Who, when, where: Obfuscation preferences in location-sharing applications

Jayant Venkatanathan, Jialiu Lin, Michael Benisch, Denzil Ferreira, Evangelos Karapanos, Vassilis Kostakos, Norman Sadeh and Eran Toch

June 30, 2011
CMU-CyLab-11-013

CyLab
Carnegie Mellon University
Pittsburgh, PA 15213

This work has been supported by the CMU/Portugal Information and Communication Technologies Institute and by NSF grants CNS-0627513, CNS-0905562, CNS-1012763. This research was also supported by CyLab at Carnegie Mellon under grants DAAD19-02-1-389, W911NF-09-1-0273 from the Army Research Office and by Google.

Key Words: Location sharing, obfuscation, security, privacy, HCI

Abstract

This paper presents a study of obfuscation practices in location-sharing systems. The study shows that users have relatively complex preferences that depend on the recipient of the location, the time of the request and location. The preferences also require multiple levels of obfuscation (ranging from disclosing no location information to disclosing the exact location) to accurately capture. For example, we find that users tend to reveal finer-grained locations to recipients with closer ties, such as family members, but coarser-grained locations to colleagues and strangers. We also find that users utilize the full range of obfuscation options, from high-level region to exact address, which further demonstrates the complexity of their preferences and highlights the importance of obfuscation as a privacy control. Additionally, we find that day of week and type of location affect users' decisions on how much detail to share.

INTRODUCTION

Location sharing applications are gaining wide adoption, with a number of commercial systems now available on the market, including Foursquare and Facebook Places. Such services are frequently used in the context of online social networks, whereby one's real-time location becomes yet another sharable aspect of one's online profile. With the increasing adoption of online location sharing services, understanding users' preferences and needs in terms of location sharing has become crucial.

In this paper, we refer to obfuscation as a mechanism by which the level of detail revealed about a location can be manipulated, for instance by revealing only the area name or city where the user currently resides. While this approach has been used in the past [7], it is not clear what strategies users adopt when setting the obfuscation level of their current location and the extent to which obfuscation can be useful as a privacy enhancing tool for location sharing. Given the significant privacy risks that can result from over sharing, it could be important to provide for users to be able to share their location only to the detail desired. While the lack of control over the extent to which users share their location might cause them to over share, the need to prevent such over sharing might be hampering the adoption of location sharing services. The study presented in this paper seeks to assess the usefulness of obfuscation in realistic scenarios and draw insights on the obfuscation preferences of users.

RELATED WORK

There is an increasing amount of work on understanding users' location-privacy needs in ubiquitous and location-aware systems relying on techniques such as diary studies [2], interviews [9], surveys [12], scenarios [16] and lab and field observations [3,14]. Previous work suggests that the recipient is an important factor determining a user's location sharing preference [13]. Various other factors such as location [3,5], activity or mood [5] and time [3] are also known to affect location sharing preferences.

There has been prior work on using obfuscation as a privacy preserving technique in the context of location based services [1,6,7,4]. It has also been shown that more expressive privacy mechanisms, for example those that depend on the recipient, time of day, day of week, and location, are needed to accurately capture users' privacy preferences [3]. As a result, it can be expected that users will be able to better express these preferences with obfuscation. However, an important issue here is to understand whether the benefits offered by obfuscation are significant, considering a potential trade-off against decreases in usability that can be incurred by allowing the user this control.

The study investigates the effect of recipient, time, and location on obfuscation preferences of participants. In a way, prior studies have investigated certain aspects of these issues [13,5]. To our knowledge, however, the effect of time on obfuscation choice has not been studied, and the interplay between these factors merit further investigation.

STUDY

A study was designed to test the following hypotheses:

H1: Choice of obfuscation level in location sharing varies with the type of recipient.

H2: Choice of obfuscation level in location sharing varies with time of day, day of week and location.

The study captured obfuscation preferences of participants for different locations they visited during the course of the day. Participants were recruited by announcements on email lists, online forums, and fliers distributed across the campus. No reward was offered to participants. The study was conducted between May and September 2010 with participants from the cities of Funchal (Portugal), Lisbon (Portugal) and Oulu (Finland). A total of 25 participants were recruited (22 male), with an average age of 26 (sd 3.84), and they were all students or staff from universities in these cities. The duration of participation for each participant ranged between 4 and 7 days.

