Exploring End User Preferences for Location Obfuscation, Location-Based Services, and the Value of Location

A.J. Bernheim Brush, John Krumm and James Scott
Microsoft Research
{ajbrush, john.krumm, jws}@microsoft.com

ABSTRACT
Long-term personal GPS data is useful for many UbiComp services such as traffic monitoring and environmental impact assessment. However, inference attacks on such traces can reveal private information including home addresses and schedules. We asked 32 participants from 12 households to collect 2 months of GPS data, and showed it to them in visualizations. We explored if they understood how their individual privacy concerns mapped onto 5 location obfuscation schemes (which they largely did), which obfuscation schemes they were most comfortable with (Mixing, Deleting data near home, and Randomizing), how they monetarily valued their location data, and if they consented to share their data publicly. 21/32 gave consent to publish their data, though most households' members shared at different levels, which indicates a lack of awareness of privacy interrelationships. Grounded in real decisions about real data, our findings highlight the potential for end-user involvement in obfuscation of their own location data.

Author Keywords
Privacy, Location, Computational Location Privacy, Obfuscation, Anonymization

ACM Classification Keywords
K4.2 COMPUTERS AND SOCIETY: Social Issues

General Terms
Human Factors

INTRODUCTION
Location-aware mobile phones, GPS car navigation systems, and other location-aware devices have enabled a wide range of location-based services, such as providing navigational assistance or letting people share their location. While some location-based services can operate using just the user's current location, others require long-term location data, e.g. trace logs of GPS data, to infer a person's location routines. For example, movement patterns can be tracked to provide a personal environmental impact report [2], and location patterns can be used to automatically program home thermostats [27]. However, while people may benefit from services that make use of location data, it is important to consider the privacy risks of such services. In this paper we particularly address long-term location tracking as there are potentially increased risks and privacy considerations compared to one-time or intermittent sharing of location data.

To better understand people's concerns about the collection and sharing of long-term location traces and whether previously proposed location obfuscation methods might address these concerns, we interviewed 32 participants from 12 households as part of a 2-month GPS logging study. In addition, we asked them for permission to share their data publicly. While location privacy is a well-studied topic, due to the challenges in collecting location data, frequently hypothetical surveys are used to ask people about location privacy considerations (e.g. [6, 9, 30]). However, in our study the participants had collected actual location data, we showed them visualizations of their own data, and we asked them to sign (if they were willing) an actual legal consent form to share their data publicly.

Our study investigated the following research questions:

1. Willingness to Share Actual, Personal GPS data. Are participants willing to share long-term GPS logs? Does this vary based on whether the data is shared to the public, corporations, or academic institutions?

2. Appeal of Location Obfuscation Methods. Several obfuscation methods have been proposed for enhancing location privacy (surveyed in [21]). With short text and graphical overviews of the obfuscations, can participants map their own privacy concerns onto them? Which makes participants most comfortable sharing their location logs?

3. Desire for Location-Based Services. Are there location-based services requiring long-term location data which participants would find compelling enough to share their data with a company in order to receive?

4. Value of Location Privacy. Comparing with Cvrcek *et al.*'s study on the value of location privacy [8], we asked similar questions about how much money participants would want in return for collecting future data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
UbiComp 2010, Sep 26 – Sep 29, 2010, Copenhagen, Denmark.
Copyright 2010 ACM 978-1-60558-431-7/09/09...$10.00.

We found that many of our participants were willing to share their anonymized actual GPS data or trade it for location-based services. Comments made by participants about the obfuscation methods suggested they were able to understand them at a high-level and identify which would best address their privacy concerns. Mixing data to provide k-anonymity was the most preferred method, followed by Deleting of data near their home and Randomization. Although we interviewed participants from the same households together, we saw little evidence that participants considered their personal location privacy to be dependent on others in their household, and frequently participants in the same household had different obfuscation preferences. Overall, our study points towards the feasibility of more end user involvement in specifying obfuscation strategies to control the spread of their private location data.

RELATED WORK

In his survey of privacy in ubiquitous computing, Langheinrich highlights the importance of considering technical, legal, and social aspects of privacy [23]. Location privacy has been defined by Duckham and Kulik as "…a special type of information privacy which concerns the claim of individuals to determine for themselves when, how, and to what extent location information about them is communicated to others. [11]." There has been extensive work [e.g. 10, 26] in the interaction between how people manage social relationships and their privacy considerations.

The most common location-based services such as location-based search and navigation rely on relatively infrequent, not-always-on location tracking. Location-sharing services (e.g. Google Latitude, Loopt, Foursquare) have many privacy risks, as highlighted by the website pleaserobme.com, which calls attention to fact that using Foursquare.com to publicly sharing location data on Twitter can make clear when you are not at home. Interpersonal privacy preferences for location sharing have been well-studied, though mostly for occasional disclosures rather than for continuous location tracks [e.g. 1, 3, 7, 18, 31].

In this work, we address an emerging class of applications that sense a user's location continuously to provide services, but where the service does not necessarily show the locations back to the users or their friends directly. These include personal environmental impact (PEIR) [2], traffic jam detection [17], routine detection [24], home heating control [27], bus route planning [19], etc. For such applications, the privacy model and the amount of data disclosed to third parties may be less visible than for social location sharing services. Note that in our study we do not address situations where the user has little control of their data such as in commercial vehicle monitoring or Shklovki's *et al.*'s study of parolee monitoring [29].

