Integrating people-centric sensing with social networks: A privacy research agenda

Ioannis Krontiris, Felix C. Freiling
Laboratory for Dependable Distributed Systems
University of Mannheim

***Abstract*—During the last few years there has been an increasing number of people-centric sensing projects, which combine location information with other sensors available on mobile devices, such as the camera, the microphone or the accelerometer, giving birth to a different dimension in sensing our environment compared to the existing wireless sensor networks approach. In this paper, we envision a new scenario, where users develop their own participatory urban sensing projects at a large scale through the use of social networks. Consequently, users can participate in campaigns created by other users, according to their sensitivities and interests, exploiting the existing enormous social interconnections offered by existing social networking tools. We place our primary concern to protecting user privacy and address the need for new solutions in location anonymity and access control under this new complex and dynamic communication paradigm.**

I. Introduction

Over the last years geo-location chips along with other sensors, such as camera, microphone or accelerometer are becoming more and more prevalent in mobile devices carried by billions of people. This provides us with a substrate for widespread public participation in data collection in the urban environment 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 "people-centric sensing", come to complement our 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.

Next to the benefits that this new approach has, it also poses new challenges. Exactly because participatory sensing is based on people, its success, like many crowd-sourcing services on the web, depends on the *willingness* of volunteers to devote their time to help with the data collection task and most of the times without direct benefits for them. Since people do not get a direct benefit with respect to their identity and location, as for example in location-based services (LBS), retaining their privacy becomes an important and necessary requirement.

One way to deal with privacy is to let users choose the privacy policies offered by service provider in some form of contract, which states what data will be collected, for what purpose and how it can be distributed. As argued by Spiekermann and Cranor [1], privacy by policy offers the minimum degree of protection and systems utilizing such solutions should make users aware of privacy risks and offer them choices to exercise control over their personal information. For example, Shilton et al. [2] recently proposed design principles for urban sensing, engaging the participants in the ethical decision-making and the negotiation of personal boundaries and identities.

On the other hand, anonymity provides higher levels of privacy, making the system tamper-proof against stronger attackers, who would not be deterred by policies and regulations. Towards this goal, techniques for achieving anonymity both on the network and data level must be combined, as there is no real anonymity on the data level, without anonymity on the network level.

In this paper we explore the research challenges for providing anonymity on the network level. In addition, we motivate the integration of sensor data with social networks and we discuss the new privacy challenges that arise from this participatory sensing paradigm.

II. Related Work

While the importance of privacy is mentioned by most urban sensing projects, there are currently only a few proposed solutions towards this direction. One approach is AnonySense [3], a general-purpose architecture for maintaining the privacy of the users in *opportunistic* urban sensing applications. The term opportunistic sensing refers to systems where the custodian's device is utilized by the system whenever its state (e.g. geographic location) matches the requirements of an application, without the custodian being aware of the sensing activity. Here we focus on *participatory* sensing, where the user consciously opts to meet application requirements for data out of personal interest.

An other approach that we also mentioned in the previous section is privacy by policy. A current research direction on providing privacy for urban sensing systems attempts to engage participants themselves to answer privacy dilemmas [2]. It has been shown that how people choose to withhold or disclose information about them depends highly on their context, e.g. identity, situation, time, or culture. Therefore, the above approach attempts to provide the tools to negotiate sharing and discretion according to

personal context and preferences. It concentrates on data level anonymity, i.e., preserving the confidentiality of user data in the application layer. In this paper we concentrate on the network layer anonymity and we are interested in hiding the network identifiers of the user in the network layer.

III. Motivating Integration with Social Networks

Our physical world contains more sensory data than we can possibly comprehend. Involving people in the data collection process can greatly narrow down observations via critical decisions, reality checks, and inferences. Which data is important? How much do we need? Humans can figure out how to collect public sensing data by making opportunistic choices on the spot, taking into consideration immediate factors not possible using digital methods. We can use these dynamics to build systems where people are the main contributors and consumers of the data, given that they are motivated by a common cause to offer their sensing possibilities. We argue that this *participatory* model is more likely to gain the trust of people in future applications, compared to opportunistic sensing, where the custodian's device is utilized by the system without the custodian being aware of the sensing activity.

We envision 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. People should be able not only to control the time and place that their personal device measures and sends information from their immediate environment, but also *be aware of what that information is being used for*. People should receive a meaningful benefit in exchange for sharing data. Meaningful benefits include compelling applications based on anonymous learning from "users like me". People should be able to enjoy the benefits of these services simply in exchange for their data.

Recruitment of participants in such 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. Web 2.0 has already initiated a new age of user-created content and participative web. The mass adoption of social-networking websites is causing a major shift in the Internet's function and design and is turning it into a tool for connecting people, who can also create content that everyone can share.

