Urban Sensing through Social Networks: The Tension between Participation and Privacy

Ioannis Krontiris and Felix C. Freiling

Abstract In this paper we elaborate on using social networking mechanisms to create incentives that increase active participation of users in the process of collecting data for urban sensing campaigns and involve them in the process of sharing and processing the data. This results in a tension between the resulted social translucence and our efforts to offer anonymous submission of reports from mobile phones. We investigate specifically three research challenges: the identification of users at the network layer, the revocation of misbehaving anonymous users and the accumulation of reputation based on user's contributions.

1 Introduction

The emergence of GPS-capable mobile devices in combination with the wide adoption of mobile phones and the spread of the Web 2.0 paradigm on the Web recently created the right conditions for a new ecosystem of mobile services. Automatically geo-referenced user generated content can now be created anytime, anywhere and shared on the Internet, where the assembled information is aggregated and interpreted. In parallel, other sensors besides geo-location chips, such as camera, gyroscope, light sensor or accelerometer started becoming more and more prevalent in mobile devices carried by billions of people. As a result, user generated content has been extended to sensed data of the urban environment. Turning users to creators, custodians, actuators, and publishers of the data they collect, provided a unique chance to create a substrate for widespread public participation in data collection

Ioannis Krontiris
Chair of Mobile Business and Multilateral Security
Goethe University Frankfurt, Germany, e-mail: ioannis.krontiris@m-chair.net

Felix C. Freiling
University of Mannheim, Germany, e-mail: freiling@uni-mannheim.de

and the chance to create collective intelligence systems to address urban-scale problems, like air pollution, noise, traffic, etc.

Such systems, often referred to as *urban sensing* or *people-centric sensing* [1] systems, come to complement previous efforts to deploy wireless sensor networks to sense our environment and extend our possibilities by taking advantage of the large scale of sensors already existing on our hands. However, the involvement of people in the process opens up new challenges; with the mobile device gathering sensor data from user's immediate environment and deriving user context, privacy concerns are rightfully raised. Over the last years there is an increasing public awareness of privacy and several studies present convincing data that such concerns have an impact on people's acceptability and adoption of these new technologies.

Then, how can scientists motivate people to adopt this new technology and participate, offering their sensing capabilities? Indeed some consumers are willing to sacrifice privacy for benefits they value [2], as for example personalized services that current location-based services (LBSs) offer. The success of people-centric sensing, however, depends on the willingness of volunteers to devote their time to help with the data collection task, like many crowd-sourcing services on the web, without direct benefits. Therefore, retaining their privacy becomes an even more prominent requirement for such projects.

Protecting privacy would not be enough to guarantee people's participation. Currently, many urban sensing projects call people to participate in projects with the goal to collect environmental data to a server. These calls have been, to put it politely, only moderately successful. The focus of research should be extended to investigate how we can make people involve more actively in urban sensing projects. *In the first part* of this paper, we elaborate on the utilization of social networks towards this goal. In particular, social networks can provide excellent tools to recruit more people in sensing campaigns, give them incentives to actively participate, as well as enable them to debate about the data and take actions. Overall, we argue that social translucence within online communities can provide many benefits for urban sensing projects. Social translucence is a term proposed by Erickson and Kellogg [3] to refer to "digital systems that support coherent behavior by making participants and their activities visible to one another".

However, interestingly enough, there is a vital tension between the above two key factors: on one hand preserving privacy and on the other hand offering social translucence. *In the second part* of this paper, we investigate three research problems that emerge from this tension. In particular, we allow users to report location-specific sensor data, while preserving at the same time their *anonymity*. That is, all data sent to the service provider do not include any identifying information of the sender. The first challenge we explore is how we can prevent an attacker from using network-level identifiers to infer the identity of users. Then we move to the data layer and discuss how to address the problem of revoking users, who are covered behind their anonymity to misbehave against the system. Finally, we investigate how to enable users to accumulate reputation points for their public profiles.

