Seven Privacy Worries in Ubiquitous Social Computing

Sara Motahari, Constantine Manikopoulos, Roxanne Hiltz, Quentin Jones
New Jersey Institute of Technology
University Heights, Newark, NJ, 07102, USA
sg262@njit.edu, manikopoulos@ADM.njit.edu, roxanne.hiltz@njit.edu,qgjones@acm.org

ABSTRACT

Review of the literature suggests seven fundamental privacy challenges in the domain of ubiquitous social computing. To date, most research in this area has focused on the features associated with the revelation of personal location data. However, a more holistic view of privacy concerns that acknowledges these seven risks is required if we are to deploy privacy respecting next generation social computing applications. We highlight the threat associated with user inferences made possible by knowledge of the context and use of social ties. We also describe work in progress to both understand user perceptions and build a privacy sensitive urban enclave social computing system.

1. INTRODUCTION

To protect users' privacy, system designers would ideally consider two aspects: users' perceptions and real risks. However, this requires an understanding we do not yet have of the relationship between technology and various types of privacy threats. Researchers have made valuable efforts to address many of the more obvious threats associated with user access control, particularly in the domains of ubiquitous and context-aware computing [3, 6], but the problem has not been completely solved.

Users usually don't have a complete understanding of the threats to their privacy. They recognize when their systems result in "privacy issues", but they usually don't know exactly what those issues are [12]. Numerous user studies have been conducted to explore people's feeling about their privacy. Many of these studies such as Westin [5] identify users' characteristics regarding their privacy [1, 2] or their attitude towards a particular fragment of information in a specific application [1, 2]. While these studies were useful and necessary, there has been no thorough measurement of users' perceptions of different types of privacy violation and the way user concerns relate to the system type and technology. Considering the E-security protection market and controversies over privacy, we think many users are worried about security aspects and the "big brother" issue. However, other aspects of privacy invasions have not been explored to the same extent, which may in fact pose an equal or greater threat to users.

In particular, users can infer unauthorized results from authorized pieces of information. This is known as the inference problem or as Mifflin defines: "The process of arriving at some conclusion that, though it is not logically derivable from the assumed premises, possesses some degree of probability relative to the premises." [10]. The inference problem is discussed in many studies too. It is mostly known as a security threat to databases [4] and sometimes a privacy risk in data mining [11]. However the problem gets more complicated in ubiquitous social computing (USC) systems that combine on-line social interactions with context-aware computing. In this domain, the sensitivity of user information has a dynamic nature based on the context.

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee. Symposium On Usable Privacy and Security (SOUPS) 2007, July 18-20, 2007, Pittsburgh, PA, USA.

Furthermore, information such as life patterns and the quality of social relations that are not kept in the database, can be inferred from available information. This leads to inference threats to the privacy of the users, which have not been explored in the studies about the inference problem. Social relations can be another source of information revelation in such systems.

2. PRIVACY AND USC

We mentioned that traditional approaches are not enough to address privacy threats in ubiquitous social computing. To explain why a more holistic view is needed, we give a real example of a privacy invasion from a centralized location-aware social computing test-bed, SmartCampus[13]. The event happened while a student thought he was safe because he was provided with the option to stay anonymous and he hid his name, but his identity was inferred from his location information. The student was using CampusWiki[13]; which allows students to create and edit location linked content. Editors of pages can be hidden or identified, as can their location. Using the CampusWiki application, the student added unpleasant comments about a course professor and revealed his location, but didn't reveal his identity. However, the professor after looking into the history of page edits related to his Wiki profile realized that the comments were added in his classroom when he was teaching. Since only two students were using a laptop during the class in question, he correctly inferred the student's identity. The result was a confrontation, which lead to the student dropping the course. This simple example shows how privacy within this context can only be preserved by taking into account social inference risks.

We grouped the threats to the user privacy in ubiquitous social computing system into seven categories.

1. Inappropriate use by Administrator's: E.g. The system admin sells personal data without permission [9].
2. Legal Obligations: The system admin is forced by an organization such as the police to reveal personal data [9].
3. Inadequate Security [9].
4. Designed Invasion (Poor Features): E.g. a cell phone application that reveals location to friends, but does this without informing the user or providing control of this feature [3, 6].
5. Social Inference through lack of Entropy: See CampusWiki example above.
6. Social Inference through Persistent User Observation: E.g. Bob is so often in Alice's office. Their relationship must be romance.
7. Social Leveraging of Privileged Data: E.g. David can't access my location, but Jane can. David asks Jane my location.

2.2 Work in Progress

Our research is focused on how to holistically address both user concerns and actual risks to their privacy associated with a real world deployment of a USC system.

2.2.1. System Instantiation

Our privacy policy and privacy management system have to address the users' concerns, but at the same time they have to deal

with the real risks, which may be unknown to the users. The CampusWiki example and our preliminary survey results mentioned below show that some serious risks remain unknown to users. In a centralized USC system like SmartCampus, we have to take care of the first and the second categories by explaining our use, data retention, distribution and sharing of policies to users. Users have to be assured that their information will not be misused by the central management. To address the security concern, we need a powerful security protection system, but we are also designing a privacy protection system to deal with the 5 other crucial categories. It includes a user interface to an access control system, where users can set their privacy preferences. These preferences are dynamic and change based on the contextual factors such as time and location.

