid but do not win an auction round. The idea is that virtual credit will keep the cost from growing out of control by sustaining competition. RADP-VPC has a parameter α, which represents the amount of credit awarded for each consecutive round a participant loses. If participant i bids b_i^j in round j, and has lost k consecutive rounds, the auctioneer treats their bid as $b_i^j - k\alpha$. If the participant wins, k is reset to 0, and they are paid b_i^j. If a participant's true valuation is higher, they can eventually win through VPC and not drop off (as long as they have enough ROI tolerance). Lower true valuations still win when there is not enough VPC artificially pushing down the perceived bids of higher true valuation nodes. This allows all but exceedingly intolerant participants to win, thus keeping ROI values above threshold.

The application that the authors consider is a sensing task which requires a set number of measurements which are collected by a service provider. These measurements are collected through mobile devices and are all the same type of measurement. Collection is facilitated by a system that is designed to adapt to the changes in users' valuations, minimize the total expenses (the amount of incentive dispensed), and maintain quality of service which includes measurement precision, age, and geographic coverage. Age is important because in some cases, data is "perishable", meaning the usefulness of the data is affected by how long ago the data was sampled. Since the system is designed for perishable data, periodically new samples must be requested. This makes a system having recurring requests, which justifies designing with recurring auctions in mind.

Recruiting participants is another competition maintenance strategy. Since participants can drop out, recruiting former participants is a useful technique. If a participant dropped out, it means that the market conditions were not yielding incentive that matched their expectations. However, since the environment and prices are dynamic, it is possible that at a future point in time the distribution of bids will have changed such that a participant could rejoin and start winning. To facilitate recruiting former participants, any participant that is no longer participating is shown the highest price that won in the most recent round. If a participant sees that its true valuation is less than or equal to this revealed price, it should rejoin. The authors assume that only participants who have dropped out will receive this information. Suggested methods of delivery are e-mail and SMS. When a participant has dropped out, it needs to decide whether or not rejoining will benefit it. To do this it needs to calculate the expected ROI of rejoining. The authors use the following definition:

Definition 1.2 Let participant i at round r have participated in p_i^r rounds prior to r, with true valuation t_i, tolerance to loss β_i, and receive revealed price φ_r. Then the expected ROI, ES_i^{r+1} is:

$$ES_i^r = \frac{e_i^r + \varphi_r + \beta_i}{(p_i^r + 1) \cdot t_i + \beta_i}$$

□

To evaluate the performance of RADP-VPC, the authors compare it to the **R**andom **S**election based **F**ixed **P**rice (RSFP) mechanism, which randomly chooses participants until quality of service is met for the round, and pays them each a fixed price. Conceptually, the authors believe that RADP is better than RSFP from the service provider's view. This is because participants make the decisions about their prices based on their knowledge about current valuations, as opposed to RSFP where the mechanism is responsible for selecting a value that will satisfy the users' price expectations. To study the behavior of the mechanisms, strategies must be defined for the users (otherwise known as agents) who participate. The authors assume risk-neutral agents which react to winning or losing by modifying their bids. The authors formulate the utility U_i for participant i as follows:

Definition 1.3 Let U_i be the utility for participant i, bidding b_i^r in round r with credit for winning $c_i(b_i^r)$, true valuation t_i, and probability $g_i(b_i^r)$ of winning by bidding b_i^r. Then

$$U_i(b_i^r) = (c_i(b_i^r) - t_i) \cdot g_i(b_i^r)$$

□

Furthermore, the attribute of "risk-neutral" is important. When considering agent behavior, a designer must take into account how they view risk. Using the standard auction terminology the authors distinguish the three risk attitudes and their corresponding objectives, namely:

1. Risk preference: Maximize $(c_i(b_i^r) - t_i)$
2. Risk neutrality: Maximize $U_i(b_i^r)$
3. Risk aversion: Maximize $g_i(b_i^r)$

As a user's bid goes up, the gain from winning a round increases but probability of winning decreases. Since the goal is to optimize $U_i(b_i^r)$, participants must be aware of this trade-off as they set their bids. This leads to a simple bidding strategy which we adopt in the simulation discussed later in Section 1.6. If a participant loses in round r, then $b_i^{r+1} \leq b_i^r$ so $g_i(b_i^{r+1}) \geq g_i(b_i^r)$ - either the bid and probability do not change, or the bid is lowered so the probability of winning might increase. Symmetrically, if a participant wins in round r, then $b_i^{r+1} \geq b_i^r$ so that $g_i(b_i^{r+1}) \leq g_i(b_i^r)$. This adaptive behavior is bounded by the constraint that for any given round r, $U_i^r > 0$ and $b_i^r > t_i$.

The descriptions of the experiments conducted can be found in the original paper. The results show that for a variety of distributions of t_i RADP-VPC results in lower total cost than RSFP. This is because in RADP-VPC, lower bid prices are favored. Because VPC prevents price explosion from happening, RADP-VPC results in a more efficient use of budget, which is reflected in the total cost being lower. The authors do not discuss how to tune the parameter α, but observe that while initially increasing it results in a higher number of active participants, after a certain α the number of participants starts to decrease. This suggests that there is an optimal value for α. If the parameter is set too low, then the addition of recruitment can still result in lower total cost, since the dynamic nature of market is advertised to participants which can then rejoin based on their ES_i^r.

The authors claim that because of VPC, if there is correlation between geographic location and true valuation, that RADP-VPC will retain additional participants and thus create a more geographically balanced set of data than a mechanism which does not account for participation. However, there will still be a bias towards lower true valuations, since data is time sensitive and $g_i(b_i^r)$ for a participant i is higher if t_i is lower. This correlation may correspond to socio-economic factors that are geographic, such as economic disparity between neighborhoods.

Since the system is supposed to provide sufficient geographic coverage, additional design considerations can be taken. Considering arbitrarily defined regions, an auction can be run in each region. This removes dependency between regions, creating several markets or auctions. Since separate auctions are run, if each auction corresponds to a group of similar true valuations, the markets are less stratified and the bias caused by having a lot of participants with significantly cheaper true valuations is diminished. An important feature in the paper's experiments is that participants do not move between regions. Depending on the definition of regions, this can be an unrealistic assumption. Alternatively, if the regions are large enough to guarantee that mobile users do not move from one region to another, the auctions may be so large that no destratification will occur.

On the topic of privacy, the authors note that a limitation of the mechanism is that participant locations are revealed whether they win or lose a round. Since true valuation encapsulates several costs including privacy and resources, losers are penalized by bidding in the round, expending effort, and providing location, but not receiving incentive. The authors suggest encrypting data until winners are decided to get around this, but approximate location (regions, in the case of multiple smaller auctions) is still provided, and encryption precludes any sort of data quality enforcement. Another option is to use RADP-VPC with a "data broker". This effectively shifts the problem downstream to a third party who can manage security and privacy. However, such an entity may be able to enforce more specific policies that better cater to a participant, and allow the participant the opportunity to participate in auctions across service providers.

Finally the authors mention several real-world challenges that face RADP-VPC. In asynchronous systems, such as the one that is described by the next paper we examine, the concept of an auction round is difficult to define. The length of time that data is useful before perishing can help tune this value, since delaying decisions longer than the data lifespan means by the time the auctioneer selects winners, the data may no longer be useful. Calibration of rounds may also be considered based on supply and demand conditions. For example, if a newspaper is looking for photographs of an individual, and only one person has a photograph, that individual can dictate whatever price they want. The newspaper can either wait for additional photographs to become available (extending the auction round), or accept that there is a limited supply of measurements (images in this example), which may result in a higher total cost for the system.

Systems may be heterogeneous, meaning multiple types of measurements are required. Formulating a mechanism and selecting winners from an auction becomes a more difficult problem because measurements may not be equal in value, the system may not require the same number of each type of measurement, and users may be at risk of having resources depleted faster than expected if they are selected for many types of measurements. One possible solution is to run a separate auction for each type of measurement, but such auctions would be unable to determine if they were overutilizing a given participant. The participant has the option of changing their t_i to reflect personal resources becoming increasingly scarce, but how this value should change is not clear and might not be a quantity that users could easily determine. This leads to another consideration, which is having a tool to automate bidding or assist in adjusting bids. Like other systems which have interfaces, a major software engineering concern is making systems easy to use and understand while still providing necessary information.

