Understanding How Visual Representations of Location Feeds Affect End-User Privacy Concerns

Karen P. Tang
Department of Informatics
University of California, Irvine
kptang@ics.uci.edu

Jason I. Hong, Daniel P. Siewiorek
Human-Computer Interaction Institute
Carnegie Mellon University
{jasonh, dps}@cs.cmu.edu

ABSTRACT
While past work has looked extensively at how to design privacy configuration UIs for sharing *current* location, there has not yet been work done to examine how visual representations of *historical* locations can influence end-user privacy. We present results for a study examining three visualization types (text-, map-, and time-based) for social sharing of past locations. Our results reveal that there are important design implications for location sharing applications, as certain visual elements led to more privacy concerns and inaccurate perceptions of privacy control.

Author Keywords
location, visualization, privacy concerns, location sharing.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors.

INTRODUCTION
The breadth of location-based services (LBSs) continues to grow and is aided by rapid advances in location sensing technology that is embedded in nearly every mobile phone today. LBSs are no longer just for navigational assistance or localized search results. Recently, LBSs have shifted away from simply *consuming* location to *sharing* locations; we refer to these LBSs as location-sharing applications (LSAs).

Many LSAs today operate by sharing the user's current location with those in her social network. To visually represent locations, these LSAs usually opt for a *map*-based visualization (e.g., Loopt, Google Latitude) or a *text*-based representation (e.g., Twitter). However, a growing number of LSAs also support sharing of location *feeds*, a historical list of a user's past locations. Figure 1 (a, c) shows examples of these feed-based LSAs, which all support *text*-based representations of location information.

The disadvantage of a text-based location feed is that it can be much more difficult to *extract* certain location properties, such as a place's spatial orientation. But the tradeoff with making these properties more salient is that information becomes more accessible to others, which introduces new privacy risks for end-users. This dilemma is similar to the privacy reaction that Facebook faced when it introduced its newsfeed feature. The newsfeed made it much easier to read through people's profile information, but the increased exposure led users desire for better privacy controls [11]. Thus, when proposing new visualizations for location feeds, it is important for visualizations to balance end-user privacy needs with the social and functional utility needed to support LSAs.

To better understand how visualizing location feeds can impact users' perceived privacy concerns, we interviewed 12 participants as part of a two-week long GPS logging study. There has already been much work done on location privacy, including work on computational location privacy [3], as well as work on designing usable privacy configuration interfaces for location sharing [7, 13, 18]. However, to date, there has not been any work done to explore users' privacy preferences and attitudes towards location *visualizations*. To address this issue, we collected actual location traces from our participants and conducted a within-subject comparison of three types of visualizations. Our study examines the following research questions:

- **Sharing Location Feeds**. Are participants willing to share their actual location feeds with others? Does this willingness change depending on what visualizations are used to present the location information to others?
- **Automated Generation of Location Visualizations**. We examine the practicality of automating visualizations for LSAs. We specifically look at the process for creating location labels using existing database sources.
- **Privacy Sensitivities towards Location Visualizations**. What makes a particular visualization more privacy invasive than another? Which visual elements, if any, lead to more privacy concerns from end-users?
- **Privacy Reasoning towards Sharing Visualizations.** How do users reason about the privacy-preserving properties of one visualization over another? Are users well-informed in their visualization choices?

We found that many of our participants were willing to share their location feeds, but different visualizations led users to *different* decisions about who to share their location feeds with. Based on interview feedback, participants

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'11, September 17–21, 2011, Beijing, China.
Copyright 2011 ACM 978-1-4503-0630-0/11/09...$10.00.

mostly preferred sharing map-based visualizations. Participants also expressed significant concerns about sharing temporal properties of their location feeds. In creating our visualization, we encountered several technical challenges that we put forth as future challenges for LSAs. We conclude our paper with a discussion of our observations about the interaction between privacy and location visualizations, paying particular attention to the implications for LSAs that wish to share location feeds.

RELATED WORK

Westin defined information privacy as an individual's claim to when, how, and to what extent information about them is shared with others [27]. Location privacy can be similarly framed. Past work has already provided a thorough discussion of how to design privacy feedback and controls so that users can more effectively specify their preferences for how their location information should be shared with others [13]. Equally important to specifying privacy preferences is the necessity for designing appropriate visualizations for conveying location information.

Work by Tufte and Cleveland have demonstrated how visualizations can influence people's interpretation of data [6, 24]. We posit there are similar effects with location visualizations. But, unlike charts, visualizing one's location (and sharing those visualizations with others) has important privacy implications for users. Prior work has identified many ways in which privacy considerations can affect one's social relationships [17]. However, these studies mostly examine privacy concerns with an implicit understanding that location is shared occasionally, rather than continuously [1, 7, 13, 14, 18]. Also, most of the previously studied LSAs have focused primarily on sharing of *current* locations. Some LSAs support varying degrees of historical logging, but it is usually not its primary goal; thus, little attention has been shown for sharing of *historical* locations.

Our study is most similar to work done by Brush et al [3], who examined users' preferences for sharing location traces. Though we both use location visualizations in our study, Brush et al. primarily used them to convey different types of location obfuscation techniques. Users then specified their obfuscation preference when sharing past locations with a commercial entity (e.g., Microsoft). In our study, location visualizations are examined in much more detail, as we probe how different visual elements influence users' decisions to share their past locations with others.