In previous research on location-based services, Tsai *et al.* [30] evaluated 89 location-based technologies and showed that the majority with privacy policies collected and saved data (e.g. locations, profile information) for an indefinite amount of time. Others have identified attacks on long-term location traces that allow the subjects to be identified despite anonymisation [15, 20, 25]. In response, obfuscation techniques have been proposed which can be applied to location traces to make such attacks harder, e.g. see surveys by Krumm [21] and Duckham [11], and in particular work on k-anonymity [14], through mixing people's data [4], and other obfuscations [16, 20]. For this paper, we showed participants their own data and the effects of obfuscation techniques such as those above on that data, and present their feedback on who they trust to receive their data with various obfuscations applied.

Another interesting view on location data and privacy is that of the monetary value placed on it. Danezis *et al.* [9] and Cvrcek *et al.* [8] looked at the value placed on continuous location traces by individuals – we use similar questions to provide a comparison point in our study.

While many of our participants consented to share their data publicly, we are not the first to publish traces of location data. OpenStreetMap is a community effort to create copyright-free map data based on GPS traces contributed by users. However, individual traces are not available independently unless marked as "public". In work on Reality Mining [12], Eagle *et al.* collected and published a database of 100 users' mobile phone data over 9 months, including the current cell ID association which can be used for location sensing [5]. While there are clear privacy implications to sharing this data, the cell ID alone inherently provides some spatial obfuscation of location, with typical urban cells spanning hundreds of meters, while our participants' GPS tracks are accurate enough to identify particular buildings (e.g. home) that they occupied, possibly raising additional privacy concerns.

STUDY METHOD

We gathered data from 32 individuals in 12 households recruited by our company's usability group to address our research questions as part of a 2 month GPS logging study in Fall 2009. None of the participants worked at our company. During the study, participants carried small GPS loggers (Royaltek RBT-2300) which passively logged data every 5 seconds for later download to a PC. The GPS data was being collected to enable research on the potential to infer location routines and predicting when people might be at home as well as this research into privacy preferences.

Participants could opt out of tracking at any time by not carrying the device or turning it off. By using independent loggers rather than mobile phone based logging, we were able to recruit a wide range of participants without relying on people having a particular mobile phone, although people did have to carry an additional device. Participants were compensated during the study with 4 free software products per household (maximum value $600 USD each) and $0.50 per person per day of recorded data to encourage continued data collection.

At the first visit to each household, we explained what data was being collected, gathered study consent forms from participants, and asked them about their daily routine and home thermostat. We also told participants that as an optional part of the study we would ask them in the second visit if they would be willing to share the data publicly without their name attached, and we left a draft GPS consent form for their review. Participants were offered no additional compensation to share their data publicly. During the study they mailed their loggers to us every 2 weeks to send in the data.

At the second visit to each household, at least eight weeks after the initial visit, we collected their last set of loggers and then interviewed participants to better understand their location privacy concerns about the data we had collected. To make sure participants were aware of the data we had collected, we started the second interview by giving each participant three personalized maps: an overview and detailed maps for the two regions they spent the most time. These maps were based on approximately six weeks of their data since the last two weeks of data were still on the loggers we collected at this visit. We asked participants how it compared to what they expected, and if there were any surprises or locations that seemed to be missing.

We continued with a semi-structured interview comprising four parts. First, we tried to understand the conditions under which participants would allow us to share their data and to whom (e.g. public, corporations and academic institutions). Second, we asked participants which location services they would trade their GPS data for. Third, modeled on Cvrcek *et al.* [8] we asked participants for the payment they would want for collecting future data. Fourth, we asked the questions behind the Privacy Score metric proposed by Tsai *et al.* [30] to allow us to judge the privacy concerns of our participants.

At the end of the interview, we gave participants an optional data-sharing consent form to allow us to share their data publicly, after obfuscation of their home location by removing data inside a random, non-regular polygon around their home or other sensitive locations. We made it clear that the data would really be shared if they signed the consent form (and, indeed, the data has since been made available online[1]). Because of this we believe the answers participants provided should represent their true feelings about when they would be willing to share their GPS data.

Our approach mixed the advantages of a survey and semi-structured interview. For consistency, we gave each participant a paper-based questionnaire. However, to ensure that participants understood the questions we were asking and in particular the different ways their location data could be modified through obfuscation methods, the researcher present walked participants through the questionnaires; explaining each section, addressing any questions, and reviewing all answers with participants. This allowed the researcher to ensure participants understood our questions and ask follow-up questions to elicit further qualitative information. While household members were shown their individual data and had individual questionnaires to fill in, the interviews were conducted simultaneously in the same room and discussions between household members were permitted.

Interviews were recorded and transcribed. Due to a technical problem with the recording device, we obtained recordings for 8 of 12 households.

The Value of Location Obfuscation

To understand the value of different obfuscation methods proposed by researchers to address participants' perceived risks of sharing their location data, we presented descriptions and examples of five different obfuscations that could be applied to their data (see Figure 2). For each obfuscation we asked participants which of the following ways they would allow us to share their data:

Public with name: You will let us share your data on a public website *with* your name associated with your data.

Public anonymously: You will let us share your data on a public website, *without* your name associated with your data.

Corporate anonymously: You will let us share your data with our corporate partners, *without* your name associated with your data.

Academic anonymously: You will let us share your data to academic researchers for non-profit research, *without* your name associated with your data.

No sharing: You are not willing to let us share your data.

The split between corporate and academic sharing was a reflection of the same split in the location privacy survey by Cvrcek *et al.* [8]. Our expectation was that if participants felt the obfuscation addressed their perceived risks of sharing their data, they would be more willing to share their data. We believe our study is the first to ask end-users about different obfuscation methods.

