Location Privacy in Urban Sensing Networks: Research Challenges and Directions

Ioannis Krontiris[*], Felix C. Freiling[†]and Tassos Dimitriou[‡]
[*]Goethe University Frankfurt, Germany
[†]University of Mannheim, Germany
[‡]Athens Information Technology, Greece

Abstract—**During the last few years there has been an increasing number of people-centric sensing projects. These combine location information with sensors available on mobile phones, giving birth to a different dimension in sensing our environment and providing us with new opportunities to create collective intelligence systems to address urban-scale problems, like air pollution, noise, traffic, etc. However, as people are directly involved in the collection process, they often inadvertently reveal information about themselves, raising new and important privacy concerns. While standard privacy enhancing technologies exist, they do not fully cover the many peculiarities of these new pervasive applications. The ubiquitous nature of the communication and the storage of location traces compose a complex set of threats on privacy, which we overview in this article. Then, we go through the latest advances in security and privacy protection strategies and we discuss how they fit with this new paradigm of people-centric sensing applications. We hope this work will better highlight the needs for privacy in urban sensing applications and spawn further research in this area.**

I. 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, light sensor, gyroscope 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 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 and collating sensorial 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 research studies present convincing data that such concerns have an impact on people's acceptability and adoption of these new technologies.

So, the question becomes: 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, like 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.

It is hard to define privacy in a way that is broadly accepted. As new technological advances open up new ways of how privacy can be affected, we need to continuously reassess our understanding of privacy and how it should be protected. Furthermore, people's concerns on their privacy vary widely. While many people wish to have control of who has access to information about themselves, differences arise about what kind of information they want to control. We also need to consider the ubiquitous nature of the communication between users and service providers, which introduces more privacy threats at the network level. Some steps to design privacy preserving solutions for both of these views (data and communication) exist and we discuss here their technical overlaps and boundaries in the context of urban sensing. But privacy is a conundrum for experts also for another reason: its quest contradicts the requirements of security. We look at this problem in the last section and discuss security solutions that allow for maintaining privacy guarantees.

II. Privacy Threats

Using a cellphone for collecting information from the environment and tagging them with time and GPS data, unavoidably reveals 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, 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. Of course, if the sensed information itself implicates sensitive personal data, like for example health data, the privacy problem becomes more obvious and we need to use techniques that protect privacy during the aggregation process of such data [2]. However, here we consider the case where mobile phones collect information from the environment, e.g. noise, pollution, etc. and we concentrate on location privacy.

Technically, location privacy can be provided in pervasive computing by assuring that the attacker cannot associate two or more of the following pieces of information: who, where and when [3]. So, the first step is to investigate which identifying information can be collected in the urban sensing paradigm and where and how it can be combined. There are two significant factors that hinder our efforts towards this direction: the diversity of urban sensing applications and the diversity of technologies used.

A. The diversity of urban sensing projects

Currently scientists experiment with the new possibilities opened by the wireless Internet and the sensing possibilities of mobile phones. As a consequence, today's suggested urban sensing projects aren't restricted to a few fixed services, but rather appear as a broad set of different, dynamic, and feature-rich services that are both exciting and helpful to citizens. With respect to privacy, we identify the following three important dimensions that we need to consider (see also Figure 1).

1) User focus: Urban sensing projects could range from community-based to very personal and self-reflective ones. Currently, most of them target the first category, where users don't 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. However, there is also an increasing number of 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.

In the first category of projects, the message sent by the mobile device could simply consist of the triplet (*location*, *time*, *data*) and be decoupled by the identity of its custodian. In this case the privacy of the sender is better protected, but not entirely. By having access to location traces, an attacker can associate place and time and analyze the corresponding movement patterns, which could reveal the identity of the sender, or significantly reduce the possible values (anonymity set). This is easy to realize since people usually spend most of their time at specific places, like their home or work.