Method

It is methodologically challenging to capture obfuscation preferences for location sharing in a realistic and reliable manner. While estimates of participants' preferences for a given location, such as one's workplace, tend to underestimate their variability [15], probing the participant to report on ongoing experiences, a technique known as the Experience Sampling Method [8], is labor-intensive and may miss important information when the participants are not able to respond. The study described in this paper uses the Day Reconstruction Method (DRM) [11], an alternative approach that asks participants to recall experiences that took place in the previous day in forward chronological order. DRM has been shown to provide a surprisingly good approximation to Experience Sampling data [10], and has been successfully used to study location-sharing preferences in a prior study [3].

System

Each participant was given an Android smart-phone equipped with GPS logging software. Participants were instructed to use this phone as their primary phone to ensure that they kept it with them at all times. Each participant used the phone for a period of between four and seven days. During this period the phone recorded participants' locations whenever they changed by more than 10 meters. Participants carrying the phones were given the option to temporarily disable the logging software should they wish to do so. Participants were instructed to upload their location data at the end of each day, and immediately answer a questionnaire online.

Experimental task

The main experimental task was an online questionnaire that participants answered at the end of each day. To do so, participants first had to register with a custom online application. During registration, participants were asked to list the names of five people from each of their family, close friends and colleagues. Subsequently, when participants logged in, the system processed the recently uploaded location data and generated an on-the-fly questionnaire in two steps.

Figure 1: Screenshot of a sample question asked for each important location and recipient group.

First, the set of "important locations" was selected from the uploaded data in the following manner: Each of the GPS readings from the uploaded data was aggregated into a location observation, if the user stood still, or a path observation, if the user moved, in a similar manner to as described in [3]. A new observation was created when a participant moved more than 100 meters from his last known location and remained stationary again for at least 2 minutes. The location observations thus obtained were the set of important locations that were subsequently used and the path observations were discarded. This ensured that transitional locations were removed from the set of important locations.

Subsequently, participants were taken through a series of questionnaire pages, where each page displayed an important location on the map along with the details of when the participant was there. For each important location, participants were asked to define how much information they would have liked to share about that specific location, at that specific time, with one specific person (chosen at random) from each of the four recipient groups (Figure 1). The information was entered in a scale that ranged from revealing (1) no information to (2) region, (3) city, (4) neighborhood and (5) exact address details. No real location sharing took place during the study.

RESULTS

During the study each participant visited on average 2.67 (s.d. = 0.83) important locations each day, and for each visit to a location the following data was recorded: time and duration of visit, privacy preference (on a scale 1-5 as shown in Fig. 1) for each of 4 possible recipient groups (family, close friends, colleagues, strangers).

To examine the effect of location on participants' obfuscation choices, each location that a participant visited was manually labeled "home", "work" or "other based on the following heuristics: (1) Since all participants were students and staff of one of three universities, their university locations were labeled "work", (2) The location that they spent the most amount of time between 9 pm and 9 am during the period of the study was labeled as their "home", and (3) all other locations were labeled "other". Based on these heuristics, the dataset contained no 'home' location for 3 participants, no 'work' location for 4 participants and no 'other' location for 1 participant.

We measured the extent to which a participant's preferences for each target group of individuals could be captured by one single obfuscation level - the level that the participant chose most for that group. We refer to this obfuscation level as the *Single Largest Choice* of obfuscation (*SLC*) of the user for that recipient group and refer to the fraction of time the participant chose the SLC as the *value* of the SLC for that recipient group. For example, if a participant chose obfuscation level 5 more often than any other obfuscation level for her family members, and she did this for a 0.4 fraction of her total location visits during the study, her SLC for "Family" is 5 (i.e. exact location) and the value of her SLC for "Family" is 0.4. Table 1 summarizes the mean SLC's over all users for each target group.

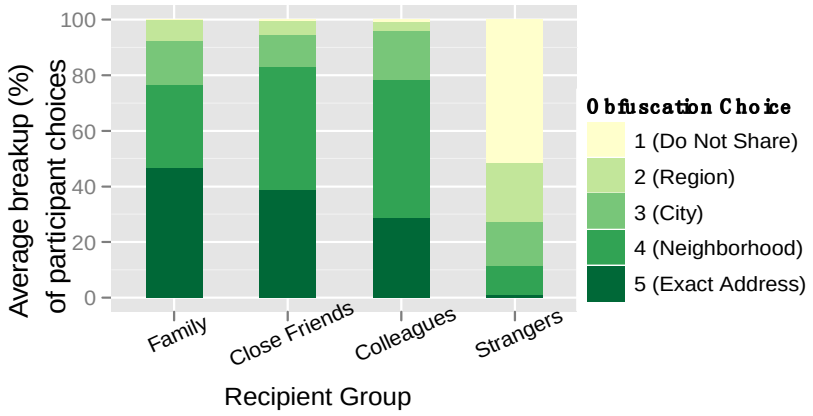

Figure 2 : Obfuscation per recipient group averaged over all participants.

