# I'm Here! Privacy Challenges in Mobile Location Sharing

Marcello Paolo Scipioni
Faculty of Informatics
University of Lugano (USI)
6904 Lugano, Switzerland
marcello.paolo.scipioni@usi.ch

Marc Langheinrich
Faculty of Informatics
University of Lugano (USI)
6904 Lugano, Switzerland
marc.langheinrich@usi.ch

## ABSTRACT

With the availability of GPS receivers in more and more mobile handsets, location sharing has become the next mobile killer-app. Services like Google Latitude or Yahoo's FireEagle are used by thousands of subscribers to share their current location in real-time with their family, their friends, or even publish it online, while Sense Network's CitySense application uses live movement traces from entire metropolitan areas to show users the "most popular places" in the city. Even though research in location privacy has proposed a range of solutions for safeguarding location information, it is unclear how existing proposals can be applied to such applications. In this paper, we set forth a categorization of location sharing applications, outline the various privacy challenges each category poses, and discuss the shortcomings of existing solutions.

## Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

## General Terms

Location Sharing, Privacy, Social Networking

## 1. INTRODUCTION

The rapid proliferation of social networking services (SNS) witnessed in the last few years seems to show no signs of slowing down. The world's largest SNS, Facebook, has more than doubled its user base in the last year – from 150 millions in January 2009[1] to 350 millions in January 2010[2] alone. It comes thus as no surprise that the growing numbers of powerful smartphones (e.g., the iPhone, the Palm Pré, or Android-based handsets) increasingly feature tightly integrated SNS client software, such as the Blackberry Facebook client or the iPhone Facebook application. MySpace and RIM, the Blackberry manufacturer, have recently announced plans to form an alliance and offer integrated mobile social networking experience. Several models already allow users to integrate their contacts from multiple SNSs into a single phone book, linking not only their numbers, pictures, and birthdates, but also displaying people's status updates almost in real-time.

This not only makes *receiving* updates from others easier, but also significantly simplifies the *dissemination* of one's own information into one's social network. Twitter has long been a truly mobile application, centering around sending and receiving SMS-based short text messages, yet modern SNS client software can not only use cheaper WiFi or 3G transmissions, but also publish a status update simultaneously across a number of SNS. In fact, the abundance of sensors in today's mobile phones actually facilitates *fully automated* status updates, where the phone autonomously publishes updates on the activities of its owner to her social network, her Twitter "followers", or her publicly accessible blog. An example of such an application is "CenceMe," which uses the iPhone's microphone, GPS, and acceleration sensor to infer the owner's "social setting" and publishes status updates like "in a meeting", "running", or "dancing" to her Facebook, mySpace, or Twitter accounts.

Slightly simpler but even more popular are applications such as Google Latitude, which continuously update the owner's location – either to other Google Latitude users or directly on the Web. Location, while only one of the many sensor readings of a modern smartphone, is probably the most powerful and versatile attribute to share. Potential applications of location sharing are, to name but a few:

- turn-by-turn navigation;
- geotagged picture upload with integrated landmark identification;
- friend-finder applications that alert one to near-by family members or friends;
- recommender systems that suggest near-by places of interest.

These opportunities for sharing one's location quickly bring privacy issues to mind. Location privacy has been a hot-button topic for quite some time now, and a large number of solutions have been proposed that seemingly "solve" the issue, e.g., by increasing the anonymity of location-based queries (k-anonymity), or by impeding de-anonymization attacks on location traces (obfuscation). Often enough, however, these solutions are presented without any concrete reference to applications or services where these techniques could be used. For example, a service that would allow one to call a taxi to one's current position would obviously not benefit from having the location obfuscated, while a tracking a person "protected" by, say, a k-anonymity value of 100,

[1]See `blog.facebook.com/blog.php?post=46881667130`
[2]See `www.facebook.com/press/info.php?statistics`

would result in an accuracy of a few dozen meters when done during rush hour in Tokyo. Clearly, location privacy mechanisms need to be discussed *in situ*, i.e., with a specific application as a point of reference (sic).