Current LSAs predominantly represent locations with text or a map (Figure 1b). Each representation has its pros and cons, particularly when applied to sharing of historical locations. On one hand, text can allow easier integration into existing information newsfeeds. However, textual representations of location are often less informative, as it buries many properties of location feeds. For example, while text affords a more compact visualization, it is harder to determine the spatiality of a place. Maps are better suited for this, but they are not without fault. For example, Figure 1d shows a mash-up of Google Maps and Foursquare's

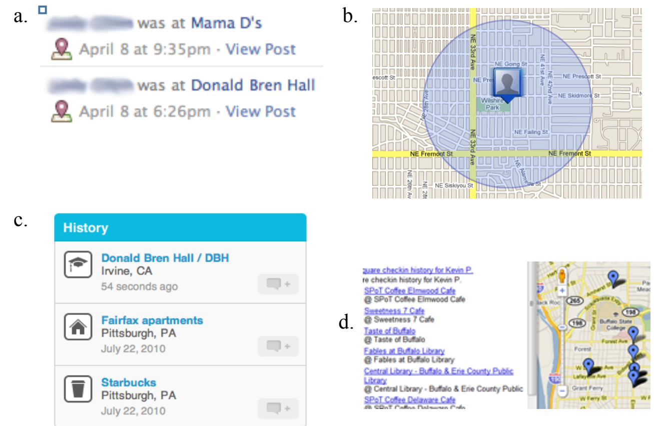

Figure 1. Examples from current LSAs that use text (a, Facebook Places; c, Foursquare) and maps (b, Google Latitude). (d) Mash-up of Foursquare (text) & Google Maps.

check-ins (text). Having the map instantly provides more spatial awareness, but it is difficult to discern the temporal properties (like the order in which the places were visited).

PROPERTIES OF SHARING LOCATION FEEDS

Adding more information to either a text- or map-based visualization is not difficult. In this section, we describe different dimensions of location feeds and how we represented them in our study. Note that our goal is not to design an optimal visualization. Instead, we focus on creating different styles of visualizations, making them as isomorphic as possible (i.e. having the same content), and then comparing how users interpret and react to them from a privacy perspective. This kind of analysis is new to the domain of sharing *historical* locations and can provide important insights for future LSAs, particularly given the rather homogenous types of LSA visualizations seen today.

When sharing current locations, we consider the following:

1) **Spatiality**: a place's physical location, in absolute terms (GPS coordinates) or where a place is located relative to a known landmark (e.g., "in south LA")
2) **Label**: how a user refers to the place (e.g., the user at "Starbucks", "a coffee house", or simply "in LA")
3) **Arrival**: when a user arrives at a place

Unsurprisingly, there is only one temporal variable that's accessible when sharing current location (arrival). When sharing location *feeds*, there is a richer set of temporal information that can be shared, including:

4) **Departure**: when a user leaves a place
5) **Duration**: how long a user stays at a place
6) **Sequence**: the order in which a user visited each place
7) **Frequency**: how often a user visits a place

There are certainly other features that one could include in a location visualization. However, for our study, we limit our exploration of the design space to these seven features.

Generating Location Labels

The second property in our list (a location's *label)* presents an interesting layer of complexity that is worth describing in more detail. Existing LSAs provide varying levels of

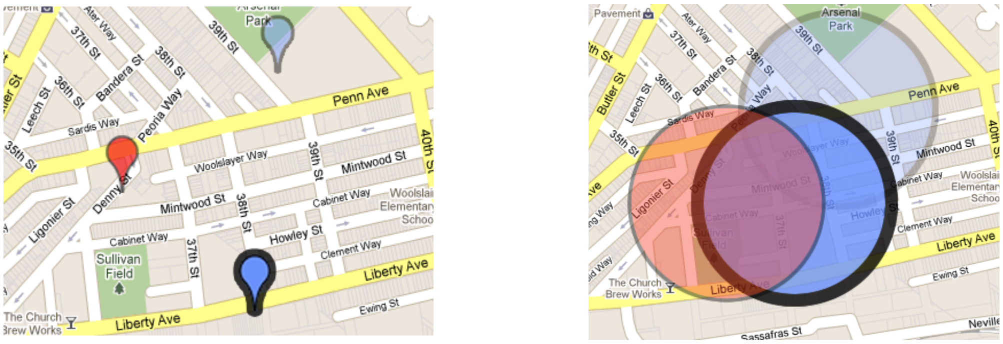

Figure 2. Shows three visualization techniques (post-pilot). The two marker styles represent same locations; (left) halos use random blurring technique, where midpoint is not necessarily actual location. Both images show transparency (to show sequence, faintest=oldest), color (red=most recent place), border width (thickest=most frequented place)

support for creating location labels. Many LSAs, like Facebook Places and Foursquare, operate by having users specify the business name or a personal label for each place they check-in at. These labels are then implicitly associated to a physical place through an address or GPS coordinates.

Past work has proposed that location labels can be classified using a taxonomy that distinguishes between geographic and semantic descriptions [16], and general and specific descriptions [22]. In our study, we use these two dimensions to create four different types of labels:

- **General geographic labels**: referencing the city and/or neighborhood that a place is in (e.g., "I'm in NYC")
- **Specific geographic labels**: a street address or an intersection (e.g., "I'm at Fifth & Main")
- **General semantic labels**: referencing the kind of place somewhere is (e.g., "I'm at a coffee shop")
- **Specific semantic labels**: a business name or specific place name (e.g., "at Starbucks", "at the White House")

Based on these categories, we note that LSAs with manual check-ins are using two types of labels: a specific geographic label (because check-ins are always tied to an address or GPS coordinate) and a semantic label (whatever string the user provides to describe their check-in). However, the types of semantic references that users provide vary greatly [5]; thus, in our study, we provide an initial exploration to see if different place descriptions affect privacy preferences and, in particular, whether labels change how users perceive different location visualizations.

VISUALIZING LOCATIONS IN TEXT, MAPS, AND TIME

Existing LSAs use either a text- or map-based visualization, and LSAs that support sharing of historical data typically only use textual representations. In our study, we study both of these types to better understand how location visualizations affect users' privacy preferences. We also include a third (time-based) visualization type.

Text-Based Representations

In text-based representations (**Figure 3**a), each row has a timestamp showing arrival and the length of stay. We chose to use relative timestamps ("5 minutes ago") for times that are < 3 hours from when the visualization is viewed. For visits that occur before then, we use absolute timestamps ("13:25pm"). This mimics current LSA behaviors (Fig 1c).