	Family		Close Friends		Colleagues		Strangers	
	mean	***sd***	***mean***	***sd***	***mean***	***sd***	***mean***	***sd***
Overall	**0.75**	**0.16**	**0.69**	**0.19**	**0.67**	**0.17**	**0.86**	**0.18**
Home	**0.87**	**0.16**	**0.81**	**0.20**	**0.83**	**0.17**	**0.86**	**0.20**
Work	**0.89**	**0.15**	**0.82**	**0.22**	**0.79**	**0.22**	**0.96**	**0.09**
Overall	**0.78**	**0.20**	**0.82**	**0.19**	**0.72**	**0.19**	**0.87**	**0.20**

Table 1. Value of the single largest choice of obfuscation for the four recipient groups

Effect of Recipient Group on Obfuscation Level choice

The details shared about locations decreased as the "strength of ties" between the participants and the potential recipients decreased (Figure 2). Perhaps unsurprisingly, participants shared most details with family and least details with strangers. Overall, participants used extreme obfuscation values in their preferences: they used mostly values 1 and 2 for strangers, and mostly 4 or 5 for everyone else (Figure 2).

The overall high values of the SLCs (the "mean" columns in table 1) reveal that participants seemed to have a default obfuscation choice for each recipient group which they chose with a

very high frequency for that group. For example, participants chose a default obfuscation level for "Family", the SLC for the "Family" group, during an average of 75% of their location visits. This high values of SLC's is likely to reflect participants' default attitudes towards each of the recipient groups. That participants tended to choose a particular obfuscation level for each recipient group with a very high frequency reaffirms hypothesis H1.

Effect of Time and Location on Obfuscation Level choice

Participants tended to be more active in sharing during office hours (between 9 am and 7 pm), even though their actual obfuscation choices did not significantly vary by hour of day (Figure 4). Furthermore, these patterns were also observed within each recipient group. In addition, most data points were recorded during mid-week (Figure 3) and after mid-day (Figure 4). This shows that during mid-week participants are more mobile and more likely to register a "significant

location" with the system. At the same time, they appeared to be more open and willing to share details about their location at those times. A chi square test showed that the relationship between day of week and obfuscation choice was significant ($\chi 2(24,2376)=82.449$, $p<0.0001$).

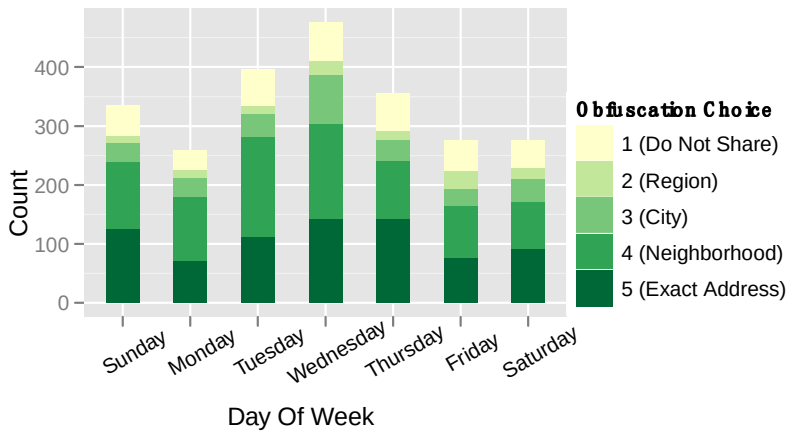

Figure 3 : Obfuscation per day of week, and total data points per day of week.