From obfuscation methods present in the literature [21], we chose 5 obfuscations to present to participants. On the questionnaire each location obfuscation method was described on a separate page with a set of accompanying pictures, all of which are shown in Figure 2. During the interview we explained the obfuscations and their implications to participants using pictures, text and our own verbal descriptions as well as answering any questions. The obfuscations presented were:

Deleting: removing data near sensitive locations

Randomizing: adding Gaussian noise to locations

[1]http://research.microsoft.com/en-us/um/people/jckrumm/GPSData2009/

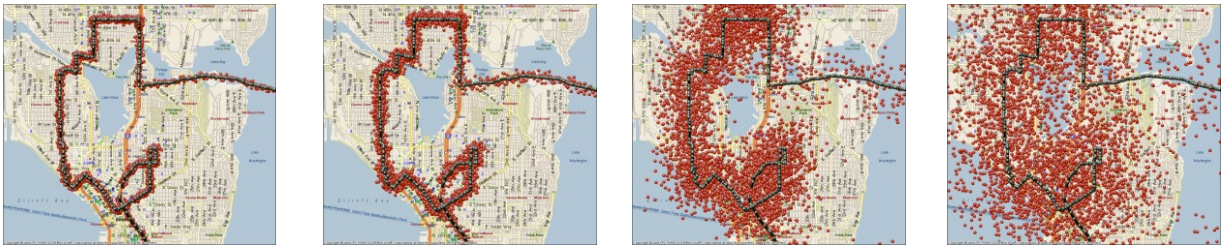
Deleting

Delete data near your home(s): Using a non-regular polygon all data within a certain distance of your home and other specific locations you select. This would help prevent someone from discovering where you live.

Condition	Public with name	Public anonymously	Corporate anonymously	Academic anonymously	No Sharing
a. Delete data within 100 meters of home(s)	☐	☐	☐	☐	☐
b. Delete data within 200 meters of home(s)	☐	☐	☐	☐	☐
c. Delete data within 500 meters of home(s)	☐	☐	☐	☐	☐
d. Delete data within 1000 meters of home(s)	☐	☐	☐	☐	☐

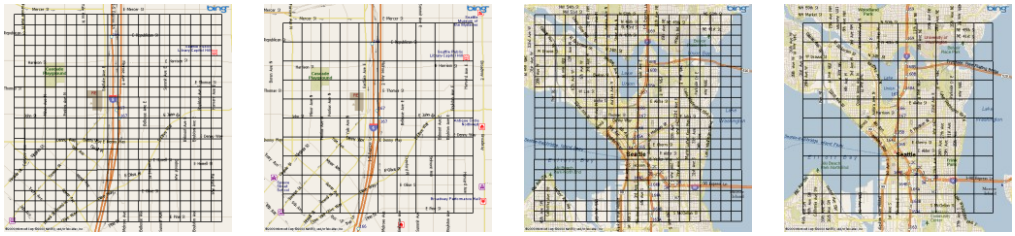
Randomizing

Randomly move each of your GPS points by a limited amount. The conditions below ask about progressively more randomization. This would make it harder for someone else to determine your exact location. See the maps below to understand the different amounts of randomization.



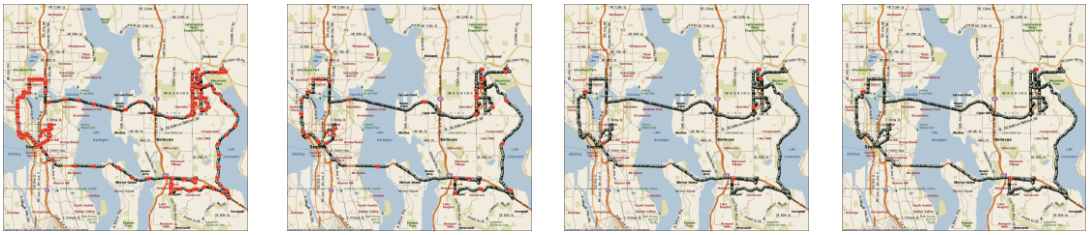
Discretizing

Instead of giving your exact location, give only a square that contains your location. Your exact location could not be determined, only that you were somewhere in the square. This would make it difficult for someone to determine your exact location. The conditions below ask about larger and larger squares.



Subsampling

Delete some of your data so there is gap in time between the points. Anyone who can see your data would only know about your location at certain times. The conditions below ask about different gaps in time.



Mixing

Mix your data with others. Instead of giving your exact location, give an area that includes the locations of other people. This means your location would be confused with some number of other people.

Figure 2: Text and graphics illustrating each obfuscation method and the various obfuscation levels available. For Deleting, one graphic was provided, and this was centered on that participant's home location, while the other obfuscations used four graphics and used the same graphics for all participants. The tickbox grid shown for Deleting is an example of grid shown for every condition. For Randomizing and Subsampling, the original trace is shown as a line and the data that would be shared as red dots.

Discretizing: quantizing locations on a lower-accuracy grid

Subsampling: providing locations at a coarser timescale

Mixing: reducing resolution to provide k-anonymity [14]

For each obfuscation we also asked about 4 different levels, so the user filled out a table similar to the one shown for Deleting in the first row of Figure 2 for each obfuscation. For Deleting, participants were asked about their willingness to share their location trace after deleting data within 100, 200, 500 or 1000 meters of their home or other pre-specified significant location. Each participant's questionnaire was personalized with circles at those distances from their house so participants could clearly see what data would fall within each distance. We reminded participants that although the picture showed the distance options using circles for simplicity, a non-regular polygon would be used to delete the data.