What is visually important about the text-based representation is that every location is treated the same. In other words, every row in the visualization is visually no different than any other row. The text-based representation emphasizes specific temporal features of location feeds: the *sequential order* of past visits and the *arrival* information for each place. Thus, we can use this visualization to probe whether these properties are an important factor in how end-users reason about their privacy concerns for sharing their historical locations with others.

What is less obvious in a text-based representation is the frequency that one visits a place and the spatial properties of each place. However, one could glean this information by analyzing the historical data, manually gathering appropriate statistics (for frequency information) or reverse geocoding location labels (for a place's spatial property).

Map-Based Visualizations

Most map-based visualizations (Figure 1d) mark a user's location using a pushpin at specific GPS coordinates. This style is conducive for describing locations with geographic labels. However, to support different kinds of semantic labels, we also use a halo marker (**Figure 3**b) that shows a user's true location as somewhere within the boundaries of the halo. The halo is more ambiguous and supports plausible deniability, which past work has shown to be an important privacy feature for LSAs [14]. When adding a halo to a map, we randomly add noise to the true location up to +/- 300m (Figure 2), reflecting a blurring of up to three city blocks (which, on average, are ~100m long). Generally speaking, this amount of blurring is most useful when using semantic labels. Given the urban area that we conducted our study in, we feel that a 3-block blurring radius provides sufficient support for plausible deniability.

While randomized blurring affects the location of the halo marker, we also manipulated the size of the halo to account for potential inconsistencies between location labels and the preciseness of the halo marker. For example, with specific geographic location labels, it does not make sense to have large-sized halo markers, as the label's precise description takes away the possibility for plausible deniability. Thus, we adjusted the diameter of the halo markers depending on the level of specificity provided by the location label. For markers associated with general geographic labels (e.g., city or neighborhood), we set the diameter to be two miles wide. For halo markers associated with semantic labels (e.g., “coffee shop”, “Starbucks”), we set a smaller diameter (0.25 miles) and used the randomized blurring technique. The exact sizes of the halo’s diameter were chosen based on the particular city in which our study was conducted and was influenced by the typical size of neighborhoods in the area as well as the surrounding area’s urban density.

We added basic interactivity to our map-based visualization, letting users click on the halo markers to show temporal features like when the user arrived and left, and how long the user stayed (per visit). The pop-up window also includes one of the four location labels we previously described. To show the remaining temporal variables (sequential order and frequency), we manipulated the marker’s visual properties. Card et al [4] and Ware [26] argue there are at least nine visual properties that can convey information: an object’s position, size, orientation, grayscale, color, texture, shape, animation, & transparency. Using these design guidelines, Figure 2 describes some of the visual mappings we used in our study.

Generally speaking, the most salient feature for map-based visualizations is the *spatial* information. By visually manipulating the marker’s transparency, size, and border width, we have also emphasized certain temporal features. However, the set of markers and their placement on the map is arguably more prominent in comparison. Thus, we can use this visualization to probe how sensitive users are to sharing the spatiality of their historical locations.

Time-Based Visualizations

We also introduced a third visualization type, as we noticed that there were still some temporal features (duration and departure) not emphasized in either the text- or map-based visualizations. In the time-based visualization (Figure 3c), we use a timeline and color-coded blocks to show when a user arrives and leaves a place. The colors of each block are randomly assigned. Similarly colored blocks indicate that the user has returned to a particular location.

The time- and map-based visualizations are isomorphic, containing the same information. Users can click on a particular block to show a map with a halo marker placed according to the same rules as for the map-based visualization. It is important to note that, in addition to the map, we also show the precise arrival, duration, and departure information to ensure that we provide the same descriptiveness in both visualizations.

In the time-based visualization, the most salient features are the colored blocks in the timeline, which show how much time a user spends at their past locations. The visualization draws the user’s attention to locations with repeated visits because of the similarly colored blocks. On the other hand, spatial information is less emphasized, as users must click on the colored blocks to see map-related information. Thus,