In this paper, we argue for taking a holistic view towards providing location privacy in location sharing systems. It is easy to imagine that today's popularity of SNS and powerful smartphones will soon lead to widespread "automated microblogging", where our phones will churn out a steady stream of "I am here!" and other status updates, and publish these on Google maps, on Facebook walls, and as Twitter feeds. Resorting to simply anonymizing or degrading the quality of these location feeds will hardly solve this issue, as this would directly impact the usefulness of these services. In order for people to value privacy-compliant solutions, these must be compatible with the original purpose of their location sharing, namely to stay in touch with friends and family, and to discover new places and interesting people. In the following, we will present a classification of location sharing applications and their goals, and outline the individual challenges for location privacy in each domain.

## 2. MOBILE SOCIAL NETWORKING

Almost all existing social networking applications have recently included location based features, and many new applications were born specifically as location-based services. In early 2009 Google released *Google Latitude*[3], a location-based mobile application to share one's current position in real-time. Latitude can be integrated with other Google tools available from the iGoogle accounts, allowing for instance to display the position of friends on Google Maps[4] or to include one's own location on a personal homepage. Google Latitude was not the first friend-finder application, nor is it necessarily the most popular. Other mobile friend-finder applications are Loopt, Foresquare, Gowalla, Brightkite, Whrrl, Buzzd, and Yelp – to name but a few.[5] Friends can locate each other, annotate real-world places with user-contributed content (reviews, photos, twitter-feeds), find new friends based on matching location profiles and interests, or even play location-based games.

Apart from directly sharing one's location in order to find friends or restaurants close-by, users might also be encouraged to disclose their location traces in order to act as human "activity sensors". Companies like the MIT spin-off "SenseNetworks" use such user-submitted location data for predictive analysis, e.g., to identify "hot spots" in a city or to cluster users into marketing segments according to their movements.[6] Another interesting project which employs location data is *Mobile Millennium*[7] from the University of Berkeley. This project was conducted to develop a system for the on-line monitoring of traffic conditions by collecting data from GPS-enabled smart-phones. Also in this case users act as sensors, allowing to build a traffic congestion model without the need of a sensor network external to the user community itself.

While the use of location based services clearly has the potential to accidentally share a particular piece of information with someone on an incident by incident basis, it is the automated inspection and categorization of movement data that poses the biggest threat to personal privacy. Early work in location profile inspection by Kang et al. [10] has demonstrated how clustering of location traces can identify places that the user would likely find important. Beresford and Stajano [2] showed how even anonymous traces can yield the identity of the user when being combined with profile information, such as the user's office number. Hoh et al. [9] simulated attacks on anonymous GPS data from 239 drivers and were able to find 85% plausible home locations on a subset of 65 drivers. In a similar experiment, Krumm [11] used anonymous data from 172 drivers and correlated the plausible home address with actual white pages data, which allowed him to find the actual home address in 13% of all cases and the actual names in 5%. Such attacks are called restricted space identification (RSI) attacks [7], as they allow an attacker to map an anonymous user's favourite place (e.g., the home or office address) to an identity using publicly available information.

Privacy is an issue to be taken into account not only because of security reasons. If people start to share, it does not necessarily mean that people want to share everything with everybody. As Dourish and Anderson [4] explain: "Privacy is not simply a way that information is managed but how social relationships are managed". To this end, a study by Consolvo et al. [3] was conducted explicitly asking people to disclose their location data. It turned out that the most important factors that people take into consideration before disclosing their location data are: *who* is requesting location data, *why* he needs it and at *what level* of detail. The implication of privacy with social relationships has therefore to be carefully considered, and for this reason location privacy strategies should take it into account.