2 Utilizing Social Networks

Currently there is an increasing number of mobile sensing applications that focus on individuals and target to deliver a service back to them, for instance computing their daily exposure to pollution, keeping track of their exercise activities, dietary habits, etc. Some of these services, like CenceMe [4], allow users to share this information through social networks. However, besides self-reflective projects, urban sensing also targets community-based ones, where users do not have a direct benefit from offering their sensing possibilities, but they are rather motivated by a common cause or interest, similar to the participative paradigm of Web 2.0. Let's take for example NoiseTube [5], which enables citizens to measure their personal exposure to noise in their environment and share this information to produce a collective noise map.

We share this vision of a sensor data-sharing infrastructure, where people and their mobile phone devices provide their collected data streams in accessible ways to third parties interested in integrating and remixing the data for a specific purpose/campaign. This new personal measurement instruments can enable an entirely novel and empowering genre of research in citizen science [6, 7] and urban planning. Paulos has recently framed some important questions within four core components of citizen science: collect, express, share and change [8].

Current research projects on urban sensing concentrate mostly on collection and expression of sensor data: which types of environmental or human conditions should we sense for a given campaign and how much should users be actually involved in the process of reporting the data? How can collected data best be represented and experienced? Here we emphasize on the last two components, namely share and change, as they are relatively unexplored and pose the biggest challenge:

- Share – How will collected data be shared? What practices of individual ownership will be appropriate and how will privacy be addressed? How can data best be shared with non-experts, urban planers, decision and policy makers, etc.?
- Change – What tools or frameworks best invite and encourage active participation? What tools and techniques will facilitate the most productive debate and ultimate positive social benefit?

The existing dynamics in social networks and online communities in general, which have found wide adoption in the Internet during the past years, is a promising direction to answer the above questions. That is, through social networks, we can involve people not only in the process of data collection, but also in sharing and acting on the data. The citizens who interact with the physical environment and contribute the data are also the members of social networking groups and therefore citizens participate in the complete life-cycle of the data.

Here we refer to social networks not so much as an individual-centered service, but more as a group-centered one and therefore as a way to form online communities. Indeed, social networking platforms, like Facebook, offer the tools to form groups not necessarily based on personal relationships, but rather on common interests [1]. Below we elaborate on some specific benefits that social networking mechanisms can offer to urban sensing campaigns.

Recruitment

Recruitment of participants in urban sensing campaigns will be a determinant factor for the success of their outcome. The organizers of campaigns, either being community groups or simply motivated individuals, should be able to attract interested and well-suited participants for a campaign, based on the needs and specifications of the case they want to make. Taking advantage of the dynamics in social-networking websites can offer urban sensing projects a powerful tool to recruit such participants.

In particular, social networks can enable organizers to identify and reach well-suited participants for data collections based on their geographic availability as well as their interests and habits. Indeed, urban sensing projects are usually initiated for a specific geographic area and refer to people with specific sensitivities. Public information on their profiles could be a selective criterion to contact specific people (or even whole groups) and invite them to participate. The number of previous campaigns undertaken and the success of a participant in them (i.e. his reputation) can also form a selection criterion [9].

Social networks also incorporate mechanisms that allow existing participants to invite their friends to join a group, or see what their friends are doing (which group they joined, in which groups they are most active), etc. This creates a “word of mouth” that can help a campaign reach the desirable number of volunteers.

Participation

We separate between getting people to join an urban sensing campaign and getting them to actually participate. Why should those who can produce the sensing data take the time to engage in such interactions? Why should they wish to? What benefits would they gain that might compensate them for their efforts? In most cases, urban sensing projects call users to volunteer and offer the sensing capabilities of their mobile devices without getting any immediate return.

The sense of community is known to be a motivation for contribution to online communities. Kollock becomes more specific and he outlines three motivations that do not just rely on altruistic behavior on the part of the contributor: anticipated reciprocity, increased recognition and sense of efficacy [10]. From those, perhaps the most important and most relevant to urban sensing is the fact that individuals desire some sort of recognition for their contributions. This is related to having some kind of online identity, basically a public profile, on which a person can build a reputation and receive attention.