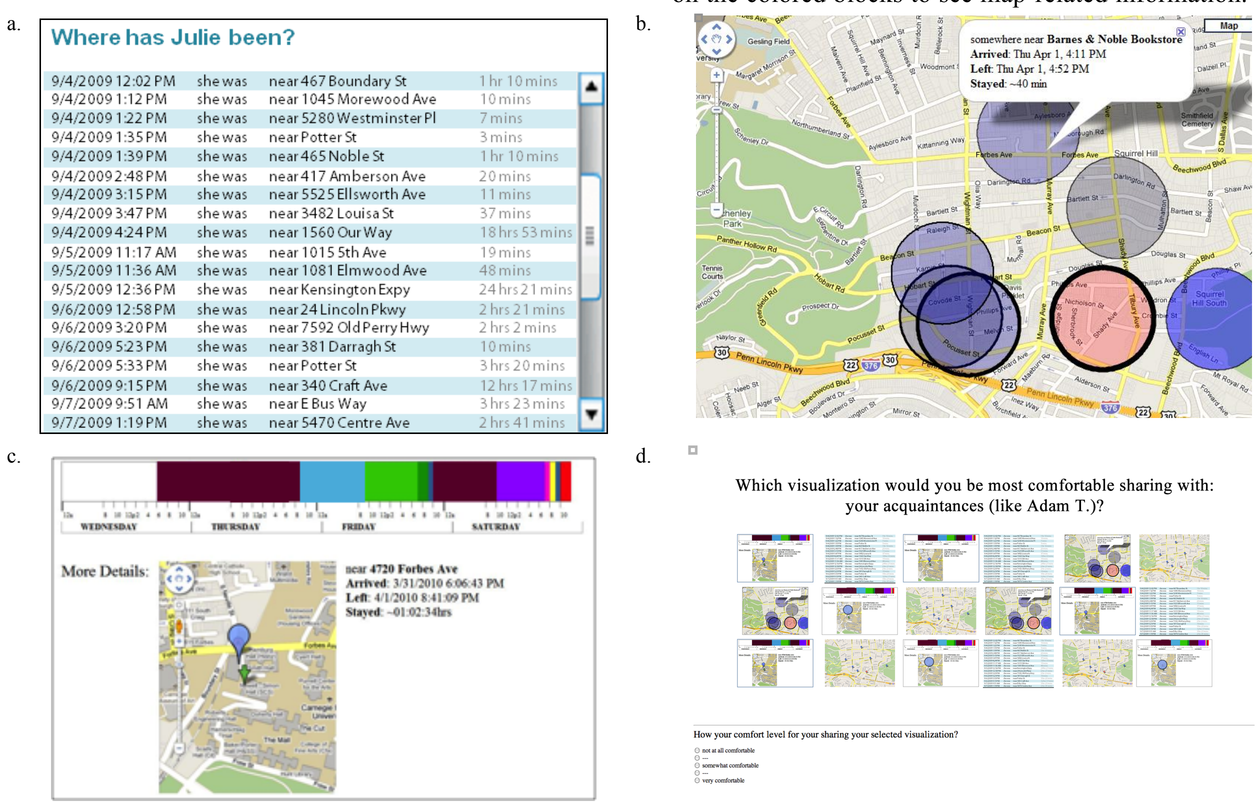

Figure 3: Examples of our 3 visualizations, post-piloting. (a) Text-based shows arrival, labels, & duration. (b) Map-based shows arrival, departure, duration, labels, spatiality, frequency, & sequence. (c) Time-based shows same features as (b). Visualizations made to be isomorphic. (d) Prompt shown to participants for choosing and evaluating visualizations.

we can use the time-based visualization to probe whether the duration and arrival/departure properties adversely affect users' privacy preferences for sharing location feeds.

PILOT STUDY FOR REFINING VISUALIZATIONS

While our text-based representation is straightforward, we wanted to make sure that our map-based visualization would not overwhelm our users or be difficult to comprehend. Thus, as a sanity check, we conducted a pilot study with 30 users to provide feedback on a sample screenshot of a map-based visualization.

Each visualization contained week-long location feeds for three individuals. The map-based visualization showed the three location trails, differentiated by differently colored markers. In the text-based representation, the three location feeds were intertwined, ordered by arrival times and demarked by differently colored text. These two visualizations were designed to be isomorphic. That is, one could convert the text-based representation into the map visualization (and vice versa) with some effort, but without needing any additional information. The order of the visualizations was counter-balanced. Participants were asked to explain each visualization to the best of their ability and choose which visualization they would be most comfortable sharing, if it were incorporated into a LSA.

Based on participants' qualitative feedback, we found that differences in transparency levels were not easy to discern. Participants reported that it was hard to pick out the most opaque marker (the most recent location) for each of the location feeds. We referred back to Card [4] and Ware's [26] suggestions and opted to use a combination of color and transparency to indicate sequential ordering of location visits. To represent the most recent location, we used a red-colored halo. All other markers were marked using a different color (blue); for these markers, we continued to use transparency levels to indicate older locations (Fig 2).

One implication for this design choice, though, was that we could now only display one user's location feed at a time (as color is now associated with sequential order and cannot be used to indicate different location feeds). In light of this change, we decided that the thickness of a marker's border (how often a user visits a place) would be computed in relative terms. In our original visualizations (as shown in the pilot study), we mapped border widths (thin, medium, thick) to certain frequencies (1-2, 3-6, and 7+ times), which were the same across different location feeds. However, since our new visualizations show only one user, we opted to calculate the border widths based on percentages instead. For example, if a user visited three places equally often (say, 5 times a week), then there would have originally been five markers, all with medium-width borders. In our new scheme, the markers each have a thin border, as the user spent equal time at each place. By using percentages, we normalize the visual features so that there are fewer thick borders in the visualization, reducing visual clutter. This change was also in response to participants' worry that our visualizations might be overwhelming with more data.

Our participants reported a clear preference for the map-based visualization for *viewing* their friends' locations (76.7%). However, when picking which visualization they would like to *share* with their friends, their preferences shifted to the text-based representation (63.3%). This statistically significant difference ($p<0.01$) suggests that there are indeed important privacy considerations for LSAs that go beyond designing adequate privacy controls and are related to how information is *visualized* to others. Armed with these results, we redesigned our visualizations and conducted a second, more controlled study.

STUDY DESIGN FOR EVALUATING VISUALIZATIONS

In our main study, our primary manipulation was between the three types of visualizations (text-, map-, time-based). For each visualization, we varied the location labels (general or specific, geographic or semantic). Within the map- and time-based visualizations, we also varied the marker type (the traditional marker style or a halo). In total, there were twenty possible combinations of these visualization variables (3 visualization types x 4 label types x 2 marker styles – 4 since text visualization does not use markers). In this study, we evaluated 18 combinations. We excluded two visualizations since they present an illogical combination, where the halo marker is paired with a specific geographic label. The intention behind using the halo is that, by providing a larger marker, users are afforded plausible deniability. However, when using the halo with a fully precise label (like an address), the halo's ambiguity is much less useful. As such, these two pairings are excluded.

Participants

Using a university-wide mailing list, we recruited twelve participants, ranging from 23-51 years old ($\mu=30.8$, $\sigma=6.2$). Five participants were female. Seven participants were graduate students, the rest were university staff members. Half the participants were from non-technical backgrounds. We did not advertise that our study was about privacy, nor did we mention this in the entrance survey. Instead, participants were recruited under the pretense that they would be evaluating different information visualizations.

Entrance Survey and Data Collection

Participants completed a 10-minute survey to collect basic demographic and social network information. For their social networks, participants provided examples (names) for four relationship groups: family, acquaintances, supervisors, and close friends. We told participants that their examples must live in the same city. We later used these names to frame the privacy questions for evaluating the visualizations. By controlling for geographical distance, we avoid it potentially biasing participants' preferences.

To ensure that participants realistically considered their privacy concerns when evaluating the visualizations, we collected two weeks of *actual* GPS traces from each participant. Participants were given mobile phones (Nokia N95s) to carry and use as their primary phone during the study period. This incentivized them to keep their phones on at all times, though participants were not penalized if they momentarily turned it off for privacy reasons.