## 3. LOCATION PRIVACY

The easiest way to apparently "solve" location privacy is to manually or automatically authorize (or not) the disclosure of location information to others. As a result, sharing location data would lead to a binary decision: friends would receive full access to location data, and strangers would be blocked. This is in fact the most basic protection that services like Google Latitude offer: one can easily share and "unshare" one's location information with individual users. This approach, however, quickly becomes limiting, as it forces people to choose between "on" and "off", between "black" and "white", without considering all those "grey" levels that a dynamic privacy negotiation process usually involves [4]. Consequently, more complex and/or powerful methods have been proposed.

### 3.1 Basic Issues

Privacy can be described as "the desire of people to choose freely under what circumstances and to what extent they will expose themselves, their attitudes and their behaviour to others" [14]. Safeguarding location information is just one of the many "data points" that make up the attitude

[3]See `www.google.com/latitude`
[4]See `maps.google.com`
[5]See their corresponding websites at `loopt.com`, `foursquare.com`, `gowalla.com`, `brightkite.com`, `whrrl.com`, `buzzd.com`, and `yelp.com`, respectively.
[6]See `www.sensenetworks.com` and `www.citysense.com`
[7]See `traffic.berkeley.edu`

and behavior of people, yet it is a particularly powerful one, as a place is often tightly connected to an activity (e.g., a shopping mall, an office), an interest/belief (e.g., a church, a political rally), or a personal attribute (e.g., a prison, a clinic).

As with any privacy assessment, one of the most basic question is: what "privacy-risks" should the system safeguard against [12]? In location-based service scenarios, one can differentiate between three primary actors:

- *Intended recipient*, e.g., the taxi company to send you that cab, or your mom so that she knows where you are. This usually involves the use of a *service provider* that offers to forward your location to the intended recipient.
- *Service provider*, e.g., Google providing you with the *Latitude* application, or a restaurant recommendation system for near-by places. In contrast to the *intended recipient*, users usually do not have a primary goal of letting the service provider know their location – it is a by-product of getting a restaurant review or staying in touch with friends.
- *Infrastructure provider*, e.g., your mobile phone company, or the operator of an indoor location system. While self-positioning systems such as GPS can work without an infrastructure provider, mobile phone users are often implicitly located in order to provide communication services (i.e., route phone calls).

In addition, a location-based system might have a number of unintended recipients, such as:

- *Accidental recipient*, e.g., your parents when you claim to be at a friend's place studying, but forgot to turn off the system when you went out drinking.
- *Illegal recipient*, e.g., a hacker intercepting your wireless location updates, or breaking into the service provider's or the infrastructure provider's records.
- *Law enforcement*, e.g., police or other government agencies accessing the service provider's records.

Clearly, for a given system, the *infrastructure provider* needs to be trusted, as location information is not available otherwise. While GPS alleviates the need for a provider, mobile phones are always traceable through their connection to a particular antenna. Obviously, withholding location information from the *intended recipient* seems of not very useful. Both *illegal recipients* and *law enforcement* are best controlled by leaving as few information as possible stored on external servers – laws that limit data retention periods might help here, as would a service architecture that would minimize data collection on those servers in the first place. *Accidental recipients* will require appropriate control tools (manual and/or automated) that can support users in minimizing such accidental disclosures. Today's *(location-based-) service providers* often offer the service for free in exchange for tracking the user's location – a technological solution withholding location data from them would need to be complemented with an appropriate alternative revenue model.

## 3.2 Existing Proposals

One of the most popular methods for location privacy is *obfuscation*. Duckham and Kulik [5], which offered a first formalization of this approach, define obfuscation as "the means of deliberately degrading the quality of information about an individual's location in order to protect that individual's location privacy." Obfuscating location information lowers its precision, e.g., showing only street or city level location instead of the actual coordinates, so that the real location remains hidden. To take Google Latitude again as an example: it allows users to manually set each recipient to only receive a city level location update. Krumm [11] shows that location can also be obfuscated by adding random noise to the actual coordinates. However, he also demonstrates how much added noise (a lot!) would be needed to significantly reduce the chances for an attack.