[1] While on online communities people usually use pseudonyms, in social networks they usually use profiles corresponding to their actual identities. However, nothing stops users from building fake profiles, so here we treat both as online identities and we do not differentiate between social networks and online communities.

Translating that for our scenario, a user, after submitting a report from his mobile phone, is given a reputation point. Reputation points assigned to each user sum up to create that user's reputation value. In addition, reputation values are public and appear on the public profile of that user to express the degree of his/her contribution to the sensing campaign.

Acting on the data

Involving people in the data collection process will unavoidably create problems connected to the behavior of these people. As Paulos put it, we need to "design for doubt" and address the range of potential underlying possibilities for sensor failures and user errors, both accidental and malicious in intent [8].

First, some data are more useful than others. Consider the case we are interested in forwarding the collected data to the authorities, responsible in taking actions based on them. An example could be that people report a problem from the urban environment (e.g. a dangerous pothole on a road surface), so that proper authorities (e.g. the relevant road agency) can take actions. However, these authorities will be reluctant to evaluate the amount of collected data. Some sort of pre-evaluation or filtering could happen from users themselves to decide on which of this data action is imperative. Let us note here that direct evaluation of data by users is not always possible (e.g. how can users evaluate submitted noise levels?), so the process should also be supported by some tools and algorithms at a lower level. Social networks also offer a very easy and open way for users to debate about the data, which could result in some social actions directly by the citizens.

3 Emerging Privacy Issues

Using a cellphone for collecting information from the environment and tagging them with time and GPS data, could be used to infer a lot of personal information, including the user's identity. This problem is often termed *location privacy*. Knowing when a particular person was at a particular point in time can be used to infer the personal activities, habits, political views, health status, profession and social interactions of that person. However, the so-called profiling is not the only threat. The location of a user could be exploited for unsolicited advertising, in order to provide advertisements of products and services available to the user's position. Physical attacks or harassment is also another threat of knowing a person's current or favorite location. Location information is therefore in many cases a particularly sensitive piece of personal data and people have now started to realize more and more the need for location privacy.

However, from our discussion so far, it is evident that in order to take urban sensing one step further, we need to introduce some kind of social translucence, which will allow users to make their contributions visible to the online community. As Erickson and Kellogg mention in their work, we speak of socially translucent

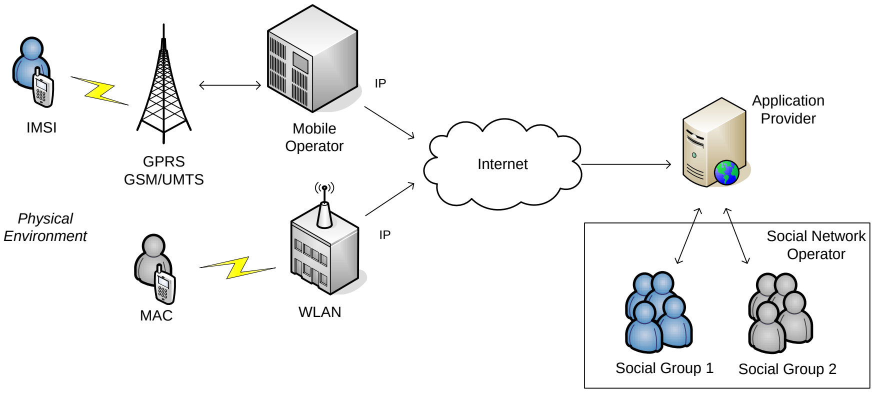

Fig. 1 Submitting de-identified data from a mobile phone does not preserve the anonymity of the custodian, as many identifiers are involved in the network layer. Also the knowledge of the social network that an anonymous user belongs to, could affect the offered anonymity.

systems rather than socially transparent systems, because there is a vital tension between privacy and visibility [3]. We need to keep certain types of one's behavior public and inhibit the publication of others.