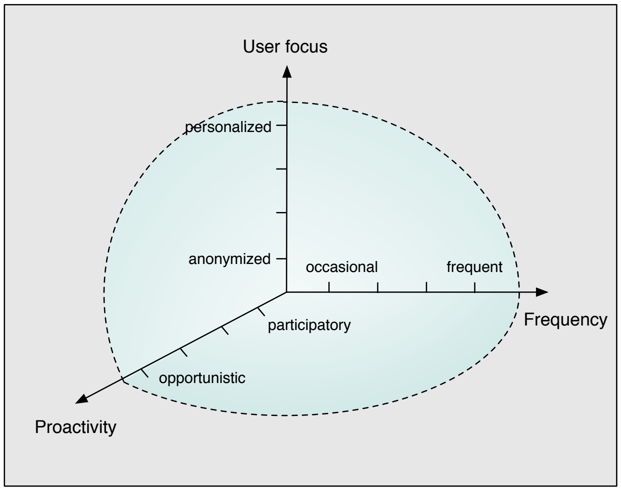

Fig. 1. The three dimensions that affect the privacy threats in urban sensing applications. As we move away from the center of the depicted sphere, weaker attackers can compromise the privacy of the participants. Towards the center, privacy is not guaranteed, but we need to assume stronger attacker models to break it.

2) Frequency: In some applications, only data at discrete points of space and time are of interest. This data, tagged with GPS and time is uploaded occasionally at the service provider, where it is aggregated with similar data uploaded by other users and represented in some visual form. Popular applications in this category are those which build a noise, traffic or air pollution map of the city.

However, some applications require continuous input of a person's movements, instead of occasional readings. Examples in this category include sharing bike routes between cyclists, where trajectories are combined with environmental data, or monitoring a person's whereabouts to estimate individual carbon footprint and exposure to air pollution.

Assuming that the application allows anonymous measurements, we need to stress here that the uploading frequency is an important factor for the privacy of the sender. An adversary could easily link anonymous independent updates to the same user, if the sample frequency is sufficiently high compared to the user density in an area. Efficient methods, like trajectory-based linking, could lead to an accumulation of path information about individual users [4]. Then extracting the endpoints of the highly probable home and work locations could result in the identification of a unique individual.

3) Proactivity: It is important to distinguish between participatory and opportunistic sensing. With participatory sensing the custodian is actively involved in the data collection process. 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. On the other hand, opportunistic sensing takes a more proactive approach. Sensor sampling occurs whenever the state of the device (e.g. geographic location, available sensors, etc.) matches the application's requirements. In this case the device is remotely tasked to collect and report sensor data, utilizing in this way the device without its custodian being actually aware of the sensing activity.

In opportunistic sensing the system needs to know which users are most likely to visit a particular location, in order to task those devices directly [5]. For example, we may assign the task "measure temperature in area X" to Alice, when she is about to enter this area. This requires knowing the current location of Alice which will also reveal the time when Alice visited area X. Thus, in this case, ensuring that users can be tasked anonymously is a harder problem to solve.

We have concentrated so far only on the nature of the sensed data. In urban sensing projects, users don't have to report this data from the location, where they actually took the measurement. That for example could depend on the availability of wireless Internet connectivity. Then, the user can be localized by identifying the wireless access point through which the network connection was made. So, we must treat separately the problem of protecting user's location from the system infrastructure and, as we will see in the next section, this creates even more risks for the user's privacy.

B. The diversity of technologies

Figure 2 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.

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*. Therefore, it is important to see what kind of identification information is revealed at each step.

1) Identification during network access: In GSM, users authenticate to the mobile operator using secret information stored in their subscriber identity module (SIM), a smart card within the GSM phone. During this process, the user discloses the unique identifier of his mobile phone (IMEI) and of his user account (IMSI). Independently of the authentication scheme, these identifiers can be used to link individual network accesses with each other.

Access through WLAN potentially needs less individualizing information compared to GSM access. However, during network access, the mobile user must disclose the hardware address of the network interface card (MAC address). This address encodes the manufacturer, type and serial number of the network card, a globally unique value that can be used to link individual accesses to the network. In contrast to IMSI and IMEI, it is often possible to modify the MAC address of the network interface by software.

2) Location data: During an active GSM connection, the mobile operator knows the wireless broadcast cell in which the user is located. A cell is usually defined by the antenna (base station) through which the user is communicating. In urban regions cells can have a diameter of around 100 meters so knowledge about the cell can be used as a source of location data. There is a comparable situation if WLAN is used as network access, because the mobile device connects to the access point with the best signal strength.

3) Transport level identification: The common protocol suite through which the network layer operates is TCP/IP. At this level the IP address is used as identifier of a communication endpoint. Usually, the IP address is assigned *dynamically* to a mobile device, i.e., the device will have different IP addresses during different network access sessions. In GSM networks, however, the IP address is chosen from a pool of addresses that belong to a particular mobile operator. Therefore, the IP address can be used to identify the network operator, which in turn implies knowledge of a certain location.