Another popular approach in the area of location privacy is "k-anonymity", introduced by Gruteser and Grunwald [6]. The actual location is substituted by a region containing at least $k-1$ other users, thus ensuring that a particular request can only be attributed to "1 out of k" people. This approach has the disadvantage that if the region contains too few people, it has to be enlarged until it contains the right number of people, while it will inadvertently shrink if many people are in close proximity (e.g., in a subway). This makes it hard for a user to predict the amount of obfuscation added to her true location.

It is also possible to extend the simple binary disclosure approach mentioned above to support rule-based mechanisms. Myles, Friday, and Davies [13] used time, location, and query originator information (among others) in their rule-based system to decide whether or not to release location information. Benisch et al. [1] found that the more of these parameters a rule-based location privacy system offers, the fewer mismatches between the privacy preferences of the user and the automated decisions of the system would arise.

Beresford and Stajano [2] propose the use of so-called *mix-zones* – areas in which no location tracking takes places and which are sized so that at any point in time, a large enough number of targets are present that can be "mixed." In a similar fashion, Hoh and Gruteser [8] have proposed "path-perturbation" algorithms that attempt to use the crossing of two or more location tracks in order to increase the chances that an attacker confuses the path of different users. Such approaches allow the frequent updating of pseudonyms in order to prevent a single association between a user and a pseudonym to "unravel" the entire pseudonymized stored location tracks of that user (called an *observation identification* (OI) attack [7]). Without these extra mechanisms, pseudonyms could not be changed, as simply changing a pseudonym at random is trivial to detect in a continuous location track.

## 3.3 Limitations

While the above methods work well in theory, it is often not clear how to apply them in practice, and in particular in the

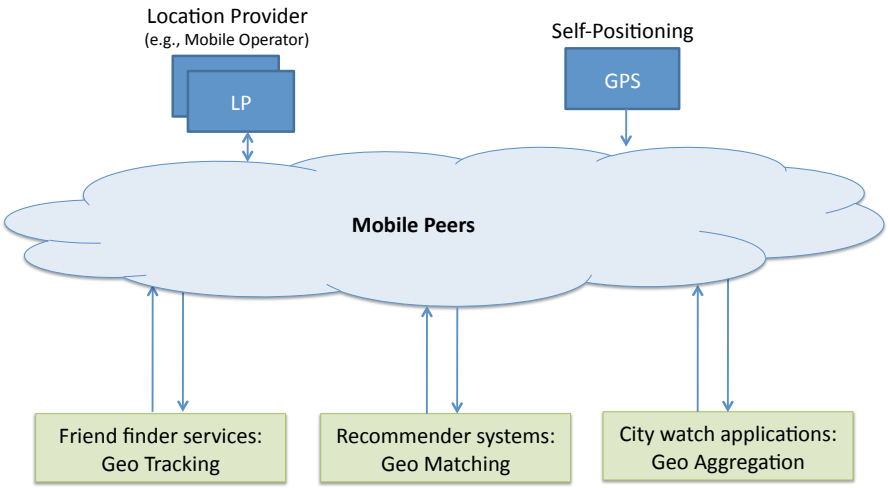



**Figure 1: Three sharing patterns for location-based social networking applications.**

context of location sharing applications.

For example, a rule-based mechanism might work reasonably well against accidental disclosure, e.g., by putting automated temporal and spatial restrictions on location updates. However, rule-based mechanisms do not address privacy risks viz. a service provider, theft, or law enforcement access. Similarly, a rule-based mechanism might affect the service quality for *intended* recipients as well, due to overly restrictive policies. At the same time, misconfigurations might still allow for accidental disclosures to happen, especially if service configuration is error-prone.