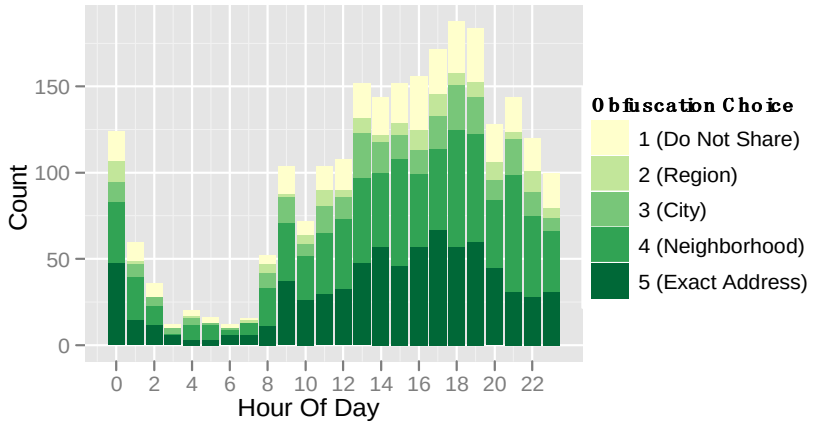

Figure 4 : Obfuscation per hour of day, and total data points per hour of day.

	Overall	Home	Work	Other
Mean	**0.39**	**0.54**	**0.61**	**0.52**
SD	**0.18**	**0.25**	**0.27**	**0.22**

Table 2. Value of the single largest choice for preference-vectors

To understand the effect of location on obfuscation choice, we examined participants' SLC values for each recipient group restricted within home and within work locations. Paired-sample t tests (comparing each participant's overall SLC value against their SLC value restricted to home or work) showed that there was a statistically significant ($p<0.05$) increase in the value of the SLCs for all recipient groups when the data was restricted only to home or work locations, with the exception of the 'Strangers' group restricted to home locations. In terms of magnitude, the highest average increase was observed when participants' SLC's for the colleagues group were restricted to "home" (from 0.67 to 0.83, see Table 1). This suggests that in addition to recipient group, location is an important predictor for obfuscation choice and the obfuscation choice of a participant is fairly predictable given recipient group and location, underlying the effect that recipient group and location have on participants' obfuscation choice.

Usage of Obfuscation

Finally, we represented the set of obfuscation choices for each visit of a participant over the four recipient groups as a tuple of values. For example the tuple (5, 3, 4, 1) signifies that the participant chose 5 for family, 3 for close friends, 4 for colleagues and 1 for strangers. We refer to such a tuple as a *preference vector*. The diverseness of preference vectors of a participant reflects the diverseness of his or her overall obfuscation choices. Participants used 7.12 distinct preference vectors on an average (min=2, max=14, sd=2.85). We also calculated the value of the SLC of preference vectors - the preference vector that a participant used most across all visits, and this is summarized in Table 2. The SLC of the preference vector for any participant was chosen during less than half (0.39) of the visits on an average and just above half when considering only home (0.54) or work (0.61) visits. This reveals that participants overall preferences did in fact vary across visits.

DISCUSSION

Overall participants tended to share more with closer ties and shared less as the strength of ties of potential recipients decreased. To a certain extent this behavior was expected and is likely to reflect participants' default attitudes towards the various recipient groups. Given that the obfuscation choice varied significantly by the day of week, this suggests that orthogonal to people's attitudes towards recipients based on their identity, people's choice of obfuscation is also affected by time. One explanation for this finding is that different activities are associated with different parts of the week, prompting these variations in obfuscation. It is also interesting to observe that the type of location as captured by our rather simplistic classification into "work", "home" and "other" did, in fact, play a significant role in determining participants' preferences.

It is important here to note that the variables time and location are not strictly independent of each other, considering that individuals tend to have daily and weekly routines. The locations that people visit for the most part can be expected to be strongly associated with time. For example, a university student is most likely to spend her weekends at home or some place that is not the university, while she is likely to visit her university on a weekday. It is thus not surprising that given one of location or time had an effect on obfuscation choice, the other variable did too.

Finally, the results show that participants used at least four different levels of obfuscation for each recipient group. This suggests that, while on average participants share more details with close family, and more during mid-day and mid-week, in fact the full breadth of obfuscation

levels was utilized by participants to express their preferences. In addition, the results show that all obfuscation levels were used throughout the day and week, suggesting that participants' preferences maintain a constant level of complexity.

IMPLICATIONS

The study shows that participants' choice of location-sharing obfuscation depends on the identity of the recipient and day of week and location. Crucially, however, the study shows that participants' use of obfuscation is quite complex and requires a full range of values to express. Therefore, simplistic on/off or high/low settings for obfuscation are not sufficient to capture the richness in people's location-sharing preferences.

The results also show that people's mobility patterns vary by hour of day and day of week. This shows that simplistic rule-based obfuscation settings whereby only time or identity can be set are insufficient to capture the richness of users' preferences. Instead, the design of a rule-based obfuscation privacy control should allow users to express rich obfuscation preferences in terms of both time, identity of recipient and location. These insights can inspire further studies and inform the design of future location sharing and location based applications.

