Location-Related Privacy in Geo-Social Networks

Geo-social networks (GeoSNs) provide context-aware services that help associate location with users and content. The proliferation of GeoSNs indicates that they're rapidly attracting users. GeoSNs currently offer different types of services, including photo sharing, friend tracking, and "check-ins." However, this ability to reveal users' locations causes new privacy threats, which in turn call for new privacy-protection methods. The authors study four privacy aspects central to these social networks — location, absence, co-location, and identity privacy — and describe possible means of protecting privacy in these circumstances.

Carmen Ruiz Vicente
Aalborg University

Dario Freni and Claudio Bettini
Università degli Studi di Milano

Christian S. Jensen
Aarhus University

Online social networks increasingly enable users to publish geo-located content in real time. Many existing services are designed specifically to enable this functionality, and other services are increasingly assimilating it. Prominent examples include Facebook, Foursquare, Twitter, Google Latitude, Flickr, Gowalla, Loopt, and MyTown. We call such services geo-social networks (GeoSNs): they combine real-time location-reporting capabilities with traditional social network functionality.

Location-aware capability in social networks enables new possibilities, such as "check-ins," where users register their arrival at a location online. A business might offer discounts to those who check-in at its location, for example, thus attracting more customers. In addition to benefiting from such discounts, users can identify popular places, then leverage their social networks for enriched functionality. For instance, a recommendation service that uses check-ins to show currently popular places in town can highlight those that are popular among a user's friends.

However, with this functionality comes increased potential for privacy violations. In this article, we discuss aspects of user privacy that are potentially at risk and examine possible solutions.

Geo-Social Network Structure

Figure 1 depicts the main concepts inherent to GeoSNs, including their relationships. A user is an individual who subscribes to a GeoSN. Users can establish online relationships with other users, and the GeoSN in turn uses such relationships to enrich the services provided. Users establish relationships

 Published by the IEEE Computer Society 1089-7801/11/$26.00 © 2011 IEEE

as a way to capture real-life relations (family, coworker, friend, and so on) or indicate affinities or common interests; they can be symmetric (as with *friendship* in Facebook) or asymmetric (such as *following* on Twitter).

A GeoSN often lets its users share user-generated content with all or some fellow users. This content might be associated with a location, called *geotagging*, and with users, called *user tagging*. For instance, a user can tag a photo uploaded to Picasa with the people who appear in it. The service will automatically extract the photo's geographical coordinates from the file's Exchangeable Image File Format metadata, if present.

A GeoSN locates a user via a location update, which takes two forms: With some services, users can update their locations only at certain, predefined places (check-ins). Other services can continuously or periodically track users, typically via GPS or a service that exploits the communication infrastructure for positioning.

GeoSNs can associate both content and users with locations and vice versa. Most GeoSNs are location-centric in the sense that they enable convenient content retrieval according to location. Thus, Flickr shows photos on a map, and Facebook Places discloses users who are currently in a particular location.

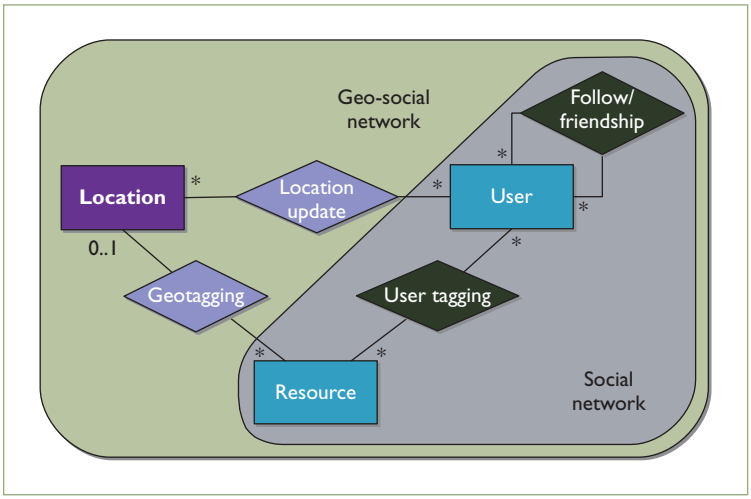

Figure 1. Geo-social network (GeoSN) concepts. GeoSNs enrich traditional social networks with location capabilities.