The phones were equipped with location-logging software to record participants' actual location traces (modeled after the software used in [2]). The software ran continuously in the background and collected GPS and Wi-Fi positioning data. To reduce power consumption, we used the phone's accelerometer to selectively sample locations. The logging data was stored on the phone and we required participants to upload their data (through a website) twice during the study: midway through and at the end.

Automatically Generating Location Labels

Before each interview, we analyzed each participant's location trace. For times when there was no GPS readings (e.g., when indoors), we used Skyhook's API [21] to translate the Wi-Fi readings into GPS coordinates. We then computed the distance and speed between consecutive GPS readings to determine if the participant was moving. Places that the participant stayed for more than five minutes were marked as "significant". Using the coordinates for each significant place, we *programmatically* generated the four label types (geographic vs. semantic, general vs. specific).

To generate the general geographic label, we queried a publicly available database to first reverse geocode the GPS coordinates to a zip code. We then used the zip code to lookup the nearest neighborhood. To generate the specific geographic description, we used Geonames [9] to perform reverse geocoding using their `findNearestAddress()` and `findNearestIntersection()` webservice calls.

To generate the semantic labels, we used the specific geographic labels (the street address & nearest intersection) to query the Google Maps API and obtain a list of the nearest POIs. Each POI result includes information about the type of place it is (e.g., "Restaurant", "Shopping") and the name of the place (typically a business name, like "Starbucks"). We record the top result, as the API dictates that it should be the POI that is closest to the address provided. As an alternate source of labels, we also queried other publicly available sources, including Microsoft's Mappoint webservice and Wikipedia. To use Wikipedia, we first scraped Wikipedia for their geo-tagged articles and then used these tags to create a local database mapping GPS coordinates to corresponding article titles (which would serve as the location label). Among the three database sources, we choose the label belonging to the POI closest to the given address, within the bounds of the halo marker.

The challenge behind automatically generating location labels is that there is no guarantee that the generated label is actually correct. In fact, there are several ways in which the label generation process is susceptible to errors.

- **Sensing Errors**: All generated labels ultimately depend on having accurate GPS coordinates. However, there are times when GPS readings are not available and we must rely on Wi-Fi readings, which can be less accurate. Also, switching between GPS and Wi-Fi sensing consumes a non-trivial amount of battery power [25]. Thus, even for the most diligent participants, there were times when their phones did not record data due to poor battery life.
- **Triangulation Errors**: For Wi-Fi readings, we rely on Skyhook's API to generate appropriate GPS coordinates. This process is, by definition, only an approximation of the user's true GPS coordinates. The accuracy of these coordinates is also highly dependent on how up-to-date Skyhook's database is. For a more detailed discussion of the shortcomings of Wi-Fi localization, we refer the reader to any one of several Place Lab papers (e.g., [19]).
- **Interpolation Inaccuracies**: Even with perfectly accurate GPS coordinates, our process relies on webservices to provide accurate reverse geocoding. However, by definition, reverse geocoding only returns a best estimate. For example, to determine the exact street number for a given set of coordinates, reverse geocoding often relies on interpolating between two known street addresses. Thus, slight variations in GPS coordinates can result in very different reverse geocoding results.
- **Sparse and/or Stale Databases**: Assuming that we can record an accurate GPS reading from the phone and we can retrieve a perfectly interpolated address via reverse geocoding, we still depend on public database sources to be up-to-date. Since we use these databases to find the nearest POIs (and their associated semantic names), we are essentially at the mercy of these services. Two problems that we frequently encountered in running this study are: 1) some databases contain out-of-date entries, and 2) databases often have fewer entries for suburban areas. In both cases, the generated label will be incorrect and, in the latter case, the generated label may end up outside the boundaries of the halo marker (resulting in no label being added to our visualization).

Based on several weeks' worth of pre-study test data, we were able to consistently and reliably translate the recorded GPS sensor readings into *geographic* place labels (with 98.7% accuracy) using the process we described. However, automatically generating *semantic* place labels proved to be much more difficult. For this study, since creating place naming algorithms (e.g., using machine learning techniques) was not the goal of this study, we opted instead to use a human-in-the-loop approach to get accurate labels.

User Validation of Location Labels

We required participants to verify our generated labels twice: midway through the two-week data collection and at the end of the study. Once participants uploaded their sensor readings through our website, we extracted the GPS coordinates, identified the significant places, and used our previously described steps to automatically generate the two labels (specific semantic and specific geographic names) for each place. We then emailed this list of labels, along with their corresponding arrival times, to each participant. In the email form, we asked participants to verify and correct, if necessary, any obviously incorrect labels. We also asked participants to provide labels for any locations that did not appear in the list (but should have). In these cases, the phone may have been off, resulting in there being no sensor readings, or our heuristics may not have generated any labels due to database errors.

This human-in-the-loop approach requires extra work from users, but results in much more accurate location labels. We felt that this tradeoff was acceptable, as we did not want participants to evaluate our visualizations based on their accuracy. Instead, we want to probe participants' *privacy reactions* to visual differences. Thus, we wanted to reduce, as much as possible, factors like label inaccuracies.

Evaluating Location Visualizations

At the end of the two-week study, we interviewed each participant about their location feeds. For each interview, we randomly ordered the visualization manipulations (for visualization type, as well as marker and label types). For the first visualization shown, we asked participants if there were any surprises or locations missing. Also, because the labels are often not as visually prominent at first glance, we made sure to point out the differences between label types when they first occurred. Otherwise, no other features were explicitly called out to the participant, though the experimenter did address any questions that the participants asked. This allowed our researcher to ensure participants properly understood each visualization and could ask follow-up questions to provide further qualitative feedback.

Recall that at the pre-study interview, we asked participants to provide names of four relationship groups (family, close friends, acquaintances, supervisors). We used these names to frame the sharing questions and, for each relationship, participants choose one of the 18 visualizations that they were most comfortable sharing (**Figure 3**d). After indicating their preferences, participants verbally explained their choices. The interview lasted one hour; feedback was recorded and later transcribed. At the end of the interview, participants were compensated with a $30 gift card.