An important design consideration in any mechanism is to make sure it is robust against collusions. Colluding is the act of one or more participants working together and behaving in ways that may result in additional gains for the colluding parties at the expense of other bidders. As a simple example of the recruitment vulnerability, consider two participants i and j that decide to work together to maximize their profit by splitting i's profit. j participates in the first round and then quits. They then receive information about the highest winning bid every round since the system tried to attract it back. If frequent participation is required to avoid unnecessary disclosure, j can consistently bid a value that is very large, or learn a less suspicious value that guarantees losing auction rounds through adaptively increasingly their bid over time. j then provides i with the highest winning bid, and i can use this information to make winning bids much closer to the disclosed bid than they would otherwise be willing to make based on only $g_i(b_i^r)$. This results in a decrease in the efficiency of the overall system. The authors do not discuss collusion protection techniques in this paper.

The authors mention a paper that attempted to develop an understanding of true valuations. The study was done at a university with a limited demographic, so the value at which users were willing to sell their data (25 cents) is not necessarily applicable to all situations. However the insight that compensation can allow participation despite privacy concerns, and that valuation is situational and a multi-disciplinary problem, is still of value [18].

Privacy and security are not discussed beyond the need to keep recruitment messages personal, and data integrity being important, with the authors suggesting trust management [19]. The paper shows challenges in mechanism design, and one solution for an auction mechanism that is oriented towards retention of participants. The framework is general, and can be applied to any homogeneous client-server sensing system. Considerations included whether or not it could apply to asynchronous systems or heterogeneous systems, and that true valuation is dynamic and reflects all user valuation including cost of resources, privacy costs, and worth of a measurement. In addition, the paper provided a formal way to look at incentive and tolerance to losses through ROI.

The next paper we examine describes a system designed to study recycling practices at a university [20]. The measurements consisted of photographs of locations where trash and recycling were deposited (such as “waste bins”) and optional tags that participants could input prior to submitting the photographs. This system is of particular interest to the chapter both because it is an example of asynchronous participatory sensing, and because it explores the effect of different incentive schemes applied to the task. While incentive does not necessarily involve application of the concepts introduced earlier regarding markets, the COMPETE mechanism which is described in the following discussion creates incentive-driven competition with the goal of promoting participation.

The authors define participatory sensing based on three requirements:

1. Users are involved in decisions about what will be collected. This is the same

belief as expressed while discussing participatory privacy [6] at the beginning of Section 1.5, which has authors in common with this paper.

2. Users contribute data collected during daily routines. The mention of daily routines is important since it suggests that participatory sensing systems are tied to the patterns in human behavior.

3. Users are connected to the context/purpose of the tasks they perform.

Five incentive schemes were considered, four of which were micro-payment schemes. Micro-payments are incentive rewarded on a smaller, more frequent level. The incentive use discussed previously in this chapter has all been micro-payments since incentive is awarded based on single actions or measurements. To make a fair comparison between the five schemes, the maximum payout from each scheme was the same, namely the MACRO amount.

- MACRO: One large payment for joining the experiment
- HIGHμ: 50 cents/valid measurement
- MEDIUMμ: 20 cents/valid measurement
- LOWμ: 5 cents/valid measurement
- COMPETEμ: Between 1 and 22 centers/valid measurement, based on how many samples taken compared to other peers. Rankings were public, which differs from the "sealed bid" approach of competitive incentive mechanisms such as RADP-VPC.

Experiments were run using 55 Android phones. The authors found that COMPETE resulted in highest number of samples, but competition motivated some users while others were indifferent or performed worse because of the competitive aspect. Micro-payment options did better than MACRO since there was a system-imposed sense of worth, where as MACRO users complained they were unsure what a measurement was worth. This shows that the campaign and mechanism influence how participants set their true valuations. Additionally, there was no incentive gain for a MACRO user for submitting a photo, so the payment scheme inherently did not encourage measurements. Looking at measurements over time, MACRO and COMPETE users became less motivated as the campaign continued. MACRO users cited a loss of novelty over time. COMPETE users "burned out" over time, with it becoming less important if they held a higher rank. HIGHμ, MEDIUMμ, and LOWμ users tended to ration out the number of measurements so that they would receive maximum incentive by the end of the campaign span. As users fell behind in quota, they would compensate later, causing slight increases. In the case of LOWμ, there is a sharper increase near the middle of the campaign, since many more measurements needed to be taken to reach the maximum reward compared to MEDIUMμ or HIGHμ.

The quality was not strongly affected by the payment scheme used - in all cases the percentage of invalid pictures is low. However, in the case of COMPETEμ, 10% or less of submissions had optional tags, while in all other schemes, an average of 50% or higher tags could be observed, with significant variation in the percentage.

Coverage was highest with COMPETEμ, where users would alter their routine to seek additional measurement opportunities. This conflicts with the participatory sensing definition suggesting data is collected during participants' daily routines. MACRO users did not alter behavior at all. The participants on the remaining micro-payment schemes would alter their behavior by going to measurement sites that they could see, but would not necessarily have measured if not on the micro-payment scheme.

From the above results, it is evident that the incentive scheme influences behavior of participants. Furthermore, the scheme was made known to the user at the beginning of the campaign, which supports the emphasis throughout the chapter on transparency and ensuring that users understand mechanisms in the system. Authors acknowledge it is an initial small scale study. It is unclear if longer periods would have resulted in participants burning out regardless of payment scheme, and if the percentage of measurements with tags would have changed. Fixed micropayments appeared to perform the best - effectively the payment scheme translated into a goal that was easy for participants to conceptualize. Authors suggest that if a mechanism could be added to decrease "participant fatigue", then COMPETEμ might perform better. The issue of "participant fatigue" is important to consider in system design, since this means a mechanism with the purpose of maintaining participation must look at long-term behavior. One way this could be done is by using techniques discussed earlier in the chapter involving the ROI model and altering β over time. ROI does not measure a cost in "interest in participating" based on the potential toll on users of participating, however modeling such fatigue might be done in a manner similar to the ROI equations. Lastly, this paper demonstrates that having a payment scheme can improve coverage spatially and temporally, but not all designs result in an improvement. Deciding if coverage is a critical factor and addressing it is a challenge that participatory sensing system designers must consider, and this paper agrees with our identification of coverage as a challenge in Section 1.3.

A sensing system that was not a participatory sensing system, but made use of incentive to solve the problem of limited resources, was **S**elf-**O**rganizing **R**esource **A**llocation (SORA) [12]. The motivation for the system was that sensor networks are comprised of low power devices able to compute and communicate. This is also the case with devices carried by humans in participatory sensing systems. The need to minimize energy used by the system and thus allow more energy to be used by the participant for normal tasks makes efficient allocation important. As discussed earlier in the chapter, the environment and human factors are dynamic, so the adaptive nature of SORA is also valuable to examine.

The nodes in the system are modeled as self-interested agents with the goal of maximizing "profit". In this paper, profit is virtual and is exchanged for virtual goods which are produced by performing actions. While this is not directly useful for participation, it illustrates a very different approach from reverse auctions to applying market mechanisms. Since this paper was published in 2005, deployments would have consisted of dedicated nodes instead of nodes carried by humans. Excluding the aspect of participation however, traditional distributed sensor networks share many of the other attributes and challenges of participatory sensing systems.

The actions that the authors describe are taking samples, aggregating stored measurements, listening for messages to forward, and transmitting messages. These same basic functions can be applied towards participatory sensing networks, though due to the use of mobile devices already connected to a service provider or wireless AP, communication tends to be less of a concern. However, as we will discuss in when exploring CarTel in Section 1.6, aggregation and delivery concerns are still applicable to deployed participatory sensing systems.

The authors describe the adaptive behavior they expect from nodes as “dynamic”, a term that was mentioned independently in Sections 1.3 and 1.4. The recurring use of “dynamic” highlights an important detail of designing for participatory sensing and in many traditional sensor network designs. Even if the system is assumed to be static, its environment is likely changing in ways that necessitate system adaption to it. By adding humans to participatory systems, the number of possible changes to which the system must react increases. The authors support this observation by indicating that a static schedule of actions, or a dynamic mix of actions on a fixed energy budget, will ultimately result in potential energy waste. This happens because different nodes are in different situations as defined by factors such as network topology and proximity to phenomena of interest. As the network and environment change, the optimal actions that any given node should take also change. The impact of a system being dynamic is significant enough that the authors credit existing work in market oriented programming [21], but assert that SORA differs in that it solves a real-time allocation problem, whereas Wellman’s work only solves a static-allocation problem.

