Privacy Implications of Location and Contextual Data on the Social Web

Aristea M. Zafeiropoulou, David Millard, Craig Webber†, Kieron O'Hara

School of Electronics and Computer Science
University of Southampton
Southampton, UK
{az4g09, dem, kmo}@ecs.soton.ac.uk

†School of Social Sciences
University of Southampton
Southampton, UK
c.webber@soton.ac.uk

ABSTRACT
Location-based applications have recently begun to emerge on the Social Web. After their appearance numerous concerns with regards to location privacy have been provoked. However, these privacy concerns seem to have effects beyond location, as other contextual information can be inferred through location information. This research addresses these implications, which keep on growing on the Social Web.

Categories and Subject Descriptors
H.1.2 User/Machine Systems: Human information processing; J.4 Social and Behavioral Sciences: Sociology

General Terms
Human Factors, Theory

Keywords
Social Web, location privacy, contextual data, Privacy Paradox

1. INTRODUCTION
Current technologies offer users the opportunity to share their real-time location data; this includes location-based applications, like Foursquare, Facebook Places and Google Latitude. These applications are pieces of a greater concept, the Social Web, identified as a set of relationships that connect people to each other across the Web [1]. Privacy on the Social Web has been recognised as a significant area of concern, in terms of user awareness of privacy issues and ability to control private information [2]. Another aspect of privacy that raises concerns is the Privacy Paradox, the relationship between users' privacy perceptions and their actual information disclosure behaviours [3]. However, there is not yet a proper understanding of how the above-mentioned privacy issues map to location and situational context.

Our work is centered on location information on the Social Web and the privacy issues that arise in such situations. We focus on user behaviour, user awareness and the existence of the Privacy Paradox in these situations, and develop the argument that location privacy should be considered as part of a greater contextual or situational privacy.

Location privacy has been defined as a particular type of information privacy that focuses on the need of individuals to decide when and how others may access their personal location information [4]. However location information is part of a person's (physical) context [4]. As a result, through location information other contextual information that refers to an individual may be inferred.

The rest of this paper is divided into several sections: section 2 includes an overview of the related work in this area, whereas section 3 refers to the motivations for carrying out research in this area. Section 4 includes the key research questions and the future work that needs to be carried out.

2. RELATED WORK
Apart from the commercial location-based applications that exist, research groups have also developed their own location-based applications. Examples of such applications are Reno by Intel Research [5], MyCampus [6], iFind by MIT [7], Connecto [8] and Locaccino by Carnegie Mellon [9]. However, most of the research initiatives have not placed privacy as their top priority. Only Reno, iFind and Locaccino have considered privacy as a principle of significant importance throughout their design.

2.1 Location Privacy Studies
Location privacy has long been studied as an area of interest in Ubiquitous Computing. However, it appears to become an emerging area in the Social Web and as a result researchers have begun to study location privacy in this area as well.

A number of studies have focused on the user's privacy attitudes as well as on the privacy settings of applications on the Social Web and Ubiquitous Computing. Benisch et al. studied people's location sharing attitudes and found out that more complex privacy settings encourage people to share more [10]. Another study, by Burghardt et al., revealed that people prefer to use combinations of simple privacy mechanisms, instead of a single one that does not meet all privacy needs [11].

Apart from the privacy settings themselves, other factors influence people's sharing attitudes. A survey focused on Brighkite users (a commercial location-based social network)[1] showed that factors like age, gender, mobility and geographic area influence users' privacy concerns [12].

It has often been indicated that users are concerned about who has access to their location data. Lederer et al. found that users wish to

[1] http://brightkite.com/

have the same privacy preferences for an inquirer in any situation than different preferences for different inquirers in a specific situation [13].

2.2 Technical Approaches

Several technical solutions have been proposed to address the privacy issues raised in this area. Previous work has approached these issues on a computational level, by developing privacy algorithms (such as identity anonymisation algorithms); on an architectural level, by designing privacy protection systems; and on a user interface level (like the studies described earlier) [14]. Another approach that explicitly aims to protect location privacy is obfuscation. Obfuscation techniques reduce the quality of the users' location data in order to achieve their aim [15].

3. MOTIVATIONS

The problem in previous approaches is the lack of focus on the contextual aspects of location privacy. Research in location privacy needs to take place through a dynamic means, by taking into account not only location but also temporal factors [16]. Our research focuses on this kind of spatio-temporal privacy but we also argue to include other types of contextual and situational information.

Location information can be used to infer various types of contextual information, even in cases where the location information is anonymous [17]. For instance, research has shown that revealing the work and home location of an unknown person, reduces the anonymity set to which the user belongs, especially if work and home are in different regions [18]. Other data that can be inferred through the publication of location data include people's activities, real-time emotional and physiological status [19] and the presence of other people in the same location (co-location). Recent research also showed that social ties can be inferred by co-located photos uploaded in Flickr [20]. Another example of a location-based inference refers to location entropy, i.e. the measurement of the variety of people who visit a certain location. The entropy of the locations a person visits can indicate the number of social ties a person has within a network [21].