In particular, we can translate the above problem in the pervasive setting of our case as follows: Is it possible to offer anonymity to the user, who submits sensing data from the physical environment, while at the same time we maintain properties connected with his online identity, like reputation or accountability? In the following sections, we first look at the problem of protecting user anonymity at the network level and then we discuss two important problems at the data level: revoking access credentials of anonymous misbehaving users and providing an anonymous reputation system. So, in a way, while at the previous section we were looking to bring the users from the physical environment closer to the online communities, now we seek ways to maintain these benefits, but also separate their physical identity from their online identity.

Network Level Anonymity

Let us assume that any identifying information has been removed from the data, so it includes only the sensing information, the GPS value and the time of measurement. This is not enough to provide anonymity to the user, if we do not first of all protect identifying information at the network level. Network identifiers can be used, either to reveal the identity of the user directly or link several reports back to the same user and therefore build a location profile of that user.

Figure 1 depicts the communication paths between the two communication ends in a generic urban sensing architecture: the mobile users and the application provider. There are (at least) two network access possibilities for the user: through a data telecommunications service, like GSM or UMTS and through a (possibly open) WLAN access point.

Providing anonymity at the first hop of communication, i.e. between the user and the mobile operator or the Wi-Fi hotspot, is a problem that falls outside the scope of this paper. Here we consider attackers who are able to observe the traffic over the Internet, between the access point and the service provider. At this level the goal is to provide communication anonymity, which means hiding the network identifiers in the network layer (i.e., IP addresses).

Since mixes were proposed in 1981 [11] as a solution for achieving anonymous communication, multiple other protocols appeared in the literature in order to provide anonymity over the Internet. In particular, low-latency anonymous overlay networks seek to provide, from the user's point of view, a reasonable trade-off between anonymity and performance. Some of the most prominent low-latency approaches include Crowds, Tor, Jap, and Onion Routing. Still, only a few of these anonymizing networks have been tested for the mobile Internet scenario and it is an area that only lately attracted research interest [12]. Even though it is not hard to adapt protocols like Tor to conform to the mobile internet constraints, other more lightweight solutions remain to be investigated [13].

In our scenario, the interconnection of users through social networks creates a different setting for the evaluation of the performance by anonymous communication networks. Here, an attacker, besides her observations at the communication layer, has also knowledge from the application layer, i.e., the identities of the users that participate in the system and how they are related, through their profiles in the social network. Users organize themselves into a community with a common goal, and these users are expected to send measurements for the corresponding campaign. There is an *a priori* knowledge of user profiles and associations that can be combined with data gathered by traffic analysis of the mix-based network.

Diaz et al. studied the problem of measuring anonymity based on profile information [14] and social networks [15] and showed that user profile information does not *necessarily* lead to a reduction of the attacker's uncertainty. The assumptions in this work include a 1-to-1 communication paradigm, where individuals communicate with each other directly, as well as a global passive adversary model, where the attacker can observe all the inputs and outputs of the anonymous communication network. Generalizing the first and relaxing the second assumption certainly creates an interesting but also challenging problem.

Revocation of misbehaving users

For most urban sensing applications it is essential to enforce access control in order to prevent service abuse and to protect against malicious attacks. Access to services for users offering the data should be granted only based on pre-established trust between users and the service provider. Given that we also want to preserve anonymity, this leads to a chicken-and-egg conundrum. On one hand, a user has to be authenticated before accessing a service; on the other hand the users ID can serve as a unique identifier that can be used to track the users whereabouts and actions.

In response to this problem, a lot of research work has focused on anonymous user authentication that targets user privacy while maintaining access security. The basic idea has been to verify the users right to access a service, while at the same time the users identifying information remains secured. This immediately creates an important requirement: the support of *user revocation*. The anonymous access to a service offers users a high degree of privacy and along with it the license to misbehave without the fear of punishment. Therefore we want to be able to deanonymize misbehaving users and limit their access to the system.