SORA applies reinforcement learning [22] by incorporating an exponentially weighted moving average (EWMA), a well-known filter. Each node computes utility $u(a)$ for an action a based on the probability of payment β_a and the price of the action’s good, p_a. β_a represents the effect of the learning, and is adjusted using an EWMA filter with sensitivity α. We omit the equations here, but they are described by Mainland, Parkes, and Welsh in the original paper. In this way nodes learn actions based on what benefits them. To influence the nodes, the system globally advertises a vector of prices that specifies how much the system is willing to pay for a particular action-produced goods. The process of deciding which action to take is based on the current global prices and the current state of the node. The state of the node is the current energy budget. Note that goods are only purchased by the system if they are useful (submitting/aggregating an interesting measurement, or routing an interesting measurement towards the base station).

The system as described so far has not addressed how to incorporate energy. As mentioned before, a fixed energy budget poses problems for resource allocation, because nodes may consume energy too quickly. To rectify this, the authors use a “token bucket model” in which the bucket can hold a maximum amount of energy, (C), and the bucket fills at a rate of (ρ). The bucket size represents the largest amount of energy that the node is allowed to use at once. If C is set to the node’s entire battery, then as in the fixed rate case, it is possible to deplete the battery rapidly, assuming ρ is not relatively large. ρ is a gradual "recharge over time" rate which may not represent physical charging, but rather it could be used to model the fact that in the case of user operated nodes, users periodically recharge their devices. In our discussion of this paper, we only consider the original design, which is that ρ is designed to limit frequent bursts, while C controls the maximum burst of energy consumption allowed at once. It is also worth noting that here, energy is a constraint and modeled as a separate budget with a separate currency. This is unlike the previously mentioned RADP-VPC, where true valuation encapsulated all perceived costs, including usage of resources such as energy, and it used the same currency as the incentive.

The application that the authors consider is tracking vehicles through use of magnetometer measurements. Such an application would be hard to consider as a participatory sensing task. Yet, if the task could be accomplished by user's devices taking pictures and running image processing, and vehicles were differentiable, the task could be framed as a participatory campaign. Additionally, the authors state that SORA is not specifically designed for vehicle tracking, so the design lessons are general and can apply to participatory design. Specifically, SORA can used for other systems as long as the actions (and resulting goods) are defined, and any dependencies are explicitly stated. Since the nodes run a simple program, they cannot make assessments independently to determine dependencies, and rely on knowing that an action can or cannot be completed based on current goods (completed actions) explicitly.

In addition to being able to adapt and let different nodes express their circumstantial differences, the authors add a design goal of allowing control. The system operator should be able to control node behavior, and this is done simply through the global price vector. Any change made to this vector is propagated to the nodes. This incurs some overhead, but the authors mention that any of several existing "efficient gossip or controlled-flooding protocols" can be used. Still, the authors suggest that price vector updates are done infrequently. Unlike adapting at the individual level, control is important because some changes may require a global view to perceive and respond to them. The control is not absolute, since nodes must react to the changed price vector through the EWMA-based learning mentioned above. In experiments, the authors found that without large changes in the global price vector, the effect was hard to observe.

In order to adapt to changes, the nodes need to periodically try actions that are not the most profitable. This risk-taking behavior is implemented by an ε-greedy algorithm, where ε is a risk taking factor (the authors use $\varepsilon = 0.05$) . The nodes behave as expected and take the action that currently is believed to maximize their profit with probability $1 - \varepsilon$. The rest of the time, an action is chosen from all possible actions, with uniform probability of choosing any given action. By having $\varepsilon > 0$, nodes can never completely be blocked from learning about an action, regardless of the EWMA α chosen. An interesting design lesson is that nodes must be given the chance to explore the system, and this exploration can lead to local adaptivity. Whether a participatory system designer tries to anticipate participants deviating from "rational" behaviors or not, humans are liable to do so. For example, it is this deviation that leads to the incentive cost explosion in the case of reverse auctions. If a designer assumes participants always act according to the expected algorithm, the system cannot be designed to be robust against such behaviors.

The authors compare their algorithm against a static action schedule, a dynamic action schedule that adjusts based on the current energy budget, and a "Hood tracker" [23] to compare against a published system. Aggregation-based methods perform worse with respect to error, however this is due to error being measured based on where the target vehicle was when the base station received a given measurement. Thus the additional time spent collecting measurements and processing them during aggregation introduced time lag, which in turn increased the distance the vehicle moved before the base station received the measurement. Any actions taken that do not result in a measurement eventually arriving at the base station contribute to wasted energy. The authors note that "In a perfect system, with a priori knowledge... there would be no wasted energy." The difference in energy efficiency between SORA and the static or dynamic methods are about 40% once $C > 1500$. This is due to SORA's learning approach is and shows that the reinforcement learning method results in much higher energy efficiency with small costs in accuracy.

Through experiments the authors examine the effects of ε and α. We do not summarize those results here, however what the authors do find is that the two parameters serve as a way to tune behavior prior to the experiment, while the global price vector allows for control during the experiment. In the case of participatory sensing, ε and α would be parameters the user could change, while the global price vector would be an example of something the system operator would change. While in both cases, parameters affect behavior, in the case of participatory sensing, the participants also express control through parameters. To prevent the two groups from working against each other, design should be oriented towards making a system easy to understand, transparent, and having operators and participants cooperate. This is in line with the philosophy suggested in by the first paper in Section 1.5. The other design lesson is that sensor network applications require addressing "extreme resource limitation of nodes" and the fact that the environment or universe is not fully known and over time it changes.

1.5 Privacy Oriented Approaches

While the designs so far have primarily focused on incentives and maintaining participation, we now shift focus to systems that were designed with a primary goal being privacy. The first paper selected serves as a transition from focusing on participation to focusing on privacy, by involving participants in the design of policies related to privacy. We then discuss two systems that are designed with privacy involved, but do not directly involve users in high-level decisions about information flow.

Privacy of participants and ethics regarding information collected by a participatory sensing system are certainly a concern. We begin by summarizing and discussing a paper that addresses these topics. For simplicity, we refer to "participatory urban sensing" as 'participatory sensing' [6].

Shilton et al. state that designers of participatory sensing systems need to *proac-*

tively take steps address to the needs and requests of users, which may be quite diverse. In addition, they bring up the idea of "social trust" by stating that users must be "significantly involved in the design process" in order to attain such trust. A definition of social trust is not provided in the paper, however the general idea is that participants in a system should be able to trust that the system will not misuse data they provide. The authors agree that user participation is an important challenge, and that addressing privacy through participation ("participatory privacy regulation", not to be confused with participatory sensing) is a way to use participants. This is an application of human involvement unlike those considered earlier in this chapter. Despite the lack of quantitative measurements, incorporating a participatory model into privacy offers insight into the complexities of privacy and participation, and is an approach that can be applied in design of such sensing systems.

In prior research that the Shilton's group did on sensing projects, they found that privacy concerns arose. These "serious privacy concerns" were identified when tasks included location tracking and image capture, which are both example sensing tasks that we independently suggested earlier in this chapter as uses for participatory sensing. According to the authors, issues about privacy were "one of the first ethical challenges". This supports our belief that in participatory sensing system design, privacy is an issue and must be addressed.

The authors note that sensing systems can be installed in which participation is passive and achieved simply by being in the same space as the sensor system. The passive nature of being "in" a participatory sensing system is backed up by other literature [24]. However, Shilton et al. suggest that participants must engage "with" the system in order to collect data that is not only useful, but ethical. Without participants being involved in design and usage, the authors indicate that data sampling may be invasive. As we will show both in the remainder of this section and in Section 1.6, in other systems, the participants are rarely viewed as designers of the system, and are instead presented with a fully developed system which may have no controls or limited controls through a set of designer selected parameters.

In the paper, a list of privacy and security techniques are briefly discussed. We provide a short list of these techniques, but exclude references, which can be found in the original paper. The value of such a list is that it illustrates a wide array of tools that exist for designers considering participatory sensing design. While we do not examine the application of these principles in other works, several appear in the selected few works that we choose to review.