For Randomizing, participants saw the four maps shown in the second row of Figure 2, illustrating a GPS trace that had been randomized by adding Gaussian noise of 50, 100, 500 or 1000 meters standard deviation. For Discretizing, participants were asked about grid sizes of 50, 100, 500 or 1000 meters. For Subsampling they were asked about 1, 5, 30 and 60 minute gaps between samples, and Mixing asked about sharing regions with 10 people (the participants + 9 others), 50, 100 or 1000 people. Although people seemed to intuitively grasp the idea of mixing their data with other people, for this obfuscation we did have to explain that it assumes that everyone is carrying a GPS device so that there would be people available to mix your data with, and the size of the region would vary with population density.

For both the choice of whom they would share their data with, (e.g. public, corporate), and at what obfuscation level they would share (e.g. 50m, 100m) we enforced that participants be logically consistent in their answers. To illustrate by example, if the participant was willing to share anonymous data publicly after Discretizing to 100m, then logically they must also be willing to share the same data to academic and corporate institutions (who could download the public data), and to those recipients after Discretizing at 500m and 1000m (less revealing than 100m).

Participants

We collected data from 32 participants (16 M, 16 F) in 12 households, aged between 21 and 59 (median 27). Five of our households were comprised of housemates, six were families with children living at home, and one was a couple without children. Five households were renting (17 people) and seven owned their homes (15 people). Four of the families included children between the ages of 12-21 (5 children), however for legal reasons we could not share their data publicly, so they are not included in our 32 person study. Most participants had used location-based applications before; 28 of 32 had used GPS driving directions, 27 had used location-based web search, and 4 had used friend finding applications (e.g. Dodgeball).

RESULTS

To familiarize participants with their GPS data we showed it to them at the beginning of the interview. The median response to "How does your data compare to what you expected?" was "No Surprises". Participants who reported some surprises easily recalled the relevant events.

Many of our participants were willing to share their GPS trace data. At the end of the interview, 21/32 participants signed consent forms allowing us to publicly share an anonymized version of the data they collected during the study with data removed around their home. Also, as shown in Table 1, when asked about trading their long term location trace to a company for services that require such long-term information, many participants regarded services as worth giving up their information for. This was true both for personally beneficial services (e.g. personal traffic, home heating), and more altruistic applications for which anonymous data would be aggregated (e.g. bus route planning, traffic jams).

However, it would be incorrect to draw the conclusion that our participants did not have concerns about sharing their GPS trace data. We used Tsai's Privacy Score metric [30] to judge the privacy sensitivity of our participants and get an understanding of their particular concerns. A higher score indicates more concern about privacy. Our participants' median Privacy Score was 5.8 (mean 5.6, SD 1) out of a maximum of 7. They were particularly concerned about unauthorized secondary use and access to their data. Participants' median response was "Strongly Agree" that online companies should never share personal information unless it has been authorized by the individual, and online companies should take more steps to make sure unauthorized people cannot access personal information in their servers. Participant responses during the interview highlighted that many participants had real concerns about providing GPS traces, and no service appealed to all participants.

The tension between willingness to share and privacy concerns was highlighted in responses to a question on whether the benefits of making location data outweighed the risks. On a 7 point Likert scale from "The benefit far outweighs the risks" (1) to the "risks far outweigh the benefit" (7), the overall median was 3 (mean 3.28, SD 1.6), close to the center of the scale.

The focus of our study, on whether obfuscation methods can be directly comprehended by end users is therefore motivated – if users can lower the perceived and actual risks of using location-based services by choosing obfuscations to apply to their data that address their individual concerns, then the benefits of more location-based services will outweigh the risks.

Service	Yes (/32)
Help cities determine where bus routes should be to help the most people [19]	*30 (94%)*
Tell you about traffic jams before you get there [22]	29 (91%)
Tell drivers where traffic is slow [17]	*28 (88%)*
Control your home thermostat to save energy when you are away [27]	23 (72%)
Help businesses locate to high traffic areas	*23 (72%)*
Personalized estimates of your impact on the environment and its impact on you [2]	22 (69%)
Weekly summary of where you go and how much time you spend there	19 (59%)
Recommend local places you might like	19 (59%)
Plan routes that stick to roads you know	16 (50%)
Show you a map of where you traveled for every day, including vacations	13 (41%)
Give ads about businesses along your intended route	8 (25%)

Table 1. Responses to "Please indicate whether you would be willing to provide GPS data to Microsoft in exchange for that service. In order to deliver these services, Microsoft would associate your GPS data with a means of contacting you."

Preferred Obfuscation Methods

After the questions about the individual obfuscation methods, we asked participants which obfuscation method made them the most comfortable overall, and why. As Table 2 shows, Mixing was the obfuscation method preferred by the most participants (15), followed by Deleting (8) and Randomizing (7). Free-response reasons given for participants' preferences fell into four categories: Keep home private ("it doesn't give the exact location of our home" – E1[2]), Obscure identity (e.g. "my identity will be collective" – E2), Obscure location ("does not tie me to a specific location" – I1), and Keep data useful ("Most useful to other people while preserving privacy" – H3). Seventeen participants (from 9 of 12 households) also mentioned that obfuscating the home location was important in additional free-response questions about concerns.