Therefore, we propose leveraging existing open Web 2.0 services like social networks, in order to offer a tool to the users for recruiting people and creating a user base for goal-based sensing projects, bringing social networks and the physical world one step closer. On the long run, the ultimate goal is to foster research and accelerate innovation in defining novel use cases and applications for the urban sensing paradigm. We argue that instead of searching for the next killer applications, we must seek incremental solutions where the combination of user-generated data and social interconnections of people can lead to a new type of knowledge and thus establishing potentially new use cases.

IV. Privacy Research Challenges

When designing a privacy protection system, one has to consider first the privacy risks that the users are subject to, depending on different attacker models. While there are well known mechanisms for understanding security risks, we lack mechanisms for evaluating privacy risks, especially in pervasive computing environments. Without the right privacy risk models it is difficult to understand at which extent the privacy technologies are needed to address those risks and develop architectures, interaction techniques, and strategies for managing them. This section targets to set the ground for evaluating existing privacy mechanisms for participatory sensing and designing new ones.

A. Challenge I: Defining appropriate attacker models

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. In such a communication paradigm, the behavior of users leaves a lot of traces. These traces are generated during data communication due to different commercial, technical and legal requirements and they can occur over the two different communication hops: between the user and the access point (mobile operator or Wi-Fi hotspot) or between the access point and the services provider. Basically, all involved stakeholders can potentially try to upset the users privacy, even by colluding with each other.

In the case a user uses his mobile operator to connect to the Internet, information like the IMSI (International Mobile Subscriber Identity) and the IMEI (International Mobile Equipment Identity) can be used to directly identify the user. In case he uses a Wi-Fi spot, the unique MAC address of his mobile device is associated to the Access Point. There are also many stakeholders in the scenario of participatory sensing: the user, the mobile network operator, the operator of the WLAN access point, organizations running the Internet backbone, and finally the social network application provider. Theoretically, all these stakeholders could try to spy simultaneously on the users identity when sensing. However, in practice there are usually smaller groups colluding with each other.

- Who is actually capable of committing attacks and which are the needed technical capabilities for these attacks?
- What would be the motivation of each stakeholder to perform an attack?

- With whom does it make sense for each stakeholder to collude in the real world?

From a theoretical point of view, the worst-case individual stakeholder, who could turn malicious, is the mobile operator. Because it is necessary for network management and billing, the mobile operator directly observes identifying information like the IMSI. It is hard to establish any form of anonymity against the mobile operator, and this is currently an open problem of research. Another alternative for the user would be to use other mechanisms of network access, like open WLAN access points. If the WLAN operator is also malicious, things become even harder. Investigating these attacker models therefore lead to interesting and challenging research settings.

Research directions and related work: We suggest first looking at all combinations of malicious entities and then identifying those that seem appropriate in the participatory sensing paradigm as well as pose theoretical and practical challenges. In particular, we propose on one hand the construction of a theoretical attacker model in order to make statements about the security of abstract models of anonymization networks in participatory urban sensing. This can allow us to find basic statements on security properties even for real systems and explore the limits of privacy provision. On the other hand, we should develop more practical-oriented attacker models, which can be used for analysis of deployed implementations and provide end-users with reasonable level of privacy protection.

B. Challenge II: Sender Anonymity and Unlinkability

Urban sensing is an emerging scenario, where a single infrastructure integrates heterogeneous technologies such as wired, wireless and cellular networks. One of the main challenges is to allow users to report location-specific sensor data while preserving at the same time their anonymity. The first approximation of the term anonymity for the system means to provide *sender anonymity*, i.e. the identity of the sender of a message must be hidden to external parties, including the receiver itself.

Towards this goal we need to investigate whether the existing solutions for providing anonymity can be applied in such a complex environment, evaluate them and propose new solutions where needed. Depending on the strength of the attacker model, we need to study the appropriate anonymity techniques and evaluate the quality of the provided protection with respect to their anonymity and performance properties. We expect that posing requirements such as usability, availability and trust will bring up needs for new solutions for this new scenario, which we need to address by proposing the corresponding measures or adjustments to the architectural design.

The determinant factor here is performance. Performance plays a much more important role in the mobile Internet than it does in the traditional wired Internet. Mobile networks generally have much lower bandwidth capabilities and more transmission errors than wired networks, a fact that causes higher latency.

The task therefore is to investigate how and whether privacy can be enhanced in the urban sensing paradigm with a reasonable tradeoff between anonymity protection and performance loss. We identify therefore the challenge of investigating and evaluating the multitude of anonymity techniques for their suitability in participatory people-centric sensing. In particular, some more concrete research question would be the following:

- How much does latency of anonymizing networks in the mobile Internet affect users' participation in anonymous people-centric sensing?
- How much does this latency affect the provisioning of services back to the users?
- To which extent can we quantify and compare the security and performance properties of available anonymity techniques, when applied to the communication paradigm of people-centric sensing?