Obfuscation solutions per definition lower the quality of service of a location-based application, and would thus negatively impact application use. Moreover, choosing the correct obfuscation level is highly context dependent, so users would need to continuously evaluate the level of precision at which they would like their current location be shared with a particular recipient. Obfuscation offers only limited protection against the risks from unwanted location disclosures to law enforcement, theft, infrastructure providers, and service providers.

The concept of k-anonymity also offers limited protection only. The biggest confusion surrounding k-anonymity is the fact that it cannot actually protect *location* information, but only protects *identity* information, albeit in a location-oriented context [12]. Given anonymous queries, k-anonymity prevents an attacker from performing an observation identification (OI) attack [2], which uses observations of a particular location to tie an anonymous query from that place to the identity of the observed person in this place. With k-anonymity, the origin of the anonymous query always contains $k$ other people whom this query could have originated with. Also, as k-anonymity simply adds obfuscation, the same issues apply as outlined above, though exacerbated by the fact that the amount of obfuscation added is variable.

Path perturbation and mix zone approaches work well with stored location tracks and can thus provide protection from unwanted access to stored location data by hackers and law enforcement. However, the actual implementation of such systems would need to happen at the infrastructure or service level, thus rendering these methods ineffective for preventing unwanted disclosures to service providers and/or infrastructure providers.

# 4. CHALLENGES

As outlined above, using existing location privacy solutions in mobile social networking scenarios is difficult. While different aspects of individual solutions might work for various issue, no single approach offers a comprehensive solution. If we want to provide users with privacy-aware alternatives to today's popular location sharing applications, we will need to carefully analyze the needs and requirements of such systems and design privacy solutions that address the identified risks without affecting service use.

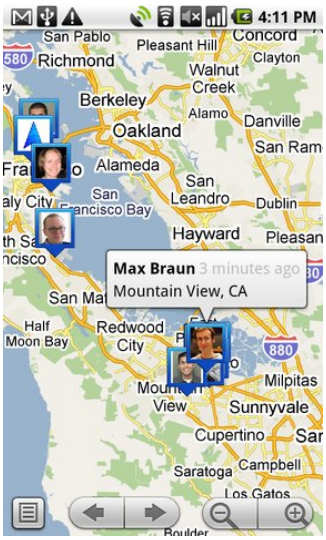
If we analyse the possible forms of location sharing, also observing the applications already available, as described in Section 2, three different location sharing methods can be identified:

- Friend finder services;
- Recommender systems;
- City watch applications.

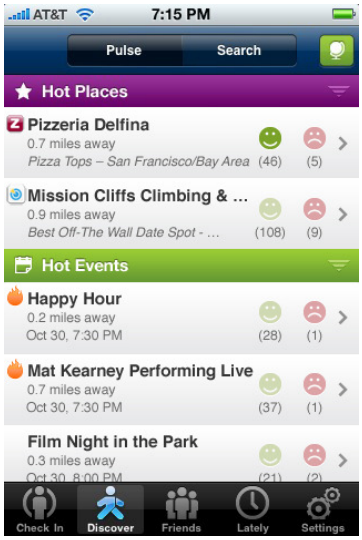
Figure 1 illustrates how these three patterns represent the applications that offer location sharing services to mobile users. Location information itself may come from a trusted location provider (LP), e.g., a mobile phone operator, or by using self-positioning technology such as GPS. Each of the three cases presented above has its own characteristics, which are analysed below in more detail.

## 4.1 Friend Finder Services

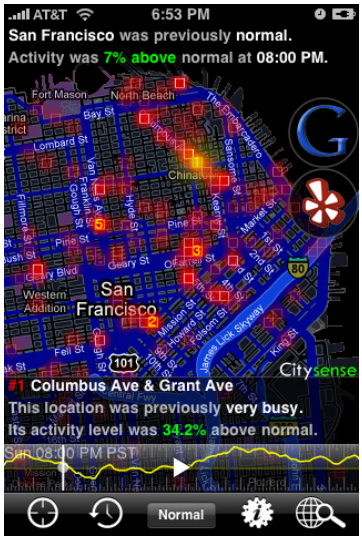
Friend finder services follow a "1:1 pattern", i.e., a single user directly shares his or her location with another user. Friend




(a) Friends shown on the map of the Google Latitude friend finder application.