These concerns were also supported by the comments participants made during the interviews. Concerned about home location, E2 explained she was not willing to share using Deleting because "the most outside circle was just ... 1000 meters, that's 1 kilometer, less than a mile so I thought that was too close." J1 pointed out an additional concern about his home location: "the information that this shows, it tells when we are home and we aren't home, that's sort of a security issue for me." Regarding anonymity, K3 said about Mixing, "once it [the mix] starts getting larger and larger there is no way to pinpoint who I am" and H1 reported "that one seemed harder to connect to an individual." F1 and F2, talking about Subsampling, felt 1 minute was too often because someone could track you, but as F1 said "5 minute intervals gives time to get away,"

[2] Participants are referred to by a letter representing the household, followed by a unique number.

Obfuscation method	Number of participants choosing method	Reasons given in freeform text			
		Obscure identity	Keep home private	Obscure location	Keep data useful
Mixing	15	8	0	1	1
Deleting	8	0	8	0	2
Randomizing	7	0	1	4	1
Discretizing	2	0	0	0	1
Subsampling	0	0	0	0	0
Total	32	8	9	5	5

Table 2. Responses to "Which type of modifications make you the most comfortable with sharing your data? (pick one)"

highlighting a desire for gaps in the data shared. These responses illustrate the three high-level concerns across participants were not disclosing home location, obscuring their identity, and not having their precise location reported, which are consistent with previous work by Tsai *et al.* [30].

Four participants stated that their reason for choosing their preferred obfuscation method was that they thought that choice would give the most value for users of the location data, e.g. corporations or academic bodies. For example, K1 said about Randomizing: "I felt like it maintained the integrity of the data while it also protected me." This somewhat altruistic desire may stem from the fact that they had already been compensated to share their location data for our research, and may have been thinking about how other researchers might make use of such data. Cvrcek *et al.* saw a similar response to their survey, where 30% of participants were interested in participating to improve mobile network quality rather than for personal gain.

Comprehension of Obfuscation Methods

As Table 2 shows, the reasons participants gave for selecting a particular obfuscation method as their favorite were for the most part highly suggestive of the properties of the individual methods as explained to the participants (e.g. Mixing is an obfuscation method that works by obscuring one's identity in a crowd of others, Deleting is oriented around keeping the home location secret, but regarding other locations as non-private, etc). This suggests that participants were able to comprehend the various obfuscation methods at a high level.

However, looking more closely at the sharing choices participants selected when asked about the individual obfuscations, we can see places were participants made choices inconsistent with their stated concerns. Although no participants picked Subsampling as their favorite method, 17/32 participants indicated they would share subsampled data at the first proposed level of obfuscation (1 minute periods). From this we can see that while participants may understand the basic operation of the obfuscation methods, the explanations given still do not allow them to understand the implications of the level of obfuscation well. In particular, with *any* level of Subsampling that we offered (1 minute, 5 minutes, 30 minutes, 60 minutes), just a single

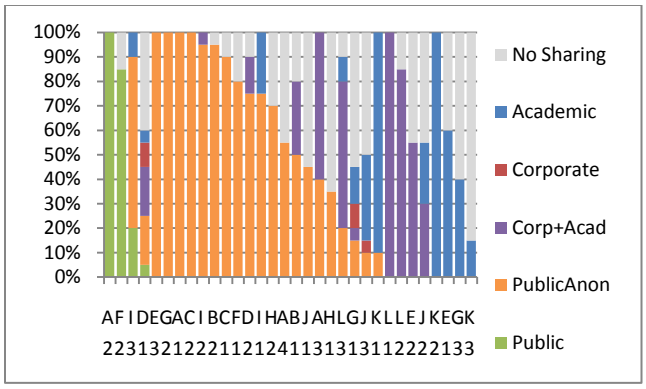


Figure 3. Comparison of participants' preferences for organizations to share their data with, across all obfuscations. X axis: participant, Y axis: percentage shared with each recipient type.

day's data would easily reveal one's home location. Thus, users concerned primarily with this (which many of our users were) should not choose Subsampling at any level.

This illustrates that while participants seem to have understood the basic operation of the obfuscation methods, the explanations given still did not allow them to understand the implications of the collection of data over time combined with the level of obfuscation. Our explanations could be improved by including more information about the impact of collecting GPS trace data over a long period. Another issue with our explanations is that since we provided a range of obfuscation levels, some participants may have assumed the highest obfuscation level equated to a high "level of privacy" (falsely in the Subsampling case). Future privacy control interfaces should be designed to avoid such assumptions.

In another freeform text question, participants were asked about whether they would prefer to use any other obfuscation or combination of obfuscations. While 24/32 gave no relevant answer and 3 said that their previous single choice was sufficient, 3 proposed a combination of Deleting and Randomizing, and 1 a combination of Mixing/ Randomizing. Interestingly, the 5 others all mentioned time in some way: 3 wished for a delay before data could be handed to a third party, 1 asked for a time limit to be placed on the use of location data, and 1 asked for the timestamps to be deleted from the data entirely. So while the effects of long-timescale data did not appear to be understood, other sensitivities due to time were apparent.

Who Participants Would Share With

We wanted to understand to whom participants would be willing to share and whether that would change based on the obfuscation method. Across the obfuscation methods participants specified to whom they were willing to share a total of 20 times (5 methods x 4 levels). For each choice there are six logically consistent answers: public with name, public anonymously, corporate+academic anonymously, corporate anonymously, academic anonymously, and no sharing. Figure 3 summarize for each participant how often they made each of the sharing choices.