- Systems that provide warnings, notification, and/or feedback about privacy
- Ways to identify vulnerabilities in information systems (which could lead to unintended data access)
- Systems that allow users to choose what data they want to submit
- Identity management
- Selective retention / "forgetting" data

- Encryption of data
- Statistical anonymization

Personal and social variables dictate how a participant shares information. For example, not revealing the location of one's home, or appearing in a particular social role such as a manager [25]. The environment affects these decisions by affecting what a participant is comfortable with. Social norms, situational pressure, and personal relationships are just a few factors that can play into individual decisions about privacy. Understanding of information flow, or beliefs about flow affects the willingness of participants to share information. If there is belief that the flow of information is very limited, the privacy risk is low. If there is an incomplete understanding of where information can go, the potential privacy risk is higher and participants may be more reluctant to contribute. Understanding information flow involves beliefs about who has access to the information, how those entities spread information, and to whom information is spread.

The paper introduces participatory privacy regulation, which is designed to allow decisions at both the individual level and in groups. These decisions develop policies about how the sensing system can collect, store, and use data. Groups are sets of multiple individuals, which are liable to have some social context. Consider a participatory system that is designed to detect concentrations of volatile organic compounds (VOCs). VOCs from sources such as paints are believed to pose a significant health risk to humans, particularly in heavy concentrations. If a system was deployed, the resulting data could be used to identify locations that could be linked back to individuals, as in the IED example from before. In addition to individual privacy being at risk, the sensing campaign might reveal high concentrations that can be traced to neighborhoods or facilities belonging to a particular company. This can result in social loss, such as a negative opinion of the group. Unlike individuals, since groups are comprised of many entities, decision making can become more complicated. This also supports our belief that privacy is a social, as well as ethical concern.

A binary 'share or do not share' system is not sufficient to meet the goals of participatory privacy regulation. Instead regulation is a process. Users decide what information about themselves can be accessed based on the context of requests. This context has "specific, variable, and highly individual meaning in specific circumstances and settings." Throughout the entire sensing campaign, privacy is an issue which is put at risk in different ways. The first place to consider privacy is in deciding which measurements are taken, and how much can be controlled about the measurements. Examples of sampling control are deciding constraints about frequency of samples, the resolution of measurements taken, and metadata about them. Once the data is submitted, privacy decisions must be made based on who can access the data, to what degree, and who can access which results. How long a system keeps data that has been submitted is another detail that must be decided. This retention decision affects the balance between privacy and verifiability of results.

Another point the authors make is that involvement in the process provides an understanding of data policies and information flow. This can help the user make decisions about valuation or willingness to participate in a given campaign. A design philosophy that can be taken away from this view is that context is given through transparency, and providing information about data management is important to participation when designing participatory sensing systems.

The authors present five principles to drive design under participatory privacy regulation. We list those principles along with abridged explanations.

1. "Participant primacy": As previously mentioned, participants should be involved with the system design to avoid participation being invasive. Estrin et al. express this by stating that all participants should also have the role of "researchers". This gives participants an understanding of the entire system from data collection to use in applications, which leaves them better equipped to make decisions about their privacy.

2. "Minimal and auditable information": The sensor platform may allow collecting far more data than necessary for the goals of the campaign. Minimizing the amount of data collected decreases the risk to privacy, and makes both the system and information flow easier to understand. "Coarse control" may allow the user to enable and disable collection, while "fine control" may allow management of retention or data submission on a case-by-case basis. Implementing fine control requires auditing mechanisms that are easy to use, which is a significant challenge.

3. "Participatory design": Participant input into the system's design should be done as a group process. Individuals have a hard time anticipating future risks of a decision based on present privacy decisions [26]. Estrin et. al suggest that communication with participants can highlight spatial or temporal regions of concern or excitement.

4. "Participant autonomy": The system's design should provide a way to avoid "the pitfall of relying entirely on configuration" by making privacy decisions part of the normal participation workflow.

5. "Synergy between policy and technology": Software and hardware alone are insufficient to solve the problems of ethics. "Institutional policies" are required, with the system serving as a tool to help facilitate policies and the enforcement thereof. Participatory policy making should include all parties involved in the system. Whether responsibility for a given policy lies with the policymakers or system is an issue that must be addressed separately for each issue during system design.

We transition to a paper that discusses k-anonymity and l-diversity [8], properties that quantify the degree of anonymity a user has in a dataset. k-anonymity is a term often found in discussions of privacy, however we selected this paper both to show the strengths and weaknesses of k-anonymity, and to study an approach that goes further by introducing l-diversity.

Huang et al. state that in "typical" participatory sensing applications data is "invariably tagged with the location... and time". This is required since in such applications, time and location are necessary contexts to extract information relevant to the sensing task. However, this is a privacy risk since it can reveal information about users, particularly if multiple submissions can be linked together. The authors then indicate that a priori knowledge of a user's locations can be used to defeat pseudonyms [27], or user identity suppression [28], and describe the effort required as "fairly trivial". Consistent with our beliefs about participatory sensing, the authors recognize that participation is "altruistic", meaning that it is a voluntary act. The violated privacy would introduce a significant cost to participants, and the risk of such a loss may deter participation. Thus, it is important that privacy is considering in designing a participatory sensing system. The emphasis of the paper is on spatial and temporal privacy. There can be other kinds of privacy, for example in a campaign related to health, it may be useful to cluster based on types of ailments. To compare to tiles in 2D space for traffic, the health example might correspond to something like blocks of International Classification of Diseases (ICD) codes.

Tessellation, a technique used in AnonySense [29] takes a real spatial location and reports an artificial region ("tile") with at least k users, instead of the actual location. This type of modification is called generalization. k-anonymity [30] on an attribute, means that at least k distinct users share a value for the given attribute. Generalization is a natural way to achieve k-anonymity by reducing the resolution of an attribute until each class shares at least k members. Because of generalization results in a decrease in resolution, the authors are motivated to show that tessellation may not be suitable for applications where higher precision in location information is required. As an example, they mention traffic analysis which may require knowing which road the measurement comes from. In order to achieve k-anonymity with an acceptable k value using tessellation, the size of tiles may encompass several roads. The system then has no way to determine which road the measurement is used for, and thus cannot effectively perform the task of traffic analysis. The authors modify tessellation to use the coordinates of the tile's center, instead of just a tile ID, when reporting. This alternate method, TwTCR, provides additional context which can help the application determine which station the report is for. For example, simply using the distance between the center and known stations may indicate that only one station is near the center of the tile. Despite the inclusion of tile centers, if the tiles cover a large area, the application may still be unable to determine which station a report is describing. The authors introduce VMDAV, which is described below, for these cases.

The authors also use microaggregation, which is another way to achieve k-anonymity. This decision is influenced by the ability of microaggregation to be used for continuous numeric attributes. Microaggregation creates equivalence classes (ECs). Within an EC, members have a common value for sensitive attributes. The authors observe that the common value is usually an average for that attribute. Clustering is done to try and maximize similarity between members of the EC, where the similarity metric for numerical attributes is often simply the L^2 norm. Unlike tessellation, which is a generalization method, microaggregation is considered a perturbation method since attributes are not generalized, but otherwise altered. The implementation used is Variable-size Maximum Distance to Average Vector (VMDAV) [31].

Having introduced both TwTCR and VMDAV, the authors evaluate if there is a reason to one over the other. Through examples, the authors show that in some situations TwTCR is better, while in other situations VMDAV is better. When the users are distributed nearly evenly across regions, VMDAV performs better. If the user distribution is dense, then TwTCR becomes the better choice. Based on both findings a third method, Hybrid-VMDAV, is proposed. This method works by using TwTCR if a cell has more than k users, and VMDAV otherwise.

To consider privacy vulnerabilities, there must be an adversary who attempts to gain knowledge through the data available in the system. The authors assume that there is an adversary with knowledge about their victims' behaviors (spatially or temporally), but do not have knowledge about the true location or time in reported information. A simple example that the authors provide is the adversary overhearing that their victim will have a medical treatment during a particular part of a specific day. If the adversary has access to the reports, whether this is by being an administrator of the sensing system or by a security exploit (such as eavesdropping), they may be able to use the information from before to determine which group the victim is in. From there, information about the group may reveal specifics, such as a cancer treatment facility being located in that tile. Finally, the adversary can conclude that their victim was treated for cancer, despite the k-anonymity, a failure known as "attribute disclosure". This shows that k-anonymity is not sufficient to protect against what are called "background knowledge attacks" in literature [32].