Privacy Threats in GeoSNs

Privacy advocates frequently warn us about the dangers of exposing location information.[1] Researchers have extensively studied the problem of protecting privacy in location-based services (LBSs), where the service delivered depends on the user's location.[2–8] GeoSNs face additional challenges. In particular, they let users post their locations in the form of a location update or a geotagged resource, so other users can access it. The dissemination of location information among users can be a concern. In addition, user tagging could allow GeoSN participants to report location information about other individuals who have little or no control over the published data.

A privacy threat occurs when an adversary can associate a user's identity with information that the user considers sensitive. GeoSNs expose users to several privacy threats due to the release of spatiotemporal information. As in LBSs in general, two major categories of threats exist: the release of sensitive location information and re-identification through location.

The former category applies to cases in which the user's identity is known, and certain location information pertaining to the user shouldn't be exposed to an adversary. Some GeoSNs let users release location information to other users at a coarser granularity than what's actually available. This is one way to avoid this type of threat, based on the idea that location data's sensitivity decreases at coarser granularities. We demonstrate, however, that current GeoSN privacy protection is inadequate against such threats.

Re-identification through location — a means of compromising the identity privacy of a user[9] — refers to an adversary's ability to reduce a user's degree of anonymity by considering location information. For example, by knowing that an (anonymous) GeoSN user was in a given place at a given time, an adversary could exclude several candidate individuals and possibly identify the user. If the user considers his or her involvement in the GeoSN (or some activities within it) to be sensitive, re-identification is a privacy violation.

We next focus on three privacy notions that might arise in GeoSNs: location, absence, and co-location privacy. We also further examine the identity privacy threat.

Location Privacy

Revealing their exact location to others is a common concern among GeoSN users. Indeed, being able to associate a user's identity with a location at a given time can reveal sensitive

information such as health problems, affiliations, and habits. Consequently, obfuscating a location that an adversary can access is necessary so that the resulting location is no longer sensitive according to the user. We've mentioned that some GeoSNs allow users to specify a generalized location (such as a city) rather than an exact location. However, in GeoSNs, this mechanism can be insufficient to guarantee that the exact location remains hidden from adversaries.

Example 1. Alice and Bob visit a pub. Alice updates her GeoSN status, writing "having a beer with Bob" and tagging the post with 10:05 p.m., Manhattan. A bit later, Bob sees his friend Charlie and invites him to join them. Charlie then updates his status, writing "just met Bob at a pub!" and tagging the post with 10:10 p.m., 24 West 35th Street. A user with access to both posts (for instance, a friend of Bob's) can infer that Alice is at a pub with Charlie and determine the pub's exact address, although Alice used a coarser location in her post.

Reporting a generalized location instead of the exact one is an intuitive way of offering users privacy. However, the example shows how near-simultaneous posts tagged with overlapping sets of users can restrict the possible locations associated with those users. An appropriate privacy-preserving mechanism should guarantee that user preferences are always satisfied, considering the information that an adversary can obtain from a GeoSN.

Absence Privacy

Publishing a user's location can let an adversary infer that the user isn't at a certain place at a given time, as Example 2 shows.

Example 2. Bob works in Manhattan until 5 p.m. One day he leaves early to meet Charlie in Central Park. Charlie updates his GeoSN status, writing "chilling in the park with Bob." If Bob's manager sees the update, he can infer that Bob was absent from the office during work hours.

It's also easy for an adversary to know how far a user is from a location and thus make reasonable assumptions about how long the user might be absent, as Example 3 shows.

Example 3. Bob lives in New York and spends his summer holidays in California with his family. After enjoying a day at the beach, Bob's daughter uploads family photos to her GeoSN profile. A person with access to those photos can infer that Bob's house is likely to be empty for some days and could use this information to plan a burglary.

In both examples, the release of location information represents an absence privacy violation. This threat also applies to traditional LBSs. However, user tagging in GeoSNs enables absence privacy violations in which multiple resources are correlated, as Example 1 shows. To the best of our knowledge, GeoSNs don't currently offer protection against this threat. In an absence privacy preference, the user could specify a location and possibly a time period during which he or she doesn't want to be reported as absent from the location.

Co-Location Privacy

In most GeoSNs, an adversary might be able to observe multiple users' presence in the same place; some users consider such co-location to be sensitive.