Overall, 49% of the time our participants indicated a public sharing option (anonymous 42%, with name 7%), and 23% of choices were "no sharing." Thus, most of our participants appeared not to see any value in specifying whether corporations or academia could see their data separately compared to the public as a whole. Of the remainder, around 15% allowed both corporate and academic use, 13% specified academic use, and less than 1% specified corporate use. Also notable, 63% (20/32) of our participants only ever specified one of the options aside from "no sharing", i.e. they did not differentiate between obfuscation types. For choosing privacy policies, this implies that in many cases the user could make just two choices – the target sharing groups and then (independently) the preferred obfuscation option(s), rather than exposing users to the full parameter set that we showed to our study participants (which could be an "advanced" option). This makes such interfaces simpler and potentially more user friendly.

Consent to Actually Share Data

Of our 32 participants, 21 signed a consent form at the end of the interview to share their anonymized data collected during the study on a public website with a non-regular polygon removed around their house. We allowed participants to specify the size of the polygon on the consent form, and the median was 500 meters, (mean 622 meters, SD 550). Two participants asked for additional regions to be deleted, one around his workplace and the other around his girlfriend's house. This data is now available online.

Of our participants, men who were renters were more likely to consent to share. Of the 21 participants, 12 were men (75% of male participants) and 9 were women (56% of female participants). 14 of the 21 were renters (78% of 18 renter participants) and 7 were home owners (50% of 14 home owner participants).

Before asking participants to consider signing the consent form, we referred them to their answers to the Deleting obfuscation for public anonymous sharing. 22/32 of participants acted completely consistently in signing or not signing the consent form and specifying the obfuscation level (12 signed, 10 did not). 7 people who shared their data changed their obfuscation amount by a single level (3 choosing less obfuscation and 4 choosing more obfuscation). One participant (G3) chose to share at a level (1500m) that was higher than those offered in the original questions (which is consistent behaviour with his "no sharing" responses at the levels offered), one (L2) changed his mind from "no sharing" and consented to sharing with 1000m obfuscation, and one (E3) originally said he would share but chose not to sign the form, stating that he had no incentive to do so.

Thus, including 8 who made only minor changes, 30/32 (94%) of our participants were consistent between the questionnaire and their actual consent forms, giving us confidence in the questionnaire data.

Value of Location Privacy

Modeled after Cvrcek *et al.*'s survey, our questionnaire told participants we were considering a future study where location data would be collected using their cell phone, and due to a fixed and limited budget we were collecting bids to take part in the future study.

While we tried to make our questions similar to the previous survey, it is important to realize that there are a number of differences that might affect our participant's responses. First, our participants had just collected data and had been paid $0.50 USD per day and received 4 software gratuities (max value $2400), so they had some notion of how we might value their data. Second, our participants had just been shown their data and interviewed about ways we could obfuscate it, so we told participants the future study would delete data around their home. Third, we were asking for pricing on behalf of a corporation, in contrast to Cvrcek *et al.* asking on behalf of an academic research institution.

This section of the questionnaire seemed to be the hardest to answer and generated the most discussion between participants. Several participants found it challenging to value their location information and some tried to ask for information about what other people had bid or discussed with other members of their household. For example, K1, a 21 year student said "It's hard to name your price when you don't know what the competitive rate is." J1 said "I told you I was struggling, I work for cheap" once he saw his family members bids. E1 asked, "Can I change my prices, I don't feel competitive," when reviewing the questionnaire.

Most of our participants, 30/32, provided bids to collect 1 month of anonymous data and share with academic research institutions and corporate partners of our institution. Our participants' median bid was $100 USD for sharing with an academic institution for one month. This is almost twice as much as the €43 median bid collected by Cvrcek survey (about $55 USD based on exchange rates in August 2006), but that could be due to many reasons such as the nationality difference and our payment level for the current study. A clearer comparison can be made concerning whether participants' bids for selling data to a corporate partner increased two-fold as Cvrcek observed. We did not observe this for our participants. The median bid for selling data to a corporate or academic recipient was the same ($100), and 18/30 participants made identical bids for both.

Cvrcek *et al.* also studied the impact on bids of increasing the study length to 12 months and instead of a twelve-fold increase saw only a two-fold increase in median bid. 26 of our participants were willing to do both. The median bid for 12 months was $500, a 5 fold increase, although the mean of $2130 and SD of $4005 highlights that the participants had a wide range of bids. Thus similar to Cvrcek *et al.* we saw that many people discount their price for collecting over longer periods and one participant even commented "I'm going to give a discount for a longer period" (B2).

One of our questions asked participants their bid for selling their location data with their name. Fifteen participants gave bids for 1 month and 9 participants gave bids for 1 year which somewhat surprised us. The median bid for 1 month was $150 (mean $1135, SD $2520) and median for 12 months was $1400 (mean $1933, SD $1406). Participants who were willing to bid for sharing un-anonymized data were predominately male (11 of 15 for 1 month bids; 7 of 9 for 12 month bids) and often renters (9 of 15 for 1 month bids; 6 of 9 for 12 month bids).

DISCUSSION

We now discuss the strength and limitation of our study methodology, the feasibility of novel privacy control user interfaces, the lack of awareness of intra-household dependencies for private data, and the remaining challenges in anonymizing long-term traces.

Strengths and Limitations of Study Methodology

Our study explores privacy in different ways compared to previous work, by showing users their actual data, showing them the effects of actual obfuscation algorithms, and by asking users to sign a real consent form rather than hypothetical. However, our study also has confounding factors and limitations that the results should be interpreted against, which we discuss below.