An approach for enhancing anonymous authentication is to use *group signatures* [16], where a vast amount of research is being carried out worldwide. These technologies can be used to verify whether or not a user is allowed access, without actually identifying the user. This is achieved by allowing a member of a group to sign a message on behalf of the group, without revealing which member produced the signature. Group signature systems can support revocation, where group membership can be selectively disabled without affecting unrevoked members.

In order to apply group signatures for mobile phones and users belonging to highly dynamic communities, we need to address a number of problems that come with this solution. For example, in online communities members continuously come and go and a solution to change and re-distribute fresh certificates to all members each time is not a viable solution. This problem has been addressed by anonymous credential systems that support dynamic membership revocation [17].

Existing group signature solutions are based on a trusted third party (TTP), which has the ability to revoke a user's privacy at any time. This becomes problematic, since users can never be assured that their privacy will be maintained by that TTP. To eliminate the reliance on TTPs, certain "threshold-based" approaches such as e-cash [18] and k-Times Anonymous Authentication (k-TAA) [19] have been proposed. In these schemes, no one, not even an authority, can identify a user who has not exceeded the allowed number of k authentications or spent an e-coin twice.

However misbehavior in urban sensing applications is not defined as overusing a service. In our case, we are interested in revoking users who upload data, which after a specific process are judged as "inappropriate". When they have been judged to have repeatedly misbehaved at least d times, they should be revoked by the system. This problem has been addressed recently by Tsang et al. [20], who proposed a d-strikes-out revocation scheme for blacklisting misbehaving users, without relying on a TTP. Unfortunately the computational and communication overhead of the protocol is not attractive for power-limited devices such as mobile phones, especially as the size of the blacklist grows.

Reputation

As we discussed in Section2, offering reputation points to people submitting data, can form a sort of recognition to their efforts. A direct process of acquiring reputation points for a given public profile would clearly compromise the privacy of the submitter, as it would link the location information inside the report with his online identity. Therefore, we need to disassociate the process of acquiring reputation points from updating the reputation value on someone's profile. One solution could be to use an e-cash system, where each user obtains e-coins from the bank for each report submission, each one corresponding to a reputation point.

With some e-cash schemes a pseudonym, which is not traceable to one's identity, is obtained by the user during the account establishment protocol. This pseudonym is visible with all coins withdrawn from the bank. Coins are thus anonymous, but linkable, which means that the bank can also link the reports and build a location profile for the user submitting them. Therefore, pseudonyms must frequently change, in order to decouple identity from location-time information.

This directly contradicts with the notion of reputation. In its most general form, reputation is memory about past performance and it is bound to each pseudonym. Consequently, changing pseudonyms results in loss of the accumulated reputation. On the other hand, not changing pseudonyms compromises the anonymity and unlinkability of the user. A solution would be a scheme that allows users to change pseudonyms as many times as needed and at the same time transfer the gained reputation to the new one. The challenge here is that the transfer of the reputation from one pseudonym to the other should happen in such a way that it does not link the two pseudonyms together [21]. Finally, something that calls for further discussion is how to make such a solution work with more permanent profiles in social networks.

4 Conclusion

Utilization of social networks could provide many benefits in urban sensing. Social networking mechanisms can help us recruit more citizens in campaigns and boost their active participation not only in the data collection process, but also in the evaluation and utilization of the data. In particular, one's reputation in the community could become an incentive for people to contribute data. At the same time, it is also important to preserve the anonymity of the users submitting data from the environment. This requires solutions both at the network and the data layer. The desired visibility in the social networks makes our efforts for such privacy solutions more challenging. User profile information could help attackers observing network traffic reduce their uncertainty. Anonymity also makes it harder to revoke misbehaving users or compute their accumulated reputation points. Some solutions on these problems have been proposed, but the pervasive nature of urban sensing and the persistent nature of social profiles create a new setting and call for further research.

References

1. A. Campbell, S. Eisenman, N. Lane, E. Miluzzo, R. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn, "The rise of people-centric sensing," *IEEE Internet Computing*, vol. 12, no. 4, pp. 12–21, 2008.
2. N. Awad and M. Krishnan, "The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization," *MIS*

Quarterly, vol. 301, pp. 13–28, 2006.