The importance of context in location information has also been emphasised by [22], who identified geographic location, material form and meaning or value to be the three main features of place.

3.1 Scenario

The following scenario illuminates the problem of location and contextual privacy by containing a number of the above-mentioned contextual data.

Alice is a regular smartphone user and on a daily basis she allows her phone to update her location information through a location-based application.

Mary, a friend of Alice, is also a smartphone user and has the exact same functionality set in her own phone.

A third party collects and stores the tracks of the locations of users of this specific application. As a result, the third party is aware of the movements of Alice and Mary.

The third party application also identifies and calculates the number of co-locations between the users. If the number of co-locations between any two users is significant, it is inferred that these two people are socially related. Apparently, Alice and Mary are often in the same location. Consequently, it is being inferred that these two users are socially connected.

Overall, this scenario demonstrates the inference of several contextual elements in practice:

- location
- co-location
- activity
- personal itinerary
- social tie

As described earlier, a location is often related with a certain activity. In addition to this, when the movements of a user between locations are recorded, the user's personal itinerary is revealed. Especially in cases where the recorded locations are recurrent, it is evident that they are part of an individual's regular itinerary.

The above-mentioned contextual elements can be classified into different degrees of data based on their inference complexity. For instance, location information is explicitly declared and consequently it belongs to the first degree of data. The second degree of data refers to data that are inferred from location data, such as activity and co-location. In addition to this, the inference of co-location information makes use of data from Alice and from another user who is known to Alice (in this case Mary's data). The third degree of data makes use of more complex heuristics, such as making inferences by combining Alice's data with the data from thousands of other users of the application who are unknown to Alice. An example application could be the identification of geographical hotspots based on the users' location tracking.

This scenario is an example of a privacy breach that tracks the movements of different users. The third party is on purpose not specified, as it could be a malicious application, the location-based application itself or even an individual. It is worth pointing out that apart from the social implications, there are also evident legal implications in this scenario.

The above-mentioned contextual aspects along with the described scenario emphasise the need for an in depth analysis of this type of location and contextual privacy. Significantly, location-based applications continue to evolve rapidly and as a result these contextual aspects are going to cause new privacy implications.

4. RESEARCH QUESTIONS

Based on this broader notion of contextual or situational privacy we identify three key research questions. First, what is the scope of this type of privacy? Second, does the Privacy Paradox apply to location data and if so does it perform in a different way than for other types of data? Third, what is the user awareness of contextual privacy and in particular what is the extent to which users are aware of the potential risks that the revelation of their contextual data may provoke?

The first question attempts to understand and analyse the contextual elements of location privacy described in section 3. Section 3 has indicated only few of the numerous contextual elements that underlie the exposure of location data. As a result it is of significant importance to identify the potential contextual elements and incorporate them in the study of location privacy.

The second question investigates the existence of the Privacy Paradox in location-based applications and whether people handle their location data in a different way than other types of data. It

examines whether people feel less comfortable sharing their location and contextual information than other information on the Social Web.

The third question explores in greater detail the users' attitudes regarding location and contextual data and studies the extent to which users are aware of the fact that the exposure of the above-mentioned contextual data may cause privacy breaches.

4.1 Future Work

The first step towards answering these questions is the design of a model that addresses the privacy implications of location and contextual data. The model aims to categorise that data identified in the literature retrieved from the latest Ubicomp and Mobile HCI Conferences. The different systems that are presented in the literature are going to be analysed and in the end a statistical analysis of the systems will be carried out, in order to shed light to important privacy related questions. An important feature of the model is the identification and inclusion of a variety of parameters that emphasise the richness of that contextual data. An example of such a parameter is the classification of data into different degrees based on the complexity of the inference mechanism in use. Other examples are the quality, fidelity and accuracy of the data as well as the data retrieval with or without user consent and knowledge.

The answers to these three research questions will provide valuable insights into location-based applications and their appropriate utilisation by the users, as they aim to understand in depth the notion of location and contextual privacy. In addition to this, they will contribute in developing better location-based applications that take into consideration the privacy of their users.

5. CONCLUSION

This paper has highlighted the importance of an in depth analysis of the privacy related issues that deal with location and contextual data. In order to achieve that, three research questions have been identified. The first question investigates the scope of location and contextual privacy, whereas the second one examines the existence of the Privacy Paradox in this specific type of data. The third question explores the user awareness regarding the privacy related issues that arise in location-based applications.

As more applications, which reveal users' location data, emerge on the Web it is of utmost importance to focus on the privacy implications of location data, understand its relationships with other forms of contextual data and develop frameworks that address these privacy issues.

6. REFERENCES

[1] Halpin, H., Tuffield, M. 2010. A Standards-based, Open and Privacy-Aware Social Web. *W3C. W3C Incubator Group Report*. http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb/.

[2] Loukides, G., Gkoulalas-Divanis, A. 2009. Privacy challenges and solutions in the social web. *Crossroads,* 16 (2), 14-18. DOI= http://doi.acm.org/10.1145/1665997.1666002.