The last three categories are the most difficult ones to control. They have not been solved in USC systems, where context-awareness and social relations can lead to novel user inferences. Furthermore, "we have no way of controlling what data is learned outside of the database, and our abilities to predict it will be limited" [7]. Particularly as we have no control over what people exchange in their social relations, but we can help them to prevent unwanted social revelations. Therefore, in addition to the direct access control module, the middleware will include inference control and social revelation control modules. In these modules we check for the entropy of the information based on the situation, the history of the queries, and the proximity of the users. Privacy settings can be adjusted automatically or by the user. For example, the users can get a warning before rejecting or accepting a query or to think twice about their preferences when there is a big inference chance.

2.2.2. Users' Perception of Privacy in USC

User privacy perceptions relate to system features and actual threats. Semi-structured interviews of perspective SmartCampus users showed that there is a great fear of one's location being made available to inappropriate people, enabling stalking and "big brother" concerns. To explore how user concerns relate to real world privacy challenges, we are conducting a user survey. We are trying to understand how aware and how concerned users are about the above 7 privacy categories and how their concerns depend on the type of information, such as identity, location, status, profile information, social relations, etc. Furthermore, we are trying to evaluate our current privacy policy based on their feedback. Our survey is not completed yet and we hope at the end we can answer the above questions. Our subjects are student volunteers from seventeen different majors. They fill out a questionnaire and answer a set of nine questions about their privacy concerns. In the first question of our survey, we introduced the above categorization to the subjects as different "situations" and gave them the same examples as above. We asked them the following question for a general commercial location-based cell phone:

A. Specify how aware you were of each situation before reading these scenarios. (Rate on a scale of one to five).

B. Rate how much it concerns you now that you've read these scenarios.

In the second question, we ask them to rate their concerns over the same threats in our application and in the last part we present our privacy policy and ask the same question again. At the time of writing this paper, the results of the first question asked, for 107 subjects, show that as we guessed, users are less aware of the last three categories and more aware of the first four categories (Friedman, χ^2=299, df=6, n=102, p<0.001). Combining the first three categories into one new variable and the last three categories into another one and performing a paired test shows a real difference between them (Wilcoxon Signed Rank u=-7.91, n=102, p<0.001). Performing the same tests on part B shows that users are generally more concerned over the threats they have more awareness about. They are most concerned over being hacked. The first and fourth categories come next and the other categories are least worrisome for them (Friedman, χ^2=64.4, df=2, n=99, p<0.001). Interestingly, we also found that subjects were less concerned about inappropriate use of their data by our campus based system administrators, than that of an administrator of commercial cell phone services such as Verzion or AT&T (Wilcoxon, u=-2.31, n=103, p=0.021). This is despite the fact that our campus administrators will have access to more personal data and less commercial pressure to ensure user satisfaction.

We have outlined how USC systems raise seven basic categories of privacy concerns, including acute social inferences. We have also highlighted the mismatch between real privacy risks and user perceptions, and some of its design implications in terms of system policies, features and privacy guidelines.

ACKNOWLEDGMENTS

This article is based on work supported by the US National Science Foundation under grants number IIS 0534520 and CNS 0454081. The opinions expressed in this material are those of the authors and do not necessarily reflect those of the NSF.

REFERENCES

[1]. Ackerman, M.S., Cranor, L.F. and Reagle, J. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences *ACM Conference on Electronic Commerce*, 1999.

[2]. Consolvo, S., Smith, I, Matthews, T., LaMarca, A., Tabert, J. and Powledge, P. Location Disclosure to Social Relations: Why, When, & What People Want to Share, *CHI 2005*.

[3]. Cornwell, J., et al, User-controllable security and privacy for pervasive computing. in *WMCSA 2007*

[4]. Farkas, C. and Jajodia, S. The inference problem: a survey. *SIGKDD Explorer Newsletter*, *4* (2). 6-11.

[5]. Harris, Louis and Associates and Westin A.F. E-commerce & Privacy: What Net Users Want *Privacy & American Business*, Hackensack, NJ, 1998.

[6]. Hong, J.I. and Landay, J.A., An Architecture for Privacy-Sensitive Ubiquitous Computing. in *MobiSys*(2004, 177-189.

[7]. Jajodia, S. and Meadows, C. *Inference Problems in Multilevel Secure Database Management Systems*. IEEE Computer Society Press, Los Alamitos, California, USA 1995.

[8]. Kim, E., et al, Perceived Benefits and Concerns of Prospective Users of the SmartCampus Location-Aware Community System Test-bed. in *HICSS* 2007, 19.

[9]. Langheinrich, M., Privacy by Design – Principles of Privacy-Aware Ubiquitous Systems. in *UbiComp 2001*, 273-291.

[10]. Mifflin, H.e. *The American Heritage Dictionary of the English Language*. Houghton Mifflin Company, NY, 2004.

[11]. Narayanan, A. and Shmatikov, V., Obfuscated Databases and Group Privacy. in *12th ACM' CCS* (2005), 102-111.

[12]. Palen, L. and Dourish, P. Unpacking "Privacy" for a Networked World *CHI2003*.

[13]. Schuler, et al. Finding Your Way with CampusWiki: A Location-Aware Wiki to Support Community Building *The ACM's Conference on Human Factors in Computing Systems CHI2007*, San Jose California, USA., 2007.