Research directions and related work: In the literature, existing solutions for network layer anonymity and unlinkability of user actions are categorized into three groups: proxies, peer-to-peer (P2P) networks and Mix networks. Here we concentrate on the last two groups to find an appropriate solution for our scenario.

1) Tor: The Tor-network is currently the network with the most number of users and related research publications. Before a client can use the network, he has to get the network information about the available servers from a cascaded cache group of dedicated directory servers. Recently Lenhard et al. [4] made performance measurements of the usage of Tor in cellular phone networks and showed that the bootstrapping phase has turned out to take significantly longer than expected in this case (about 232.9 seconds). So, downloading the relay descriptors forms the bottleneck in this phase.

2) AN.ON: Formerly initiated by the German project AN.ON (ANonymity ONline), the project is sometimes also referred to as JAP, which is in fact the name of the client side-software. In contrast to Tor, here a cascade is accepted by the central authority only if the nodes are providing a fair amount of bandwidth, resulting in a better quality of service. Also AN.ON does not require the downloading of directory information. Even though it does not have forward secrecy and does not supports arbitrary TCP traffic (except HTTP and HTTPS), we consider it a solution worth being investigated for our scenario.

3) Ant-routing: Another promising direction is anonymizing P2P systems, which are designed for anonymous file exchange and they are based on ant routing algorithms for ad-hoc networks. The main representatives are the anonymizing network Ants and Mute. Even though they have very limited academic coverage so far and a

very small number of users, they are more attractive as a research direction for our scenario, in terms of performance. As these networks are mere peer-to-peer networks, all the participants also act as intermediary nodes. Any new user needs to learn some of these identities in order to connect to the network. For this reason, participants only get to know small parts of it upon arrival of a new node and there is no central point where all information is gathered together at any given time.

C. Challenge III: Integrating with social networks

In the typical situation studied extensively in the bibliography for anonymous communication, N users send messages to each other through a mix-network and anonymity is based on creating uncertainty concerning the identity of the subject who originated or received a message. As the number of users in the system increases, the probability of being linked to a particular action decreases. The theoretical analysis is usually based on the assumption that senders choose the receivers of their messages uniformly at random.

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 groups 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.

Clauß and Schiffner argue that an adversary with access to more information is always able to reduce anonymity [5]. But later, Diaz et al. [6] showed that user profile information does not *necessarily* lead to a reduction of the attacker's uncertainty. So, the corresponding question of interest is the following:

- Does the knowledge of user profiles and interconnections through social networks reduce the offered anonymity when integrated in the people-centric sensing paradigm?

Research directions and related work: To answer the above question we need to evaluate (quantify) the offered anonymity by an anonymous communication network. This has been proved a very difficult task so far. One method of measuring anonymity is based on the entropy of the probability distribution linking an action to all possible subjects that may be related to it [7]. However, the combination of several sources of information in entropy-based anonymity metrics is not yet well understood.

Diaz et al. studied the problem of measuring anonymity based on profile information [6] and social networks [8]. In these papers an 1-to-1 communication paradigm is followed, where individuals communicate with each other directly. Furthermore a global passive adversary model is assumed, 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 very challenging problem.

V. Conclusion

In this paper we motivate and discuss the integration of social networks into people-centric participatory sensing. We concentrate more on the privacy challenges in such a setting and emphasize on the importance of first defining the right attacker models. Then we elaborate on the suitability of existing anonymization techniques and we discuss on the need to study the effect that social interconnections of users have on their anonymity. We believe that future pervasive computing systems that call for people's participation in the interaction with our environment will pose similar challenges and certainly addressing these challenges will lead to a more solid understanding of privacy.

References

[1] S. Spiekermann and L. Cranor, "Engineering privacy," *IEEE Transactions on Software Engineering*, vol. 35, no. 1, 2009.

[2] K. Shilton, "Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection," *Communications of the ACM*, vol. 52, no. 11, pp. 48–53, 2009.

[3] C. Cornelius, A. Kapadia, and N. Triandopoulos, "AnonySense: privacy-aware people-centric sensing," in *Proceeding of the 6th international conference on Mobile systems, applications, and services (MobiSys '08)*. Breckenridge, CO, USA: ACM, June 2008, pp. 211–224.

[4] 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)*, June 2009, pp. 324–341.

[5] S. Claußand S. Schiffner, "Structuring anonymity metrics," in *Proceedings of the second ACM workshop on Digital identity management*, 2006.

[6] 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)*, 2007, pp. 72–75.

[7] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Proceedings of 2nd International Workshop on Privacy-Enhancing Technologies*. Springer-Verlag, April 2002.

[8] 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.