(b) Restaurant recommendations from the Loopt recommendation service.




(c) Hot-spots inside a city, from Citysense.

**Figure 2: Screenshots of applications belonging respectively to the friend finder, recommender system and city watch categories.**

finder services allow users to locate their friends, e.g., to visualize them on a map, or to receive alerts when they are near to each other. The map-based visualization of Google Latitude is shown as an example in figure 2 a).

Clearly, users will need adequate controls to manage their location sharing behavior. While rule-based systems offer high expressiveness, they require high maintenance and possible expert users. One option would be to tie location sharing behavior to other channels of communication, and to reduce the frequency and accuracy of location sharing in tune with two people growing "out-of-touch". This would introduce a natural decay into such sharing systems, which could only be countered by more frequent communication (e.g., via SMS, email, or physical encounters).

However, while 1:1 sharing systems clearly run the risk of publishing location information to an *accidental recipient*, we believe that such risks are small by comparison. The real danger of such systems stems from the comprehensive recording of location information by *service providers*: every time a location update is shared, the service provider gets an update and is thus able to create detailed behavioral profiles of its customers. Ideally, a privacy-aware 1:1 location sharing system would be able to share location information even *without* a central service provider receiving a copy of the entire movement track. One option would be a complete decentralized, peer-to-peer system, albeit at the expense of higher communication costs. Alternatively, one could imagine making use of a central service provider, yet using encrypted location updates that would only allow the provider to *detect* when two people are in each other's vicinity, but without knowing where this is. Obviously, this would entail a different business model, potentially a paid service or ad-financed. Also note that simply lowering the number of updates sent to the service provider would not avoid its ability to build movement tracks of a user: even a once-every-hour or once-per-day location update would allow the creation of privacy invasive location profiles.

## 4.2 Recommender Services

Recommender systems follow a "1:n" pattern, i.e., the movements of a single user are shared with an unknown group of other users, who do not have to have a prior relationship with the originating user. The goal of a location-based recommender system is to suggest new shops, restaurants, or events based on the user's preferences and tastes, as manifested by the user's movements. Figure 2 b) shows a screenshot of Loopt, recommending the best nearby places and events. The recommendation themselves would be equally collected from the movements of other users, just like an online bookstore is able to make recommendations such as "people who buy the books you buy, also buy the following books..." Recommendations could not only include places and events, but also people with similar interests.

A location-based recommender system has thus has to match a user's individual movement history with traces from other users, find overlaps, and identify from these overlaps new places (i.e., stores, events) that the user should explore. In its simplest incarnation, a location-based recommender system would simply collect the spatio-temporal records of all its users in order to find pairwise overlaps between users. As outlines above, simply pseudonymizing those tracks would not prevent thieves or law enforcement to later identify an individual and his/her movements from such traces [9], nor would this prevent the service provider itself from creating movement profiles for other purposes (e.g., advertising).

A more challenging way to perform such a service would be to avoid the tracking of the users by the service provider, performing the matching – i.e. finding commonalities among

users – without having access to the full location tracks of its users. Cryptography could again be used to facilitate matching seemingly meaningless properties that would only make sense to the individual users, who would be the only one able to translate the information back. Alternatively, matching could be done in a split-fashion with the help of personal devices, which would extract some kind of aggregate "measure" and have the service perform matching on such abstract distances.