Two types of disclosure can happen as a result of the information stored in the system:

1. Identity disclosure - linking a specific record to an individual. (k-anonymity protects against this.)

2. Attribute disclosure - private information revealed from the "semantic meaning" of an attribute.

Again using the cancer patient example previously mentioned, the adversary cannot know which of the k records in the tile with the treatment facility belong to their victim, so there is no identity disclosure. However, since the tile is known to be in the region of a treatment facility, which is semantic information, the adversary learns that their victim has a specific medical condition. This is an example of attribute disclosure. In addition to the background information attack, there can be homogeneity attacks. These use "monotony" of attributes to gather information. In the cancer example, background information is used (the time at which the treatment is done) to narrow down possible records, and homogeneity is used to determine that all remaining records share a tile ID which contains the cancer treatment facility.

To rectify the above problem and further protect against disclosure, the authors suggest use of l-diversity. The concept of l-diversity is that within any group, there are at least l distinct values for a sensitive attribute. This prevents the conclusion in the above example, since with higher l-diversity, the adversary cannot conclude that his victim is specifically the report in the cancer treatment facility. In more general terms, monotony corresponds to a low l value, so by increasing diversity, homogeneity attacks are increasingly difficult to perform, if at all possible. The authors note that l-diversity can be applied to a k-anonymity algorithm, and refer to an existing l-diverse implementation of VMDAV, LD-VMDAV [33].

The structure of the system used consists of several components:

1. Mobile nodes (MN) - the actual mobile sensing devices (usually on humans, but could be in cars)

2. Registration authority (RA) - Validates other components, gives security elements necessary for authentication (such as certificates).

3. Task server (TS) - Allows communication from applications to MNs, and does not allow tasks that would violate the MN's required privacy (k,l).

4. Report server (RS) - Combines reports, then submits them to the application.

5. Mix Network (MIX) - Responsible for facilitating anonymous communication. De-couples the identity of the sender from the report so any component/application cannot figure out which reports come from a particular component/application.

6. Anonymization server (AS) - Generates tiles for TwTCR and equivalence classes for VMDAV.

The AS is responsible for facilitating privacy by taking requests from users regarding their required privacy (k and l values), and returning an anonymized value that the MN can then report to the system. The authors assume that "the AS is owned by a third-party operator and is isolated from attacks," and that the AS does not collude with other components to compromise user privacy. The authors also require that the AS has periodic updates about the locations of MNs to effectively provide privacy, and assume that MNs will trust the AS. However, the authors recognize that in actual deployments blind trust in a third party entity also constitutes a single point of failure and thus is not reasonable. To fix this, they suggest using Gaussian perturbation with a normalization factor p. The formal details are not covered in this chapter. The authors recognize that perturbation on its own can be defeated, but suggest it as an added layer of security and use Gaussian perturbation for simplicity.

Authors measure the performance of algorithms in simulations using the Dartmouth campus traces [34]. Performance is measured based on information loss (IL) and positive identification percentage (PIP), which are defined in the original paper. PIP is application specific, since a "correct association" may refer to the system identifying a single attribute or a tuple comprised of several attributes. A notable result is that the performance is not affected by the percentage of users that participate, meaning the algorithms can scale arbitrarily without performance loss. Hybrid-VMDAV achieves higher PIP and lower IL than either VMDAV or TwTCR, however even Hybrid-VMDAV's PIP is only around 35%. The authors explain that the metric used in experiments was a simple Euclidean distance, and that choosing a more advanced metric could improve performance. Choosing a metric that is suited to the data depends on the environment, application, and resulting attributes, and thus is something that must be considered during system design. Alterations during algorithm execution would result in a new set of tiles or equivalence classes being computed, which would then invalidate older measurements or leave confusion as to which set of anonymized attributes a given report referred.

The authors examined the effect of the Gaussian perturbation on location and found that by increasing p, IL increased. This is expected since perturbation adds noise to the data, and added noise results in higher loss of information. Proper selection of p can keep IL low and balance privacy and system performance. How to tune this parameter is not discussed in the paper, but, as with other designs, finding suitable values for parameters should be part of system design and is likely dependent on the application and current state of the network and environment.

As established in the previously discussed paper by Huang et. al [8], k-anonymity is not a sufficient solution for preventing attribute disclosure. PoolView, a participatory sensing application designed with privacy as a goal, develops further privacy measures to overcome the shortcomings of k-anonymity [9]. In PoolView, data collected from users is viewed as a time-series of values called "streams".

PoolView is designed with the belief that there is no trust in the system outside the nodes, and places the responsibility of data protection upon the nodes before data submission. The authors observe that anonymity does not work if there is location information since the resolution of data may have to be quite low in order to provide anonymity through approximate location. Preventing location information in a particular region may still indicate identity, and may create large areas with no measurements by overlapping with areas where position is not measured.

Perturbation cannot protect because correlation between data and correlation between data and context. This correlation can be exploited by statistical tools such as Principal Component Analysis (PCA). Tools can be used to make application-specific noise so reconstruction cannot happen for an individual stream, but for information about the community it can. Since the individual is not at risk using such tools, the authors indicate participants have no need for anonymity. Nodes run by participants send information to pools after agreeing on a noise model. The pools can then be accessed by applications to gain information with little error about aggregate statistics, but cannot obtain information about individual measurements without high error. As in several other systems discussed in this chapter, the application is designed to be simple with the belief that this leads to usability.

The authors consider the application of collecting traffic statistics (which are computed after measurements have been taken), and an ongoing average weight of participants (which happens as data from streams arrive). These are only two examples of time-series data that could be used by PoolView. However, it is worth noting that another design decision that must be made is when data can be accessed. Designing a system that allows partial data to be used means the system cannot have a reliance on knowing the full data in order to process or present results.

The system works by having the pool send a user the noise model when the user joins. The noise model has a distribution of parameters, and the user generates a particular instance of parameters from this distribution. This noise is then added to the user's measurements prior to their submission to the pool. Having a distribution of parameters means the communal noise can be guessed and removed to get aggregate information. The accuracy of this method is higher when there are more participants, since the theoretical parameter distribution will be more closely matched. Getting the model from an untrusted source is risky since the model could be designed to make the stream vulnerable. For example, if the model is a constant, then the stream is simply a shifted version of the actual time-series, which as previously indicated, is not adequate protection. More complicated models, such as noise with a known spectral range, can be used since a filter can then be applied to remove the noise. This is not a problem, because noise model must fit the phenomenon. In other words, even if the noise model describes a person gaining weight over time, but the participant's time-series indicates weight loss, this is still acceptable. The participant's time-series must be one that the noise model can generate using the parameters available with a probable set of values. If this cannot be done, the model is a poor description of the phenomenon. User can test with curve-fitting before deciding to submit a stream to a pool. A tool is provided in the PoolView application based on two user parameters, p_1 (the fitting error threshold), and p_2 (the threshold on probability that data was generated by noise model).

In order for a pool to estimate the actual distribution of a series, the authors show the system must solve a deconvolution problem. The formal math is omitted from this chapter, but provided in the original paper. The approach used to solve this formulation is the Tikhonov-Miller method [35]. In solving, two variables arise. The authors refer to the first as the "regularization coefficient", λ, which comes from needing to provide an error bound, ε. The second parameter is in the method's formulation and is represented by ν. A larger value of ε means a larger upper bound on the reconstruction error. The error decreases as the number of participants increases, but even with low numbers of participants the error is reasonably small.

Since the noise is uncorrelated and the resulting signal-to-noise ratio is relatively low, PCA does not work on PoolView's approach. Through experiments, the authors show that PoolView is resilient against PCA, while perturbation by white noise is not sufficient to protect individual measurements. The authors also recognize that the model is available and explain why this does not pose a problem. One attack is to try to estimate the parameters used. Using MMSE (minimum mean squared error), if the noise model is close to the actual phenomenon, and there are many candidate noise streams, the authors deem the approach robust. This creates two ways for a server to be malicious - send a model that does not match the phenomenon (in which case MMSE may give a low-error result), or a very narrow set of possible parameters for the model. If a user suspects a model of being inadequate, they can opt not to submit to the pool. A participant's parameters should not be at a tail of distribution since either the user is unusual or the distribution is not representative. This poses a risk of participants thinking a valid model is untrustworthy if they are anomalous. The authors suggest that in social settings, a user may know if they are representative of the community or not, and this can assist in their decision to trust or not trust the pool. Another vulnerability is that the pool could change models. To solve this, users can simply make a model a permanent decision and not allow the model on a stream to change.