In WLAN, the dynamic IP address assigned to the user is often a *local address* which is hidden behind the access point. Through the technique of network address translation (NAT) communication initiated by the user will use the IP address of the access point when it is routed into the Internet. The translation to local IP address can only be performed by the access point itself. The reason for both NAT and dynamic IP addresses was the shortage of IP addresses in the current version of the IP protocol (version 4). This will change in the future once version 6 of IP is introduced. Roughly speaking, there will only be static IP addresses in IP version 6.

III. Privacy Protection Strategies

There are currently two general approaches in the research literature for providing privacy: one based on legislation and one based on technology. We refer to these approaches as privacy-by-policy and privacy-by-architecture, respectively [6]. This characterization is based on the diversity of the threat models discussed in Section II and the privacy expectations and concerns of the users themselves.

A. Privacy by Policy

Privacy policies are trust-based mechanisms that aim to protect location information and any other collected personal data from accidental disclosure or misuse. In the U.S. and Europe, these privacy policies are influenced by the Fair Information Practices (FIPs), originally codified in the 1970s. A subset of them was later tailored by the Federal Trade Commission to e-commerce, emphasizing to properties such as notice, choice, access and security. They aim to inform the individuals about the data collected, offer them choices as to whether they wish to share this data for other purposes, give them access to their data so that they can review or delete information and finally, protect the security of the information.

Even though these codes are still considered a gold standard for privacy protection, they pose two main limitations. The first one is the assumption that corporations can be trusted

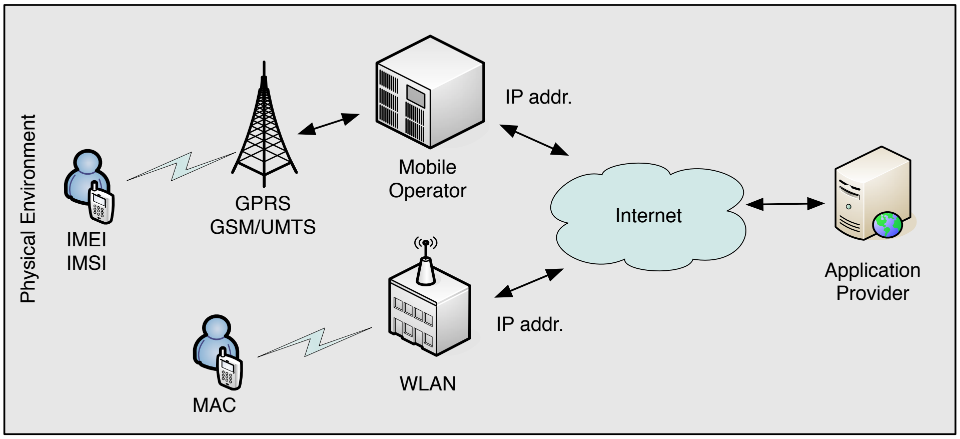

Fig. 2. In urban sensing a single infrastructure integrates heterogeneous technologies such as wired, wireless and cellular networks.

to handle user's personal information and that policies and regulations are generally enforceable. Privacy policies are ultimately vulnerable to disclosure of personal information, accidental or malicious. Second, not all people have the same privacy preferences. In location-based systems these preferences vary with place, social context and even culture, assigning to privacy a specific, variable, and highly individual meaning.

Against the second problem, the latest research directions on providing privacy for urban sensing systems attempt to engage participants themselves to answer privacy dilemmas [7]. In pervasive technologies like urban sensing, people are implicated as the primary participants of the data collection process. It has been shown that the way people choose to withhold or disclose information about them depends highly on their context, e.g. identity, situation, time, or culture. Therefore, urban sensing projects should give them the possibility to negotiate sharing and discretion according to the context and their preferences. Towards this direction researchers identify the need of new tools, including: techniques to allow participants to mask location data; local processing on phones to avoid upload of sensitive information; and unique interfaces that encourage selective sharing.

However, in order to weight the costs and benefits of sharing information and make informed, context-specific decisions, people should have a very good understanding of what is being sensed or transmitted and be well informed of the privacy risks. Otherwise, these technologies could raise potential privacy and security conflicts.