Our participant group was quite small with 32 people in 12 households, due to the effort of collecting actual GPS data. The size leads us to report few statistical results, though we do report qualitative data as well. The group has some diversity in ages, genders, occupations and household types, but is geographically restricted to Seattle, WA, USA.

Our participants are people who had agreed to collect and share their GPS data with researchers in return for a fee. While they did not know when they were recruited that making the data public would be an optional part, they were obviously comfortable in sharing their data to an extent. However, we did find that many preferred not to share under some circumstances and that the level of sharing and what was important to protect were inevitably subject to thoughtful discussion.

Participation in the study may have also biased their opinion towards location-based services – since they were willing to give their location traces up for a sum of money, giving it up for a service is not farfetched. There may be particular bias towards efficient home heating control, since we described this as a motivating application. With the "altruistic" services, the participants' may have been biased by participation in the study, so they may have been more open to freely contributing.

There was no financial incentive to answer one way or the other in questions on sharing of their data – the participants had already been paid and further sharing was purely

voluntary. However, there may have been some bias due to the fact that we were physically there – while no pressure was placed to answer positively or negatively, they knew our research was based on such location data.

Enhancing Privacy Control User Interfaces

As Tsai *et al.* point out, location privacy policies are not easy for end users to understand and often do not address participants perceived risks. While our users had simple concerns such as "don't reveal the location of my home", they may not easily be able to map them onto legal terms in click-through disclosure agreements. In our study, we showed participants their own data and the effect of various obfuscation methods and levels, and participants used this in order to make a more informed choice about which obfuscation method best mapped onto their individual concerns. While our presentations did not capture the time factor as well as we would have liked, this can be remedied by further development of the visualizations, e.g. by using a dynamic visualization that presents behavior over time, or by presenting aggregations of longer periods of data and highlighting correlated data. Given such improvements, we believe that our study suggests that users can comprehend obfuscation methods' effects on their location data, and use this knowledge together to address their individual privacy concerns effectively.

The presentation of actual location traces requires a mechanism for showing the user their data with various obfuscations applied. While this can be done locally on the location-gathering client (e.g. the mobile phone), another method is to have one's location data managed by a trusted cloud service, e.g. Shilton *et al.* [28]. Such a service is responsible for monitoring the information about an individual, sharing it with authorized third parties, and modeling what can be mined with that information to ensure privacy rules are not broken, making it an ideal place to calculate candidate obfuscations for users.

Intra-household Privacy Dependencies Not Addressed

While we asked each household member independently about their sharing preferences, in actual fact if one household member's location trace became known, this has implications for other household members whose locations (e.g. home location, typical POIs visited) or behaviors (e.g. home/away state) may be correlated.

During the study, only household B discussed the potential interrelationship between their locations. Despite this, they gave different answers to questions (e.g. see Fig 2), and signed consent forms to share their data publicly with very different obfuscation levels – 100m (B2) and 1000m (B1). In fact, all of the 9 households where at least one member signed a consent form had different responses on the consent forms.

Household members also frequently expressed different preferences throughout the questionnaire, with 8 of 12 households differing on the preferred obfuscation method, and 8 of 12 households differing in their most-commonly-specified data recipient (no sharing, public, academic, etc). This implies that there may be relatively little awareness of the interrelationships between location data and of the importance of coordinated action to secure data jointly if it is to be secure for anyone. An analogy may be made to a shared PC – if any one user deliberately or inadvertently installs spyware, that spyware may be able to read data from all users of that PC.

Anonymizing Long-Term Traces Remains Challenging

While we explained five location obfuscation techniques drawn from the literature to participants [21], in practice there is significant work remaining to achieve robust long-term protection of private data using any of these techniques. To give just one example, Deleting of data around the home might seem safe, but as the work of Golle and Partridge shows [13], knowing a work location and only an approximate home location might be enough to uniquely identify an individual, and other behavior patterns are inferable from their movement patterns [20, 24].

We believe that safely obfuscating long-term location traces is a challenging problem that warrants further exploration. While solving these issues is beyond the scope of this paper, in our study we have explored the feasibility of informing users directly of the obfuscations available and allowing them to choose and thereby target their individual privacy concerns. In contrast, a non-user-involved system would likely necessitate using a generic obfuscation that imposes a more constrained notion of what constitutes sufficient privacy.

CONCLUDING REMARKS

In our study we showed 32 individuals from 12 households plots of their own location data, and also gave them brief visual, text and face-to-face explanations of five different obfuscation methods drawn from the research literature. 21 of our 32 participants signed consent forms to share their anonymized GPS data publicly.

Participants preferred different location obfuscation strategies: Mixing data to provide k-anonymity (15/32), Deleting data near the home (8/32), and Randomizing (7/32). However, their explanations of their choices were consistent with their personal privacy concerns (protecting their home location, obscuring their identity, and not having their precise location/schedule known). When deciding with whom to share with, many participants (20/32) always shared with the same recipient (e.g. public anonymous or academic/corporate) if they shared at all. However, participants showed a lack of awareness of the privacy interrelationships in their location traces, often differing within a household as to whether to share and at what level.

Our results suggest that we may be able to provide privacy control interfaces with simple explanations to empower users to make an informed choice about obfuscation based on their own privacy concerns. Future work could explore improving the explanations and visualizations, looking at how various obfuscations affect the quality of location-

based services delivered, and putting obfuscation controls in between real users and real applications.