The authors claim that when making a model, there must be some belief about the phenomenon already. Otherwise sensor calibration and validation would be impossible, and no hypothesis could be formed. This poses a roadblock to using PoolView as it has been described for exploratory campaigns. The authors indicate that, in literature, there are methods for extracting a model as a sensor network learns more [36]. However, since the pool must make the modeling decision, it can only adapt based on aggregate as opposed to individual measurements. The authors also suggest that a low number of updates to the model could be sent out, but this must be balanced against the security risk of receiving an updated poor model, a risk previously mitigated by not allowing the model to change.

The authors indicate other techniques which include allowing clustering but protecting privacy via rotation, randomized response for non-quantitative data, and secure multi-party computation. Citations for these approaches are in the original paper. Secure multi-party computation is not feasible because the high communication overhead, which is not scalable. Scalability is a key requirement of sensor networks. In addition, the approach does not work with dynamic joining/leaving, such as what might happen if the system allows coarse privacy control as described at the beginning of this section.

1.6 Participatory Sensing Systems

So far we have discussed what participatory sensing is, and gone over many issues that arise when thinking about design for participatory systems. We have also introduced the idea of incentives, auctions, and how they relate to participation. Several approaches to privacy have also been discussed. We now examine two systems that are notable for taking several existing ideas and putting them together.

One often cited participatory sensing system is CarTel, which measured traffic and vehicle information through participants and devices interacting with their cars [1]. Cars were equipped with a GPS, making them into mobile sensors. Due to the fact that cars were driven by humans, human mobility and patterns are integrated into the system by design. One application of CarTel described in the paper was road traffic analysis. The authors found that users had heuristics which influenced their routes. They also found that travel times indicated routes were "reasonably predictable." As an extension, they observe that this means models of traffic delays should be possible to build. In the context of the chapter, it is important to understand that this means even vehicular mobility has patterns and these are influenced by human decision making.

In addition to humans being drivers for vehicles, they may act as data mules and physically transfer data by having an intermediate device such as a USB flash drive or operating a wireless AP that can be utilized by nodes in the system. We do not include an extensive discussion on delay tolerant networks (DTNs), however many of the ideas about multihop communication and muling are a common topic in DTNs. This suggests an avenue of research is studying human involvement in participatory DTNs. The authors mention that "unplanned in situ Wi-Fi networks can in fact be used with a delay-tolerant protocol ... such as CafNet..." with further details from their study in a separate paper [37]. The authors also observe that they are not the first to consider Wi-fi usage, citing Infostations as an example [38].

CarTel uses existing work in “mobile systems, sensor data management, and delay-tolerant networks”, and serves as a synthesis of ideas. This approach is worthwhile since, as mentioned several times in this chapter, traditional networks share many of participatory sensing’s challenges and a large body of research already exists on addressing these problems in the context of distributed or wireless sensor networks.

The system is composed of three main components

1. Portal (configuration, control, data storage)
2. ICEDB (delay-tolerant continuous queries)
3. CafNet (carry-and-forward network, delay-tolerant networking protocol)

The portal is the simple interface through which users are able to run applications which can issue queries to nodes and allow users to see the results. Queries contain information about which kinds of data are needed, at what resolution, frequency, and prioritization. The authors recognize that data may exceed bandwidth, and some data may be more important, or receiving summaries prior to the full data set may be desired. Intra-query and inter-query priorities are supported. Once queries are processed, data is sent back to the portal which stores data from nodes in a database. Applications use traditional SQL queries to extract data from the database as it becomes available. This design is of interest since it recognizes that there may be delivery delays, but applications may need results before all data has arrived at the portal. This allows perishable data to be consumed as it is available instead of potentially expiring while the system waits for more measurements, and removes a point of synchronization while still allowing a client-server architecture.

A web interface through the portal lets users look through the data. The authors emphasize visualizations with geo-coded attributes, which indicates the importance of location in their anticipated sensing campaigns. Since location plays a significant role in CarTel’s data, the authors claim that traditional search methods such as only using temporal locality may not be useful. Instead their interface incorporates location-based searching in the design by having operators and areas of interest that are defined graphically. The data available based on these criteria is then displayed. This sort of data can be viewed as a privacy risk. For example, with sufficiently dense geo-traces to infer identity, a user’s driving habits could be examined. Several illegal driving activities such as speeding or trespassing could be in the recorded data, and through this inference be tied to a particular user. To address privacy, the portal only allows users to look at their own data. However this means someone with access to the portal’s backend, or identity spoofing, could still compromise users. An alternate that the authors suggest is anonymously reporting data or reporting aggregates. Aggregated queries pose less of a privacy risk, but do not provide as in-depth data exploration opportunities. The authors note that the a limitation of CarTel is that there is not a way to aggregate results across users while maintaining privacy. Furthermore, they acknowledge that the correct queries could allow inference of users’ locations through targeted aggregate queries. Adapting CarTel to facilitate such queries is left for future work.

Due to bandwidth and connectivity constraints, it is not always possible to deliver all data in-order. Authors recognize that data may have different utility based on the application, so local and global prioritiziation are implemented. This gives each data tuple a score, with global prioritization done by sending a summary of the data and receiving information from the portal. The advantage of the global prioritization is that the portal can see data from all nodes and may be able to make dynamic decisions that an independent node could not. Global prioritization is justified in much the same way as global price vectors, described in the discussion of SORA in Section 1.4, were. From both these examples, it is apparent that in designing systems, adding a way for the system to influence or control network behavior allows global knowledge about measurements and node states to be utilized.

Local prioritization uses two constructs:

- PRIORITY - numeric priority with higher numbers indicating more important data. The authors suggest smaller data such as detecting an event be used for high priority, using raw GPS as an example of low priority data. This of course depends on the application.

- DELIVERY ORDER - specifies the ordering of results, this can be done by a field name or a user defined function that computes a score based on a tuple. Authors suggest the example of bisecting data for the purpose of constructing a curve - the rough shape of the curve is more important to deliver before smoothing the curve with additional data points. Simply using ORDER BY (a SQL construct that orders by a column's values numerically or lexicographically) cannot perform contextual ordering like that used in the bisecting function.

Global prioritization only uses SUMMARIZE AS, which specifies a mapping of tuples to summarized tuples by using grouping and aggregation functions. The portal computes an ordering, then sends the ordering to the node in question. The node then returns data ordered in the same way as the server sends it.

To keep general, the authors describe the use of software-based "adapters" which handle configuration of specific sensors on nodes and package measurements in a standardized format to be inserted into the portal database. Adapters are stored on the portal, and when required by an application they are sent to nodes. Nodes can run multiple adapters to regulate different sets of sensors for different applications. Similarly, CafNet has a layer called the "**M**ule **A**daption **L**ayer" (MAL) which allows any device to be used to transport data providing the appropriate MAL driver exists. While we do not discuss adapters or most technical implementation details, the authors go into much greater detail in their paper.

CarTel is a valuable system not only because it demonstrates a real-world deployment of a participatory sensing network, but because it also provides insight that can be applied to design. It is not necessary to create a system from scratch, but instead by studying past solutions, new systems can be synthesized based on requirements.

In CarTel, the goal was usage of distributed nodes through a simple interface with low latency. The use of cars in a participatory sensing system shows that participation can involve more than just carrying a dedicated node with limited power around on foot. Additionally the discovery that there are patterns in vehicular mobility is valuable in addressing coverage and limitations of participatory sensing even with vehicles.

Having explored the work of many other groups, we now examine the **P**rivacy, **P**ower, and **P**articipation-aware **A**uction **M**echanism (P3AM) which we developed in previous work [39]. Since this chapter emphasizes participatory sensing system design, we describe our approach first.