B. Privacy by Architecture

As we discussed in the previous section, privacy by policy cannot protect from stronger attackers, who would not be deterred by policies and regulations. A consensus has not been reached in the privacy research community on how realistic these stronger attacker models are. Cryptography researchers and privacy rights organizations tend to agree that we should prevent access to location information at all costs, making it tamper-proof against both i) malicious hackers with the desire to intrude on other people's privacy, and ii) against companies profiling and accumulating users' location information for profit maximization.

Privacy-by-architecture works under such stronger attacker models and covers both of the above concerns. The basic goal here is to actively design for non-identifiability of users and provide stronger privacy guarantees, in the sense that even if an attacker has access to the necessary information, no personally identifiable data can be created or recreated with reasonable effort. In general, to achieve this goal some degree of noise needs to be introduced into the data set and thereby distort its contents and usefulness.

1) Anonymity-based techniques: This class includes all solutions based on the notion of anonymity, which is aimed at making an individual (i.e., her identity or personal information) not identifiable. Early solutions suggested the use of static pseudonymous IDs, but soon it was realized that it might be trivial to infer the true identity behind each pseudonym, by linking all user entries together. Therefore, pseudonyms must frequently change, in order to decouple identity from location-time information. In general, methods in this class do not guarantee that the process of linking a pseudonym to an individual is impossible, but that it requires a large effort.

2) Obfuscation-based techniques: As we mentioned in Section II-A2, even when identifying information is removed from the reports, associating pairs of time and location data might still prove rich source of information for inference about a user's location and activity. To make it harder to link reports back to the same user, an approach is to obfuscate location and time information, lowering their precision or accuracy and adding enough "confusion" in the data. As an example, consider AnonySense [8], a privacy-preserving architecture for realizing participatory sensing applications. AnonySense uses the concept of *tessellation* for protecting the location privacy of contributing users.

In tessellation, a point coordinate is generalized to a plane in space, which is referred to as a tile. The sensor reports uploaded by users contain the tile ID and a time interval ID, rather than the absolute location and time. This generalization is guided by the principle of k-anonymity, which ensures that at least k users are located in the same tile within a time interval. Hence, it is difficult for an adversary to distinguish between the k users, based on the location or timestamp within the reports. The problem with this solution is that it requires the presence of sufficiently large number of active users, or else the tiles must be made impractically big.

IV. Network Level Anonymity

As we saw in Section II, protecting the privacy of the user demands not only solutions at the data layer, but also at the network layer. Often, techniques for achieving anonymity on the network and data level are combined, as there is no real anonymity on the data level without anonymity on the network level. 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 has not been addressed extensively. So, here we consider only 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 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.

Performance plays a much more important role here 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 even higher latency. Expensive cryptographic operations on the mobile phone also contribute in the degradation of the performance. The resulted latency significantly affects the user experience, and users are known to be impatient and willing to wait only a short time, especially in scenarios where they don't get a direct benefit.

Most of the urban sensing projects could tolerate some latency, when it comes to *the delivery of the message*, at least compared to browsing the Internet. For example, Anonysense [8], the only urban sensing project so far that employs an anonymizing network, integrates Mixmaster. Mixmaster is the primary anonymizing network for sender anonymity in e-mail messaging and belongs to the high-latency approaches. These approaches seek to provide a strong degree of anonymity at a possibly increased delay. For example, instead of flushing all messages at each iteration, Mixmaster keeps a subset of messages in the proxy until next round, meaning that messages may be delayed for hours or even days. This is clearly too much for some urban sensing applications, like those that build real-time maps of noise and air pollution in the city, but maybe acceptable by others, like sharing bike routes or photos.

V. Privacy-Preserving Security Protocols

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. Authentication gives users and service providers assurance that no intermediate devices have tampered with the data and that they are indeed interacting with the intended parties, and not some malicious entities.

There are many approaches to privacy, as highlighted in Section III, however, most of the times these approaches lead 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, preferences and actions. So, a question arises: If privacy is to be preserved through user anonymity, how can a service provider be convinced that an anonymous user is trustworthy?

In response to this, 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. In what follows we review some of these techniques and briefly discuss their merits in achieving privacy-preserving authentication in people centric applications.