3. T. Erickson and W. A. Kellogg, “Social translucence: an approach to designing systems that support social processes,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, no. 1, pp. 59–83, 2000.
4. E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, “Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application,” in *Proceedings of the 6th ACM conference on Embedded network sensor systems (SenSys ’08)*, pp. 337–350, 2008.
5. N. Maisonneuve, M. Stevens, M. Niessen, and L. Steels, “Noisetube: Measuring and mapping noise pollution with mobile phones,” in *Proceedings of the 4th International Symposium on Information Technologies in Environmental Engineering (ITEE 2009)*, May 2009.
6. A. Irwin, *Citizen Science: A Study of People, Expertise and Sustainable Development*. Routledge, 1995.
7. E. Paulos, R. Honicky, and B. Hooker, *Citizen Science: Enabling Participatory Urbanism*, ch. 28. 2008.
8. E. Paulos, “Designing for doubt, citizen science and the challenge of change,” in *Proceeding of First International Forum on the Application and Management of Personal Electronic Information (Engaging Data)*, (Cambridge, MA), October 2009.
9. S. Reddy, K. Shilton, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, “Evaluating participation and performance in participatory sensing,” in *Proceedings of the International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems (UrbanSense ’08)*, November 2008.
10. P. Kollock, *The Economies of Online Cooperation: Gifts and Public Goods in Cyberspace*, ch. 9, pp. 220–239. 1999.
11. D. Chaum, “Untraceable electronic mail, return addresses, and digital pseudonyms,” *Communications of the ACM*, vol. 24, no. 2, pp. 84–90, 1981.
12. J. Lenhard, K. Loesing, and G. Wirtz, “Performance measurements of Tor hidden services in low-bandwidth access networks,” in *Proceedings of the International Conference of Applied Cryptography and Network Security (ACNS ’09)*, pp. 324–341, June 2009.
13. I. Krontiris and F. C. Freiling, “Integrating people-centric sensing with social networks: A privacy research agenda,” in *Proceeding of the IEEE International Workshop on Security and Social Networking (Sesoc)*, 2010.
14. C. Diaz, C. Troncoso, and G. Danezis, “Does additional information always reduce anonymity?,” in *Proceedings of the 2007 ACM workshop on Privacy in electronic society (WPES ’07)*, pp. 72–75, 2007.
15. C. Diaz, C. Troncoso, and A. Serjantov, “On the impact of social network profiling on anonymity,” in *Proceedings of the 8th international symposium on Privacy Enhancing Technologies*, 2008.
16. D. Chaum and E. van Heyst, “Group signatures,” in *Advances in Cryptology - EUROCRYPT ’91*, pp. 257–265, 1991.
17. D. Boneh and H. Shacham, “Group signatures with verifier-local revocation,” in *Proceedings of the 11th ACM conference on Computer and communications security (CCS ’04)*, (New York, NY, USA), pp. 168–177, ACM, 2004.
18. J. Camenisch, S. Hohenberger, and A. Lysyanskaya, “Balancing accountability and privacy using e-cash (extended abstract),” in *Proceeding of the 5th Conference of Security and Cryptography for Networks (SCN ’06)*, pp. 141–155, 2006.
19. I. Teranishi, J. Furukawa, and K. Sako, “k-times anonymous authentication,” in *Proceedings of the 10th International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT ’04)*, pp. 308–322, December 2004.
20. P. P. Tsang, M. H. Au, A. Kapadia, and S. W. Smith, “BLAC: Revoking Repeatedly Misbehaving Anonymous Users Without Relying on TTPs,” 2010.
21. E. Androulaki, S. G. Choi, S. M. Bellovin, and T. Malkin, “Reputation systems for anonymous networks,” in *Proceedings of the 8th international symposium on Privacy Enhancing Technologies (PETS ’08)*, (Leuven, Belgium), pp. 202–218, 2008.