LIMITATIONS

The participants recruited for the study were all university staff and students, mostly male, and most of them were young. Hence it is possible that the results observed here might not be directly generalizable to broader demographics. For example, a door-to-door computer technician working for a tech-support firm might share his location in greater detail with his colleagues, in order to facilitate co-ordination, than a university staff working in a university campus. It is not obvious that we would observe precisely the same results for a more diverse sample of participants. In addition, no real location sharing took place among participants. It would be interesting to observe whether the results observed would vary if participants actually shared their location real-time with their contacts, as is the case in location sharing applications.

CONCLUSION

The results of the study described in the paper show that the obfuscation sharing preferences of users depend on the recipient of the information, and the location and time of request. Participants tended to reveal finer-grained locations to recipients with closer ties, such as family members, but coarser-grained locations to colleagues and strangers. Their preferences also varied with the day of week and the type of location, such as home or work. Finally, participants varied their obfuscation preferences across location visits, showing that they had diverse overall preferences.

ACKNOWLEDGEMENT

This work has been supported by the CMU/Portugal Information and Communication Technologies Institute and by NSF grants CNS-0627513, CNS-0905562, CNS-1012763. This research was also supported by CyLab at Carnegie Mellon under grants DAAD19-02-1-389, W911NF-09-1-0273 from the Army Research Office and by Google.

REFERENCES

1. Ardagna, C., Cremonini, M., Damiani, E., De Capitani di Vimercati, S., Samarati, S. (2010). Location privacy protection through obfuscation-based techniques. Proc. Annual IFIP WG 11.3 Working Conference on Data and Applications Security.
2. Barkhuus, L. (2004). Privacy in location-based services, concern vs. coolness. MobileHCI 2004 workshop: Location System Privacy and Control.
3. Benisch, M., Kelley, P.G., Sadeh, N., Cranor, L.F. (2010). Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs. Personal and Ubiquitous Computing, Forthcoming.
4. A. B. Brush, J. Krumm, and J. Scott. (2004) . Exploring end user preferences for location obfuscation, location-based services, and the value of location. In Proc. of Ubicomp 2010.
5. Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J., Powledge, P. (2005). Location disclosure to social relations: why, when & what people want to share. Proc CHI 2005. ACM Press (2005)
6. Duckham M., and Kulik, L. (2005). A formal model of obfuscation and negotiation for location privacy. Pervasive 2005, pp. 152-170.
7. Gandon, F. and Sadeh, N. (2004). Semantic Web Technologies to Reconcile Privacy and Context Awareness. Journal of Web Semantics. Vol. 1, No. 3, 2004.
8. Hektner, J. M., Schmidt, J. A.,& Csikszentmihalyi, M. (2007). Experience Sampling Method: Measuring the Quality of Everyday Life. Sage Publications Inc.
9. Hong, J. I. and Landay, J. A. (2004) An architecture for privacy-sensitive ubiquitous computing. MobiSys 2004, 177-189.
10. Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., and Stone, A. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. Science, 306(5702):1776-1780.
11. Kelley, P. G., Hankes Drielsma, P., Sadeh, N., and Cranor, L. F. (2008). User-controllable learning of security and privacy policies. AISec 2008, 11-18.
12. Khalil, A. and Connelly, K. (2006). Context-aware telephony: privacy preferences and sharing patterns. CSCW 2006, 469-478.
13. Lederer,S., Mankoff,J., Dey, A.K. (2003) Who wants to know what when? Privacy preference determinants in ubiquitous computing. Proc CHI 2003. ACM (2003)
14. Sadeh, N., Hong, J., Cranor,L., Fette,I., Kelley,P., Prabaker,M., and Rao,J. (2009). Understanding and Capturing People's Privacy Policies in a Mobile Social Networking Application", Journal of Personal and Ubiquitous Computing, Vol. 13, No. 6.
15. Schwarz, N., Kahneman, D., Xu, J., Belli, R., Stafford, F., Alwin, D., et al. (2008). Global and episodic reports of hedonic experience. Calendar and Time Diary Methods in Life Course Research: Methods in Life Course Research, Sage Pubns, 157.
16. Wagner, D., Lopez, M., Doria, A., Pavlyshak, I., Kostakos, V., Oakley, I., Spiliotopoulos, T. (2010). Hide And Seek: Location Sharing Practices With Social Media. MobileHCI 2010, 55-58.