Blind signatures schemes are just like ordinary signatures in which the contents of the message are not revealed to the signer of the message. Typically, to produce such a signature on a message m, the user first *blinds* the message by combining it with a random quantity r and then forwards the blinded message m to the signer. Once the message is signed, the user proceeds to remove the blinding factor, thus obtaining a signature on the original message. Blind signatures can be used in the urban-sensing scenario as a means to provide for *authentication tokens* by which a user can hide their identity and obtain access to a particular service. Care has to be taken, however, so that a malicious user cannot effectively mount a chosen message attack by obtaining signatures on arbitrary messages or simply re-use the tokens.

Another approach for enhancing anonymous authentication is to use *group signatures* (and the simpler *ring signatures*), 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. Thus the owner of the system can tell that a group member created the message, but not exactly which member. One problem in the case of group signatures arises from the fact that anonymity can be *revoked* by an authority, called the group manager, who, in case of

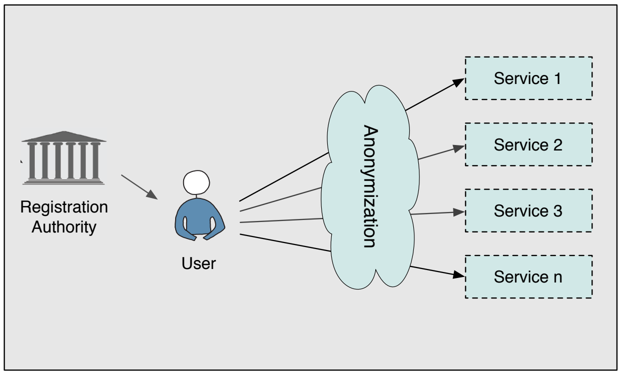

Fig. 3. A single registration allows the user to enjoy secure access to different services with enhanced privacy.

disputes or unauthorized access, can identify the user – this last concept is not provided by ring signatures. This may act as a deterrent for malicious user behavior, but on the other hand, the revocation capability can be used by malicious group managers and service owners to track the actions of legitimate users as well.

Now we will look at a different paradigm in which users can enjoy maximum privacy, provided they use a particular service a predefined number of times. This concept is called *anonymous k-times authentication* [9]. This scheme has two distinct features that make it very attractive to the urban setting model. First, it ensures that no one, not even an authority, can identify a user who has not exceeded the allowable number of authentication attempts. Second, it allows anyone to trace, without help from the authority, dishonest users who are overusing a particular service. Thus this scheme is more preferable than the identity escrow/group signature schemes, in which the authorities have the ability to trace users. Additionally, the scheme is not confined to one particular service but can be applied to multiple services, provided the access threshold has not been exceeded, as shown in Figure 3.

On the negative side, this scheme is more complicated since the role of the group manager is separated into two roles; one to be used during the *signup phase* and one to be used during the *tracking phase*. However, each role can be distributed among multiple entities, guaranteeing user privacy even in the case of colluding authorities. An added benefit of this separation of powers is that it provides for *accountability* in addition to achieving privacy and security: The users true identity will be revealed when the user is overusing a service, i.e. for more than k-times.

VI. Conclusions

Is sensing data with always-on mobile phones a new opportunity to promote science and community research or is it a new embedded surveillance tool that we carry around in our everyday life? The answer to this question depends on who has control of the collected data and what privacy assurances users are given. As we start exploring this new paradigm of urban sensing, we realize that the diversity of the applications and the implicated communication technologies define multiple threats and open new front-ends that we need to defend. What urgently needs to be done is to build privacy solutions while this new technology is still in its infancy and we have the opportunity to ensure that the dangers will be averted.

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] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, “PriSense: Privacy-preserving data aggregation in people-centric urban sensing systems,” in *Proceedings of the 30th IEEE International Conference on Computer Communications (InfoCom ’10)*, San Diego, CA, March 2010.

[3] M. Langheinrich, “Privacy in uniquitous computing,” in *Ubiquitous Computing*, J. Krumm, Ed. Chapman & Hall, CRC Press, 2009.

[4] M. Gruteser and B. Hoh, “On the anonymity of periodic location samples,” in *Proceeding of the 2nd International Conference on Security in Pervasive Computing*, Boppard, Germany, April 2005, pp. 179–192.

[5] A. Kapadia, D. Kotz, and N. Triandopoulos, “Opportunistic sensing: Security challenges for the new paradigm,” in *Proceedings of the First International Conference on Communication Systems and Netwprls (COMSNETS ’09)*, Bangalore, India, January 2009, pp. 1–10.

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

[7] 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.

[8] 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.

[9] 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)*, December 2004, pp. 308–322.