RESULTS

In total, we identified 139 unique places, with each participant visiting, on average, 11.6 unique places (σ=3.1). In response to the completeness of our visualizations, nearly all of our participants responded that there were no omissions in their visualization. Only one participant was surprised by the inclusion of some places on her visualization, though she quickly recalled the relevant events and realized that the visualization was correct.

We were able to automatically generate geographic labels for all 139 identified places. We generated semantic labels for 71.3% of the places. Following the human-in-the-loop approach, participants corrected 10.2% of those labels. Locations that the participant was not able to label were removed and not included in any of the visualizations.

Sharing Location Feeds

As Table 1 shows, depending on who the visualization is shared with, participants generally preferred either the text- or map-based visualization. Prior work has shown that, when sharing *current* locations, users are significantly influenced by who is asking [7, 15]. Our results are similar, but apply to sharing of *historical* locations.

By choosing different types of location labels, participants implicitly control the granularity of their shared location feeds. Consolvo et al. showed that users are likely to share location at a granularity useful to the requester [7]. In our study, we found that participants did *not* consistently choose the same granularity for each group. For example, when sharing with family members, participants were split between sharing specific geographic (addresses, 66.7%) and general geographic labels (neighborhood or city). For acquaintances, we found that participants were split between general semantic (business types, 66.7%) and general geographic names. Only close friends and bosses elicited consistent granularity preferences of general semantic labels and general geographic labels, respectively.

These results suggest that, when there are potential power dynamics involved (e.g. sharing with family members or supervisors), participants prefer the lowest granularity (general geographic labels). When with more intimate relationships (such as close friends), they are comfortable sharing a more descriptive location (general semantic).

When explaining their preference, participants most commonly cited concerns about their physical privacy, where they wanted to avoid being unexpectedly found.

> *"I like being able to go someplace and know that other people can't find me. If you have [specific geographic labels], then you can't really do that anymore. It makes it super easy for people to bother you whenever they want. So yeah, I'd rather share the [general geographic labels]. If someone really knows you, then they might still know where to find you. But those people are OK…they know you well enough to figure it out, so it probably means they're a good friend so if they find you, it's not a big deal. But that's why you need something general, so you don't have to worry about those other people." –P3*

> *"I can see how [specific geographic labels] might help spontaneous meetings [and that] might be cool. But I think I'd only like it every once and while. I'd rather someone just call me if they want to find me. Giving them [general semantic labels] means they'll still need to ask me to find me. Sure, it's more work for me, but I prefer to know that someone is looking for me, rather than have them just show up and surprise me." –P9*

Previous studies have mentioned the issue of power dynamics [20] and users' fear of unknowingly being tracked [23]. Our results suggest that the same concerns for sharing historical locations but also, more importantly, that there are ways to overcome these privacy concerns by *visually* manipulating how location information is shared.

Visualization Preferences and Reactions

There were two main visualization preferences that we observed. First, participants unanimously disliked sharing the time-based visualization. Based on interview feedback, participants felt that it was too privacy-invasive. In particular, several participants showed concerns that visualizing the time information might make others more curious about a place than they might have otherwise been.

> *"If [others] see that I was somewhere for a long time, like for more than a day, they're going to want to click on [the timeline]. You know, so that they can find out more. I mean, that's what I would do if I were looking at someone else's timeline. I'd be curious why someone would spend so much time at a place." –P2*

Similarly, many participants felt that revealing the temporal information may lead others to draw incorrect inferences about what they may have been doing at a place.

> *"You just don't know what kinds of conclusions people are going to jump to when they see how much time you spend at certain places. I mean, what if I was at home for a whole week? Maybe I was feeling sick. But someone else might think 'gosh he's such a lazy bum' and totally get the wrong picture of me." –P5*

> *"Actually, it's not that I mind people knowing I was [at the store]. It's more that I mind people knowing that I was there for so long. I guess I don't want to have to explain to people what I was doing, if they ask. Isn't it enough that they know where I was at?" –P11*

Our second observation was that, when choosing a map-based visualization, participants' preferences for which marker to use was often not consistent with their choice of location labels. For example, one would expect that when choosing either general geographic (city/neighborhood) or general semantic labels (business type), participants will prefer a visualization with halo markers, since using general names and the halo provides more plausible deniability. For example, combining a specific semantic label (like Starbucks) with a halo marker is nearly as informative as a pushpin marker with an exact address, especially if the person viewing the map is familiar with the area and knows that there is only one Starbucks there.

Yet, the majority of our participants selected this mismatched label+marker combination for sharing their location feeds with their close friends and acquaintances. To understand this preference, recall that halos are, by definition, randomly centered. The participant's true location is only guaranteed to be somewhere inside the halo and not necessarily at the center of the marker. Because of this design, it is possible for two locations to be somewhat far from each other, yet appear close together with the halos (Figure 2). Thus, some participants mentioned that their preference for pushpin markers was because they felt that the pushpins conveyed more mobility than halo markers.

> *"I know [the pushpin markers] are kind of revealing, but I like that the points are more scattered around. It feels like [the halo markers] just make my life seem kind of boring. They often overlap and makes it look like I'm not really moving around much." –P10*

Some participants also selected this combination for sharing their location feeds with family. However, this percentage is much smaller than with close friends and acquaintances.

DISCUSSION

We now discuss the strengths and limitations of our study methodology, the implications of how visualizations can impact end-user privacy preferences sharing location feeds, and the remaining challenges for LSAs to design privacy-aware visualizations for socially sharing location feeds.

Strengths and Limitations of the Study Design

In our study, we use participants' *actual* location traces, so that participants can better reflect on the potential consequences of sharing their location feeds with others. But, despite this realism, there are some confounding factors that our results should be considered against.