[3] Norberg, P. A., Horne, D. R. and Horne, D. A. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs,* 41 (1), 100-126. DOI= 10.1111/j.1745-6606.2006.00070.x.

[4] Duckham, M. and Kulik, L. 2006. Location privacy and location-aware computing. Dynamic Mobile GIS: Investigating Change in Space and Time, pp. 34—51.

[5] Smith, I. and Consolvo, S. and Lamarca, A. and Hightower, J. and Scott, J. and Sohn, T. and Hughes, J. and Iachello, G. and Abowd, G.D. 2005. Social Disclosure of Place: From Location Technology to Communication Practices. *Pervasive Computing. Lecture Notes in Computer Science*. Springer, pp. 134—151.

[6] Sadeh, N., Gandon, F., and Kwon, O. B. 2006. Ambient intelligence: TheMyCampus experience. *T. Vasilakos and W. Pedrycz, editors, Ambient Intelligence and Pervasive Computing. ArTech House*.

[7] Huang, S., Proulx, F., and Ratti, C. 2007. iFIND: a Peer-to-Peer application for real-time location monitoring on the MIT campus. *International Conference on Computers in Urban Planning and Urban Management (CUPUM*).

[8] Barkhuus, L. and Brown, B. and Bell, M. and Sherwood, S. and Hall, M. and Chalmers, M. 2008. From awareness to repartee: sharing location within social groups. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (CHI '08). ACM, New York, NY, USA, 497-506. DOI=10.1145/1357054.1357134 http://doi.acm.org/10.1145/1357054.1357134.

[9] Toch, E. and Cranshaw, J. and Hankes-Drielsma, P. and Springfield, J. and Kelley, P.G. and Cranor, L. and Hong, J. and Sadeh, N. 2010. Locaccino: a privacy-centric location sharing application. *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing* (Ubicomp '10). ACM, New York, NY, USA, 381-382. DOI=10.1145/1864431.1864446 http://doi.acm.org/10.1145/1864431.1864446.

[10] Benisch, M. and Kelley, P.G. and Sadeh, N. and Cranor, L.F. 2010. Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing. Springer* 1—16, 1617-4909. DOI= http://dx.doi.org/10.1007/s00779-010-0346-0.

[11] Burghardt, T. and Buchmann, E. and Müller, J. and Böhm, K. 2009. Understanding User Preferences and Awareness: Privacy Mechanisms in Location-Based Services. *On the Move to Meaningful Internet Systems: OTM 2009, Springer* 304—321. DOI = http://dx.doi.org/10.1007/978-3-642-05148-7_21.

[12] Li, N., Chen, G. 2010. Sharing location in online social networks. *Network, IEEE*, vol. 24, no. 5, pp. 20-25. DOI= 10.1109/MNET.2010.5578914.

[13] Lederer, S. and Mankoff, J. and Dey, A.K. 2003. Who wants to know what when? Privacy preference determinants in ubiquitous computing. *CHI '03 extended abstracts on Human factors in computing systems* (CHI EA '03). ACM, New York, NY, USA, 724-725. DOI=10.1145/765891.765952 http://doi.acm.org/10.1145/765891.765952.

[14] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., et al. 2008. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing , 13* (6), 401-412.

[15] Duckham, M. and Kulik, L. 2005. A formal model of obfuscation and negotiation for location privacy. *Pervasive*

Computing, Lecture Notes in Computer Science, Springer, pp. 152—170.

[16] Duckham, M., Kulik, L., & Birtley, A. 2006. A spatiotemporal model of strategies and counter strategies for location privacy protection. *Lecture Notes in Computer Science , 4197/2006*, 47-64.

[17] Krumm, J. 2008. A survey of computational location privacy. *Personal Ubiquitous Comput. , 13* (6), 391-399.

[18] Golle, P., & Partridge, K. (2009). On the Anonymity of Home/Work Location Pairs. *Proceedings of the 7th International Conference on Pervasive Computing (Pervasive '09)* (pp. 390-397). Nara: Springer-Verlag.

[19] Riboni, D., Pareschi, L., & Bettini, C. 2009. Privacy in Georeferenced Context-Aware Services: A Survey. *Lecture Notes in Computer Science , 5599/2009*, 151-172.

[20] Crandall, D.J. and Backstrom, L. and Cosley, D. and Suri, S. and Huttenlocher, D. and Kleinberg, J. 2010. Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences, vol. 107, no. 52, pp. 2436}, issn 0027-8424.

[21] Cranshaw, J. and Toch, E. and Hong, J. and Kittur, A. and Sadeh, N. 2010. Bridging the gap between physical location and online social networks. *Proceedings of the 12th ACM international conference on Ubiquitous computing* (Ubicomp '10). ACM, New York, NY, USA, 119-128. DOI=10.1145/1864349.1864380 http://doi.acm.org/10.1145/1864349.1864380.

[22] Guerin, T. F. (2000). A Space for Place in Sociology. *Ann. Rev. of Sociology* , 463–496.