ACKNOWLEGEMENTS

We thank our participants and George Danezis.

REFERENCES

1. Anthony, D., Henderson, .T and Kotz D., Privacy in Location-Aware Computing Environments," *IEEE Pervasive Computing*, vol. 6, 2007, pp. 64-72.
2. Agapie, E. et al. Seeing Our Signals: Combining location traces and web-based models for personal discovery. *Proc. Hotmobile 2008*.
3. Barkhuus, L. and Dey, A. K. Location-Based Services for Mobile Telephony: a study of users' privacy concerns. *Proc. Interact 2003*, 207-212.
4. Beresford, A. and Stajano, F. Location Privacy in Pervasive Computing. IEEE *Pervasive Computing*, 2(1):46-55, 2003.
5. Chen, M., et al. Practical Metropolitan-scale Positioning for GSM Phones. *Proc. UbiComp 2006*, Springer-Verlag.
6. Colbert, M. A Diary Study of Rendezvousing: Implications for Position-aware Communications for Mobile Groups. *Proc. GROUP 2001*. ACM Press 15-23, 2001.
7. Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J. & Powledge, P. Location Disclosure to Social Relations: Why, When, & What People Want to Share. *Proc CHI 2005*, ACM Press, 81-90.
8. Cvrcek, D. et al., A Study on the Value of Location Privacy. *Proc. Workshop on Privacy in the Electronic Society*. 2006, ACM 109-118.
9. Danezis, G., Lewis, S. and Anderson, R. How Much is Location Privacy Worth? *Proc. 4th Workshop on the Economics of Information Security*. Harvard University, 2005.
10. Dourish, P. and Anderson, K. Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena. *Human-Computer Interaction* 21(3), 2009, 319-342.
11. Duckham, M. and Kulik, L. Location privacy and location-aware computing, In *Dynamic & Mobile GIS: Investigating Change in Space and Time*, CRC Press, 2006, 34-51.
12. Eagle, N., Pentland, A. and Lazer, D. Inferring Social Network Structure using Mobile Phone Data. *Proc. National Academy of Sciences* 106(36), 2009, 15274-15278.
13. Golle, P. and Partridge, K. On the Anonymity of Home/Work Location Pairs. *Proc. Pervasive 2009,* 390-397.
14. Gruteser, M. and Grunwald, D., Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. *Proc. MobiSys 2003*, ACM Press, 31-42.
15. Gruteser, M. and Hoh, B. On the Anonymity of Periodic Location Samples. *Proc. 2nd International Conference on Security in Pervasive Computing*. 2005, 179-192.
16. Hoh, B., et al., Preserving Privacy in GPS Traces via Uncertainty-Aware Path Cloaking. *Proc. ACM CCS 2007*.
17. Horvitz, E., Apacible, J., Sarin, R. and Liao, L. Prediction, Expectation, and Surprise: Methods, Designs, and Study of a Deployed Traffic Forecasting Service, *Proc. UAI-2005*.
18. Iachello, G., et al., Control, Deception, and Communication: Evaluating the Deployment of a Location-Enhanced Messaging Service, *Proc. UbiComp 2005,* Springer-Verlag, 213-231.
19. Kostakos, V. Using Bluetooth to capture passenger trips on public transport buses. arXiv:0806.0874, 2008.
20. Krumm, J. Inference Attacks on Location Tracks. *Proc. Pervasive 2007*, Springer-Verlag, 127-143.
21. Krumm, J. A survey of computational location privacy. *Personal and Ubiquitous Computing* (2009) 13:6, pp. 391-399.
22. Krumm, J. And Horvitz, E. Predestination: Inferring Destinations from Partial Trajectories. *Proc. UbiComp 2006*, Springer-Verlag.
23. Langheinrich, M. Privacy in Ubiquitous Computing. In: *Ubiqutious Computing Fundamentals*, Edited by John Krumm, 96-156.
24. Liao, L., Patterson, D., Fox, D. and Kautz, H. Learning and Inferring Transportation Routines. *Artificial Intelligence*, 2007.
25. Mulder, Y., Danezis, G., Batina, L., Preneel. B., Identification via location-profiling in GSM networks. *Proc. Workshop on Privacy in the Electronic Society 2008*. 23-32.
26. Palen, L. and Dourish, P. Unpacking "privacy" for a networked world. *Proc. CHI 2003*, ACM Press, 129-136.
27. Scott, J., Krumm, J., Meyers, B., Brush, A. J., and Kapoor, A. Home Heating Using GPS-Based Arrival Prediction. *Microsoft Research Technical Report* MSR-TR-2010-19, Feb 2010.
28. Shilton, K., Burke, J., Estrin, D., Hansen, M., Govindan, R., & Kang, J. Designing the Personal Data Stream: Enabling Participatory Privacy in Mobile Personal Sensing. *Proc. TPRC*. September 2009.
29. Shklovski, I., Vertesi, J., Troshynski, E. & Dourish, P. (2009) The commodification of location: Dynamics of power in location-based systems. *Proc. UbiComp 2009*, ACM Press. 11-20.
30. Tsai, J., Kelley, P., Cranor, L., and Sadeh, N. Location-Sharing Technologies: Privacy Risks and Controls. *Proc. TPRC 2009*.
31. Tsai, J. Y., Kelley, P., Drielsma, P., Cranor, L. F., Hong, J., and Sadeh, N. 2009. Who's viewed you?: the impact of feedback in a mobile location-sharing application. *Proc. CHI '09*. ACM Press, 2003-2012.