The initial decision was to consider the general task of a entity that wants to collect measurements with spatial and temporal information in each report. The information flows to a “data sink” which could be the controller, or a data broker which could sell the data independent of the system. In addition to viewing the system as synchronous, we considered that nodes would be some type of phone and thus the existing cellular networks’ infrastructure would be usable. This meant we did not focus on considering how data delivery happened, eliminating the need to design a protocol like CafNet, which was described in CarTel earlier in this section. Additionally, this allowed us to view cellular towers as “data sinks”, and assume that the service provider takes responsibility for the data upon arrival. This responsibility includes any privacy mechanisms, whether they be policy-based or a system such as the ones described in Section 1.5. Either type of privacy mechanism can be done on top of the system we describe, and does not affect using incentive for participation. Using cellular service providers was also advantageous because it meant instead of a single controller, there could be distributed controllers. In our experiments, we assume that towers do not communicate with each other, thus allowing for diverse smaller auctions based on locations and mobility patterns of the participants. Lastly, by using service providers, the system has a pre-established channel for distributing incentives.

The work by Lee and Hoh on RADP-VPC in Section 1.4 guided our general approach to using incentive. Like their work, we use risk-neutral adaptive bidding behavior, ROI to model participant tolerance to loss, the idea that bids should encapsulate perceived costs of the bidder. However, our approach does not use the idea of virtual participation credits, and instead of a single bid value to express concerns, P3AM takes a user’s valuation and modifies it. The modifications come from system defined curves describing the impact of battery level (to model node resources) and the time since a measurement was last accepted (to model privacy), and blends these with the participant’s valuation to produce a bid price. P3AM also has a parameter $P_{cheapest}$ which allows P3AM to operate differently than a first price auction by prioritizing a percentage of wins based on ROI. This incorporates participation preservation by the ROI model’s definition, while still allowing bidding to affect the probability of winning and the amount won in an auction round by a user. User understanding is something we believe is important, so P3AM is designed to be transparent and easy to understand. The functions of battery level and privacy are supposed to be simple functions to allow users to easily visualize the effects of a particular bidding strategy. System designers may consult directly with users when designing the incentive scheme, leading to participant involvement like that of participatory privacy discussed in Section 1.5.

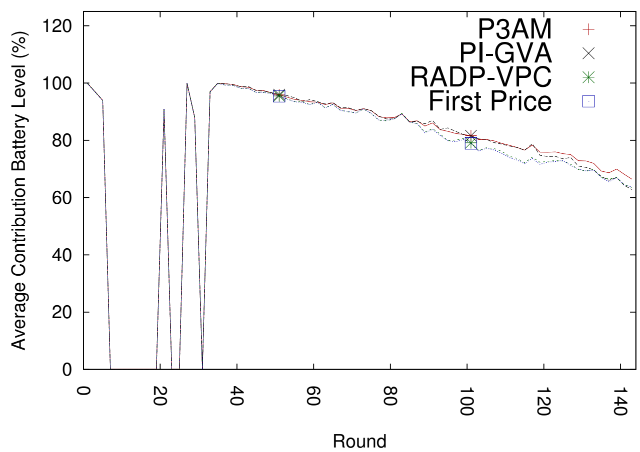
We also consider a second-price auction, PI-GVA [40]. With slight modifications to account for the reverse auction, PI-GVA is a valuable mechanism to compare to because it is designed for recurring auctions and is designed to be incentive compatible. Incentive compatibility makes the bidding decision simple for users who trust the system, since their true valuation will give them the highest utility. However, understanding the mechanism is difficult so users must know and trust that bidding their true valuation is the best action. Adding in factors such as battery level may change optimal bidding strategies and further complicate a second-price approach, so P3AM uses a first-price (where price is either the bid or the RoI of the user) approach.

In our previous study [39], experiments were run over a variety simulation parameters to examine how first price auctions, PI-GVA, RADP-VPC, and P3AM behaved. One of the assumptions we made that is not realistic was the use of random mobility. To define how participants move in a more realistic manner, we examined mobility traces of taxi movement in San Francisco [41, 42]. Taxis were chosen since their routes are indicative of paths taken by many individuals in the population, and thus are a good baseline for human mobility in a populated region. We consider a taxi to represent a participant with a single data source only when the taxi is active (i.e. a ride with a customer is in session). Since the taxi traces were originally spread over 3 weeks, we broke each trace apart into individual days of the week and overlaid them on a new 24 hour trace. In other words, there would be one trace for all Mondays for a given taxi, another for all Tuesdays, and so on. This yielded a set of movements that still kept correlation of days and hour of the day but was less susceptible to unusual trips being characterized as probable. 7606 such trips corresponding to the traces are used in simulation with one participant per trace.

The effects of mobility were clearly seen in the average battery level of contributors, shown in Figure 1.1. The fact that holes are present shows that using real traces is important to understanding availability of data sources with respect to both participation and location. The regions of holes is about 30 rounds in length which corresponds to 5 hours. This is roughly the 01:00 a.m. to 06:00 a.m. time period in which we would expect less participants to be active. Broadcasting a stationary location (likely a participant's home) for 5-8 hour span provides minimal information to the system while greatly increasing the potential privacy loss for that participant. These observations indicate that the use of taxi traces as a model of human mobility for populous regions is realistic, and highlights the importance of understanding human mobility when designing a participatory sensing system.

Even with extremely low payout settings and very intolerant ROI β, less than 5% of simulated participants stopped participating during the course of the experiments. As a result, the battery level behavior, average price per measurement and number of samples collected were dependent primarily on data source movement. Unlike the random mobility case where various parameters in mechanisms had effects on behavior, the only effect seen in the trace-based experiments was that the average price per measurement using PI-GVA grew sharply after the "region of holes" described

Figure 1.1: Average Contributor Battery Level: Each data point is the average (across all experiments) of the battery levels of all participants winning an auction during a given round. If no data source contributed during a round, there is a "hole" for that round.

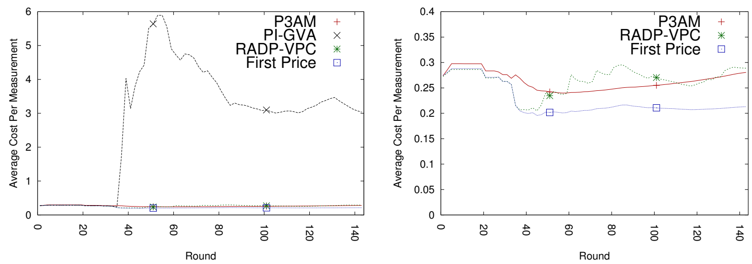


above. By the end of the day, PI-GVA's average price per measurement was still approximately 10 times higher than all other mechanisms, which had very similar average prices to each other. The average price per measurements are shown in Figures 1.2(a) and 1.2(b). Node based parameters, namely tolerance and true valuation, can cause the average price per measurement under PI-GVA to grow at a much faster rate than any of the other mechanisms we studied.

Due to the difficulty in creating participant dropout from ROI being unsatisfactory, we did not produce a trace-based case of explosion of incentive. However, the fact that parameters needed to be drastically different to produce such an explosion, and that using the same parameters as in the random mobility case we observed very different behavior, indicates that parameters are highly specific to the nodes' behavior. Since this is a participatory sensing application, this translates to needing to understand human behavior's impact on a system. Using testbeds or simulations is an approach that can be used to tune parameters prior to a full-scale deployment [12].

1.7 Conclusion

Participatory sensing is a type of sensor network applications which uses humans. These uses may include providing mobility to sensing and processing platforms (by



(a) (b)

Figure 1.2: Average Price per Measurement: Each point is the average across all experiments of the incentive awarded in exchange for a sample, regardless of if the sample was at a point of interest or not. (a) View of the full curves (b) Zoomed in view to compare P3AM, RADP-VPC, and first-price mechanisms

carrying hardware), acting as senors and processors (by deciding when to submit a report or by annotating data), and designing policies for sensor systems. Designing a system that involves human participation requires understanding challenges that arise because of human behavior such as patterns in mobility and difficulties in maintaining involvement.

In this chapter we have discussed only a sampling of the existing literature to explore some of the lessons in designing participatory sensing applications. The increased availability of powerful and versatile mobile hardware, coupled with a wide array of potential applications suggest great potential for growth in the field. Designing systems requires an understanding of the application, and making design decisions about various challenges, such as those described throughout Section 1.3. While our focus has been on incentivizing systems to maintain participation, a real-world deployment requires addressing many issues and combining work done in a variety of independent problems. As the topic of participatory sensing becomes more popular and more mature, we expect to see more effective systems that are more advanced and find innovative ways to address the multitude of design challenges.