First, all the disclosure decisions posed to our participants were hypothetical. Participants did not actually share their visualizations with others. However, to help ground their decision making, we provided real names of people in their social network when framing disclosure questions. Also, based on interview feedback, we felt that participants were indeed thoughtfully answering and carefully reflecting. Second, our sample size was quite small. This is due to the effort and equipment needed to collect GPS data. A larger sample size would be helpful in future studies, in addition to recruiting more users from outside the university.

Sensitivity Towards Temporal Features

Relationship Group	Visualization Type			Label Type				Marker Type		% of Participants	Comfort Rating (median)
	Text	Map	Time	Specific Geographic	General Geographic	Specific Semantic	General Semantic	Pushpin	Halo		
Family Members											
Combination #1		X			X				X	50.0%	5
Combination #2		X		X				X		33.3%	4
Combination #3		X			X			X		16.7%	4
Close Friends											
Combination #1		X					X	X		83.3%	4
Combination #2		X					X		X	16.7%	4
Acquaintances											
Combination #1		X					X	X		66.7%	3
Combination #2		X			X				X	33.3%	3
Supervisors											
Combination #1	X				X			N/A		83.%	4
Combination #2		X			X				X	16.7%	4

Table 1. Visualizations preferences, grouped by the different relationship groups. Comfort ratings indicate how comfortable participants felt sharing a particular visualization combination with a specific relationship type.

When examining location sharing, past work has almost exclusively focused on sharing of *current* locations [1, 7, 13, 14, 18] and has only recently begun to consider *social* location sharing [22]. In our study, we provide initial insights into how users make privacy decisions about sharing *visualizations* of their *historical* locations. Based on participants' responses, we found that certain dimensions of location feeds are much more privacy sensitive than others.

In our visualizations, we choose to emphasize 7 location variables: spatiality, name (label), arrival, departure, duration, frequency, and sequential order. Some of these emphasize the physical properties of a place (spatiality and label), while the rest are temporal features. Generally speaking, participants felt that temporal properties were more privacy sensitive. In fact, of the five temporal features, participants reported that they were least comfortable sharing duration (median=1) and arrival (median=2). These self-reported comfort scores are based on a 5-point Likert scale, where 1=not at all comfortable.

Participants' discomfort is because they felt that some temporal features were often tightly associated with what they were *doing* at a particular place. This suggests that users are more concerned about sharing their *activities* rather than their location. This finding is interesting as prior work suggests that users are willing to share their activity information [22]. We reason that our results are different because, in Tang et al's study, participants were allowed to manually specify a place's label. In our study, labels are automatically generated and are always tied to the physical properties of a place. As such, activity information is never explicitly conveyed through the visualization; it is only inferred based on the label, spatiality, and time information.

Technically, as the visualizations are isomorphic, these inferences should be possible with any of the three visualization types. However, participants had the strongest (negative) reaction to the time-based visualization because it exposes the very information that is most suggestive of one's activity at a particular place. Thus, our results suggest that, when sharing location feeds, users should be given an option to *not* share certain temporal information. Otherwise, LSAs should allow users to additionally tag locations with their own activity information, so that there will fewer chances that others will improperly judge the user.

Perceived Perceptions of Control

In the interviews, participants hinted at issues of control and how some visualizations took control away from them.

> *"I can't really control what the timeline [in the time-based visualization] will look like to other people, so I just don't feel comfortable sharing it." –P7*

> *"I chose this visualization because I feel like it's the most flexible. Like, it shows just enough that people can know what I'm up to, but it's vague enough that I can spin another story, if I want to." –P5*

While P5's comment is somewhat related to the issue of plausible deniability, it is also indicative of the importance for LSAs to support *flexible* visualizations that empower users to feel in control of *how* their locations are perceived by others. In our study, we automatically generated labels for users' past locations. This automation led users to feel less control over *what* location information might be shown to others. Thus, to compensate for this loss, users chose visualizations that allowed them to share information only in broad strokes, though they still wanted it to be helpful.

> *"I don't think I'd share [the text visualization] with my close friends. It's just not that useful in that format, but I also don't want to share everything on the map either. But, by using the general labels, I feel like the map won't be too revealing." –P10*

P10's comment suggests that having the right spectrum of location labels is an important feature for LSAs to consider when supporting automated sharing of location feeds.

But labels are not the only way that participants try to control their location disclosures. All the participants agreed that the text-based representation was the least revealing of the three visualization types. This is an interesting response, as the visualizations were designed to be as close to isomorphic as possible. However, most participants believed that presenting location in text form was less susceptible to further probing by others.

> *"The text[-based visualization] just seems more innocent. You can't immediately tell if a place is sketchy unless you look really hard and google it or something. And people are usually way too busy so they aren't going to bother with all that work." –P4*

P4's belief is that, by removing sensitive location information to be one-click away, he is afforded more privacy. This perceived privacy control is important to note, as it reinforces the link between *information control* and *information access*, both important concepts that LSAs should heed in supporting privacy-sensitive sharing. Specifically, when users feel like it is harder to access information, then they are more likely to feel comfortable sharing a particular visualization. However, this also suggests that LSA should remind users that access is not always as elusive as they may believe.

Social Dimension of Perceived Privacy Concerns

Recent studies in social location sharing have suggested that impression management factors into users' decisions about how they name their locations [22]. We have similar findings in our study and have evidence to suggest that users give significant consideration how a visualization may be interpreted by others. In particular, several participants opted for pushpin markers over the halo markers for this very reason, commenting that the pushpins gave them the sense that they were more mobile and, consequently, would appear more interesting to others.

This type of feedback strongly suggests that there are indeed ways for visualizations to manipulate end-user perceptions about *what* and *how* information is shared. The consequence of choosing pushpin markers is that, at least visually speaking, readers are given a much more precise indication of where a user was located. Thus, LSAs again must carefully consider how they design visualizations so that they can responsibly inform users of both the social benefits, as well as the potential privacy issues.