### 4.3 City Watch Applications

Location-based sharing can also be used to monitor the behavior of an entire city – in our framework, this could be called an "n:n" location sharing pattern. Instead of individual recommendations, such a system would simply collect location data from many people – an entire city, ideally – and identify macro- and micro-trends of behavior. The purpose of such a system is to aggregate data for analysing behaviours of large groups of people, e.g., to monitor traffic in order to locate or even predict traffic jams, or to find the "hot spots" inside a city on a Saturday evening, in order to give people an idea where to find the most popular clubs, restaurants, or events. This is exemplified in figure 2 c), where a screenshot of Citysense shows the hot spots in the city of San Francisco.

In such systems, no individualization is needed, yet a simple approach would still collect location tracks from all users in order to detect not only aggregate numbers but also spatio-temporal causality (e.g., "after *that* restaurants people go to *this* club"). While mix zones and path perturbation approaches seem highly applicable, the challenge is to find useful path lengths that still offer useful causality without running the risk of allowing RSI or OI attacks (see section 3 above). Another option again would be the use of a split-solution, in which user-owned devices would perform intermediate computations, potentially in a peer-to-peer fashion.

## 5. CONCLUSIONS

Location information – especially when collected over extended periods of time – is data that allows for significant inferences over people's lives. Existing online social networking platforms (SNSs) have already demonstrated the risks associated with the increased publication of personal information, yet so far these affected mostly factual data (e.g., names, hobbies, people you know). The use of location information and other contextual data will greatly increase the risks associated with SNSs.

While several proposals for providing location privacy exist, they are not directly usable in social location sharing settings. A thorough investigation of the problem of privacy-aware location sharing in the context of three specific application settings (friend finder service, recommender system, and city watch application) is needed to both improve the security and privacy of users of such systems, and to stimulate subsequent research in technology, policy, and social and legal sciences.


## Acknowledgements

This work is partially supported by the Swiss National Science Foundation (SNF) under Grant No. 200021-129674.


## 6. REFERENCES


[1] M. Benisch, P. G. Kelley, N. Sadeh, T. S, L. F. Cranor, P. H. Drielsma, and J. Tsai. The impact of expressiveness on the effectiveness of privacy mechanisms for location sharing. Tech Report CMU-ISR-08-141, Carnegie Mellon Univ., Dec. 2008.

[2] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[3] S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge. Location disclosure to social relations: why, when, & what people want to share. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 81–90, New York, NY, USA, 2005. ACM.

[4] P. Dourish and K. Anderson. Collective information practice: Exploring privacy and security as social and cultural phenomena. *Human-Computer Interaction*, 21:319–342, 2006.

[5] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In *Proceedings of Pervasive 2005*, pages 152–170, Munich, Germany, 2005. Springer.

[6] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *MobiSys '03: Proceedings of the 1st international conference on Mobile systems, applications and services*, pages 31–42, New York, NY, USA, 2003. ACM.

[7] M. Gruteser and D. Grunwald. A methodological assessment of location privacy risks in wireless hotspot networks. In *Proceedings of the First International Conference on Security in Pervasive Computing (SPC 2003)*, volume 2802 of *LNCS*, Berlin, 2003. Springer.

[8] B. Hoh and M. Gruteser. Protecting location privacy through path confusion. In *Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SecureComm 2005)*, pages 194–205. IEEE Computer Society, 2005.

[9] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46, 2006.

[10] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of locations. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9(3):58–68, 2005.

[11] J. Krumm. Inference attacks on location tracks. In *In Proceedings of the Fifth International Conference on Pervasive Computing (Pervasive), volume 4480 of LNCS*, pages 127–143. Springer-Verlag, 2007.

[12] M. Langheinrich. Privacy in ubiquitous computing. In J. Krumm, editor, *Ubiquitous Computing*, pages 95–160. CRC Press, Sept. 2009.

[13] G. Myles, A. Friday, and N. Davies. Preserving privacy in environments with location-based applications. *IEEE Pervasive Computing*, 2(1):56–64, 2003.

[14] A. F. Westin. *Privacy and Freedom*. Atheneum, New York, USA, 1967.