[1]Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden, "Cartel: a distributed mobile sensor computing system," in *Proceedings of the 4th international conference on Embedded networked sensor systems*, SenSys '06, (New York, NY, USA), pp. 125–138, ACM, 2006.

[2] "Mobile millennium." `http://traffic.berkeley.edu`.

[3] "Urban atmospheres." `http://www.urban-atmospheres.net/`.

[4] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G.-S. Ahn, and A. T. Campbell, "Bikenet: A mobile sensing system for cyclist experience mapping," *ACM Trans. Sen. Netw.*, vol. 6, pp. 6:1–6:39, Jan. 2010.

[5] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet," *SIGOPS Oper. Syst. Rev.*, vol. 36, pp. 96–107, Oct. 2002.

[6] K. Shilton, J. A. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Participatory privacy in urban sensing." Retrieved from: http://www.escholarship.org/uc/item/90j149pp, April 2008.

[7] M. Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks," *International Journal of High Performance Computing Applications*, vol. 16, no. 3, pp. 293–313, 2002.

[8] K. L. Huang, S. S. Kanhere, and W. Hu, "Preserving privacy in participatory sensing systems," *Computer Communications*, vol. 33, no. 11, pp. 1266 – 1280, 2010.

[9] R. K. Ganti, N. Pham, Y.-E. Tsai, and T. F. Abdelzaher, “Poolview: stream privacy for grassroots participatory sensing,” in *Proceedings of the 6th ACM conference on Embedded network sensor systems*, SenSys ’08, (New York, NY, USA), pp. 281–294, ACM, 2008.

[10] J.-S. Lee and B. Hoh, “Dynamic pricing incentive for participatory sensing,” *Pervasive and Mobile Computing*, vol. 6, no. 6, pp. 693 – 708, 2010. Special Issue PerCom 2010.

[11] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, “Examining micro-payments for participatory sensing data collections,” in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp ’10, (New York, NY, USA), pp. 33–36, ACM, 2010.

[12] G. Mainland, D. C. Parkes, and M. Welsh, “Decentralized, adaptive resource allocation for sensor networks,” in *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation - Volume 2*, NSDI’05, (Berkeley, CA, USA), pp. 315–328, USENIX Association, 2005.

[13] S. Reddy, D. Estrin, and M. Srivastava, “Recruitment framework for participatory sensing data collections,” in *Proceedings of the 8th international conference on Pervasive Computing*, Pervasive’10, (Berlin, Heidelberg), pp. 138–155, Springer-Verlag, 2010.

[14] V. Krishna, *Auction Theory*. Academic Press/Elsevier, 2009.

[15] G. Pickard, I. Rahwan, W. Pan, M. Cebrián, R. Crane, A. Madan, and A. Pentland, “Time critical social mobilization: The darpa network challenge winning strategy,” *CoRR*, vol. abs/1008.3172, 2010.

[16] J.-S. Lee and B. K. Szymanski, “A novel auction mechanism for selling time-sensitive e-services,” in *E-Commerce Technology, 2005. CEC 2005. Seventh IEEE International Conference on*, pp. 75–82, 2005.

[17] J.-S. Lee and B. K. Szymanski, “Stabilizing markets via a novel auction based pricing mechanism for short-term contracts for network services,” in *Integrated Network Management, 2005. IM 2005. 2005 9th IFIP/IEEE International Symposium on*, pp. 367–380, 2005.

[18] J. Grossklags and A. Acquisti, “When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information,” in *Proceedings of the Sixth Workshop on the Economics of Information Security (WEIS 2007)*, pp. 7–8, 2007.

[19] P. Gilbert, L. P. Cox, J. Jung, and D. Wetherall, “Toward trustworthy mobile sensing,” in *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications*, HotMobile ’10, (New York, NY, USA), pp. 31–36, ACM, 2010.

[20] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micropayments for participatory sensing data collections," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, (New York, NY, USA), pp. 33–36, ACM, 2010.

[21] S. H. Clearwater, *Market-based control: A paradigm for distributed resource allocation*. World Scientific Publishing Company, 1996.

[22] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1. Cambridge Univ Press, 1998.

[23] K. Whitehouse, C. Sharp, E. Brewer, and D. Culler, "Hood: a neighborhood abstraction for sensor networks," in *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, MobiSys '04, (New York, NY, USA), pp. 99–110, ACM, 2004.

[24] E. Byrne and P. M. Alexander, "Questions of ethics: participatory information systems research in community settings," in *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, SAICSIT '06, (Republic of South Africa), pp. 117–126, South African Institute for Computer Scientists and Information Technologists, 2006.

[25] L. Palen and P. Dourish, "Unpacking "privacy" for a networked world," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, (New York, NY, USA), pp. 129–136, ACM, 2003.

[26] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in *Privacy Enhancing Technologies* (G. Danezis and P. Golle, eds.), vol. 4258 of *Lecture Notes in Computer Science*, pp. 36–58, Springer Berlin Heidelberg, 2006.

[27] G. Calandriello, P. Papadimitratos, J.-P. Hubaux, and A. Lioy, "Efficient and robust pseudonymous authentication in vanet," in *Proceedings of the fourth ACM international workshop on Vehicular ad hoc networks*, VANET '07, (New York, NY, USA), pp. 19–28, ACM, 2007.

[28] K. P. Tang, P. Keyani, J. Fogarty, and J. I. Hong, "Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, (New York, NY, USA), pp. 93–102, ACM, 2006.

[29] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, "Anonysense: privacy-aware people-centric sensing," in *Proceedings of the 6th international conference on Mobile systems, applications, and services*, MobiSys '08, (New York, NY, USA), pp. 211–224, ACM, 2008.

[30] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[31] J. Domingo-Ferrer and J. Mateo-Sanz, "Practical data-oriented microaggregation for statistical disclosure control," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 14, no. 1, pp. 189–201, 2002.

[32] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discov. Data*, vol. 1, Mar. 2007.

[33] H. Jian-min, C. Ting-ting, and Y. Hui-qun, "An improved v-mdav algorithm for l-diversity," in *Information Processing (ISIP), 2008 International Symposiums on*, pp. 733–739, 2008.

[34] D. Kotz, T. Henderson, I. Abyzov, and J. Yeo, "CRAWDAD data set dartmouth/campus (v. 2009-09-09)." Downloaded from http://crawdad.cs.dartmouth.edu/dartmouth/campus, Sept. 2009.

[35] A. Tikhonov and V. Y. Arsenin, *Methods for solving ill-posed problems*, vol. 15. Nauka, Moscow, 1979.

[36] C. Guestrin, P. Bodik, R. Thibaux, M. Paskin, and S. Madden, "Distributed regression: an efficient framework for modeling sensor network data," in *Information Processing in Sensor Networks, 2004. IPSN 2004. Third International Symposium on*, pp. 1–10, 2004.

[37] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and S. Madden, "A measurement study of vehicular internet access using in situ wi-fi networks," in *Proceedings of the 12th annual international conference on Mobile computing and networking*, MobiCom '06, (New York, NY, USA), pp. 50–61, ACM, 2006.

[38] T. Small and Z. J. Haas, "The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way)," in *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, MobiHoc '03, (New York, NY, USA), pp. 233–244, ACM, 2003.

[39] B. O. Holzbauer, B. K. Szymanski, and E. Bulut, "Socially-aware market mechanism for participatory sensing," in *Proceedings of the first ACM international workshop on Mission-oriented wireless sensor networking*, MiSeNet '12, (New York, NY, USA), pp. 9–14, ACM, 2012.

[40] J.-S. Lee and B. K. Szymanski, "A participation incentive market mechanism for allocating heterogeneous network services," in *Proceedings of the 28th IEEE conference on Global telecommunications*, GLOBECOM'09, (Piscataway, NJ, USA), pp. 2206–2211, IEEE Press, 2009.

[41] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD data set epfl/mobility (v. 2009-02-24)." Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility, Feb. 2009.

[42] M. Piorkowski, N. Sarafijanovoc-Djukic, and M. Grossglauser, "A Parsimonious Model of Mobile Partitioned Networks with Clustering," in *The First International Conference on COMmunication Systems and NETworkS (COMSNETS)*, January 2009.