Challenges: Scalability and Value of Automated Sharing

Our study examines the privacy concerns relating to *automated* sharing of *historical* locations. One may question whether this type of location sharing is worth studying, given current trends in LSAs. Foursquare, a LSA with one of the largest user base, has over 6 million users and has reported over 1 million check-ins per day [10]. While these numbers do suggest Foursquare is growing, it also shows that the users are only contributing, on average, about one place every week. LSAs are often seen as a service that mainstream has not fully embraced yet [12].

One potential reason for this is that there may not be enough location data being recorded and shared, so services are seen as less useful. To address this, LSAs could easily begin incorporating automated location updates, and some LSAs already in fact support this sharing mode. Thus, it is critical for us to be ahead of the technology curve and begin understanding how we can appropriately design LSAs to share these location feeds. Designing *privacy-sensitive visualizations* and understanding their effect on users' perceived privacy concerns has not yet been discussed for LSAs, though we have seen previous visualization studies done in other domains (e.g., [8]). Based on our observations, we have found that there are indeed important domain-specific issues to consider when considering how users are influenced by visual representation of locations.

Moving forward, there are still many challenges to address. For example, our study only considers a small section of the design space for location visualizations. There are many other variations that should be further explored. In addition, our visualizations were all designed to support only a single user and only two week's worth of data. For future work, these visualizations should address scalability issues so that they can support multiple users and varying time windows.

CONCLUSION

In summary, our study has focused on users' perceptions of different visualizations for sharing location feeds. While we have not exhaustively explored the design space of location visualizations, we believe that this study is a step forward in understanding how visual representations of location information can affect a user's perception of privacy. Based on our findings, we describe several design implications for LSAs that wish to support sharing of past locations.

REFERENCES

1. Anthony, D., Kotz, D. and Henderson, T. (2007). Privacy in Location-Aware Computing Environments. IEEE Pervasive Computing, 6 (4), 64-72.
2. Benisch, M., Kelley, P.G., Sadeh, N., et al. (2008). The Impact of Expressiveness on the Effectiveness of Privacy Mechanisms for Location Sharing. Technical Report, CMU-ISR-08-141R, Carnegie Mellon University, Institute of Software Research.
3. Brush, A.J., Krumm, J. and Scott, J. (2010). Exploring End User Preferences for Location Obfuscation, Location-Based Services, and the Value of Location. In Ubicomp '10, 95-104.
4. Card, S.K., Mackinlay, J. and Shneiderman, B. Readings in Information Visualization: Using Vision to Think Morgan Kaufmann, San Francisco, 1999.
5. Cheng, Z., Caverlee, J., Lee, K., et al. Exploring Millions of Footprints in Location Sharing Services. In ICWSM '11.
6. Cleveland, W.S. The Elements of Graphing Data. Wadsworth, Monterey, CA, USA, 1985.
7. Consolvo, S., Smith, I., Matthews, T., et al. (2005). Location Disclosure to Social Relations: Why, When, & What People Want to Share. In CHI '05, 82-90.
8. Cranor, L.F., Guduru, P. and Arjula, M. (2006). User Interfaces for Privacy Agents. ACM Trans. Comput.-Hum. Interact., 13 (2), 135-178.
9. Geonames. http://www.geonames.org/export/ws-overview.html
10. Grove, J. Foursquare Nearing 1 Million Checkins Per Day. http://mashable.com/2010/05/28/foursquare-checksin/
11. Hoadley, M.C., Xu, H., Lee, J., et al. (2010). Privacy as Information Access and Illusory Control: The Case of the Facebook News Feed Privacy Outcry. Electronic Commerce Research and Applications, 9 (1), 50-60.
12. Hong, J.I. Privacy & Security in the Location-Enhanced World Wide Web. In Workshop on Privacy at Ubicomp '03.
13. Hsieh, G., Tang, K.P., Low, W.Y., et al. Field Deployment of Imbuddy: A Study of Privacy Control and Feedback Mechanisms for Contextual Im. In Ubicomp '07, 91-108.
14. Iachello, G., Smith, I., Consolvo, S., et al. Control, Deception, & Communication: Evaluating the Deployment of a Location-Enhanced Messaging Service. In Ubicomp '05, 213-231.
15. Lederer, S., Mankoff, J. and Dey, A.K. (2003). Who Wants to Know What When? Privacy Preference Determinants in Ubiquitous Computing. In CHI '03, 724-725.
16. Lin, J., Xiang, G., Hong, J.I., et al. Modeling People's Place Naming Preferences in Location Sharing. Ubicomp '10, 75-84.
17. Palen, L. and Dourish, P. (2003). Unpacking "Privacy" for a Networked World. In CHI '03, 129-136.
18. Sadeh, N., Hong, J.I., Cranor, L., et al. (2008). Understanding and Capturing People's Privacy Policies in a Mobile Social Networking Application. Pers & Ubiq Comp, 13 (6), 401-412.
19. Schilit, B.N., LaMarca, A., Borriello, G., et al. (2003). Challenge: Ubiquitous Location-Aware Computing and the "Place Lab" Initiative. In WMASH '03, 29-35.
20. Shklovski, I., Vertesi, J., Troshynski, E., et al. (2009). The Commodification of Location: Dynamics of Power in Location-Based Systems. In Ubicomp '09, 11-20.
21. Skyhook Wireless. http://skyhookwireless.com/locationapps/
22. Tang, K.P., Lin, J., Hong, J.I., et al. (2010). Rethinking Location Sharing: Exploring the Implications of Social-Driven Vs. Purpose-Driven Location Sharing. In Ubicomp '10, 85-94.
23. Tsai, J., Kelley, P., Cranor, L., et al. (2009). Location-Sharing Technologies: Privacy Risks and Controls. In TPRC '09.
24. Tufte, E.R. The Viual Display of Quantitative Information. Graphics Press, 1983.
25. Wang, Y., Lin, J., Annavaram, M., et al. (2009). A Framework of Energy Efficient Mobile Sensing for Automatic User State Recognition. In MobiSys '09, 179-192.
26. Ware, C. Information Visualization: Perception for Design. Morgan Kaufman, San Francisco, 1999.
27. Westin, A.F. Privacy and Freedom. Atheneum, NY, 1967.