Example 4. Alice and Bob meet in a bar, and they don't want to reveal that they've met. While they're there, Alice sees her friend Charlie, who decides to send a geo-located status update saying that he just met Alice. Later, Bob sees his friend Dan, who also updates his status, saying that he saw Bob in the bar. A person with access to Charlie's and Dan's profiles (for instance, Bob's jealous wife), can deduce that Alice and Bob are probably in the same bar.

Disclosing this information to an adversary constitutes a co-location privacy violation. This threat also applies to traditional LBSs. Intuitively, the location information an adversary can obtain by observing multiple users might reveal sensitive information about the relationship among those users. In GeoSNs, user tagging might enable others to discover such relationships even without directly observing the users involved: in the example, having access to Charlie's and Dan's profiles is enough to violate Alice's and Bob's privacy.

None of the GeoSNs we analyzed supports co-location privacy, and specifying privacy preferences related to co-location can be challenging. A basic preference should state those people a user doesn't want to be co-located with, and it might possibly include locations

or time periods for which the preference should apply. Additional aspects that might be important include other users' presence at the same location and the frequency of co-location events. For example, if Alice and Bob attend the same event together with many other users (or with Bob's wife), Alice might not consider their co-location to be sensitive; the same might hold true if the co-location only happens sporadically.

Identity Privacy

Some GeoSNs, such as proximity-based dating services and location-based social gaming, let users employ pseudonyms instead of real names, which hides their real identities. Maintaining user anonymity is particularly important for certain services, as Example 5 illustrates.

Example 5. Bob is using an anonymous location-based dating service, and he's concerned about others discovering it. One night, Bob tells his wife that he's working late and accesses the dating service from his office. Bob's wife, who is suspicious, accesses the same service and looks for people located at Bob's workplace. If Bob's wife knows that Bob is the only one present at the office at that moment, she can deduce that he's using the service — hence violating Bob's privacy.

To avoid user re-identification, some proposals offer anonymity by extending the principle of *k*-anonymity to the geo-spatial setting, which we discuss in further detail in the next section.

Preserving Privacy in GeoSNs

We've seen that publishing geotagged and user-tagged information in GeoSNs enables new attacks. Consequently, existing techniques for traditional LBSs need to be reengineered to enable privacy in GeoSNs.

Existing Techniques for LBSs

Several techniques aim to address location-related privacy threats as they occur in traditional LBSs. We can classify[9] techniques for protection against the release of sensitive location information as *query enlargement*,[10] *fake locations*,[4] *progressive retrieval*,[8] and *encryption-based techniques*.[7] Techniques for protection against re-identification through location alter a reported user location so that it can be associated with at least *k* users (the principle of *k*-anonymity).[2,3,5]

In LBSs, we generally apply these techniques to location-based content retrieval, while, in the following, we focus on techniques that we can adapt to support the publication of location information in GeoSNs. For most GeoSNs, fake location publication isn't an option because it would make the associated resources useless; hence, query enlargement and encryption are the most promising techniques.

Query enlargement techniques can employ *spatial* and *temporal cloaking*. Spatial cloaking generalizes a location into a region. The idea is that an adversary then knows only that the user is located somewhere within that region. Techniques differ in the regions used and how those regions are computed. Spatial cloaking can be applied to obtain regions large enough to lower the information's sensitivity to an acceptable level.

Similarly, temporal cloaking alters the temporal data a user reports. This occurs usually by delaying a service request or, in the context of content-based GeoSNs, delaying the publication of certain content so that an adversary has uncertainty about the actual time associated with it. For instance, suppose a user doesn't want to reveal that he or she was in a particular place at a certain time. The service can eliminate the threat by generalizing the temporal information to a larger interval, such as an entire day.

Finally, encryption-based techniques encrypt the location information (or all the information being exchanged). Recent proposals use private information retrieval (PIR) techniques to allow a service provider to answer a service request without revealing to anybody, including itself, the location data in the request.[7] Other proposals use multiparty secure computations to provide location privacy in proximity services.[10] Techniques differ in the encryption functions and protocols used, and consequently in their costs. Using encryption can introduce additional system costs and might reduce the data's utility to service providers. For example, it might be difficult with current proposals to support location-based queries over encrypted data in a scalable manner.

Existing GeoSN Features

GeoSNs' features and fundamental purpose might render existing techniques applicable

only to limited degrees. Let's examine how such features affect the described techniques' applicability in the GeoSN setting.

Exact location required. Some GeoSNs require using exact locations. Examples include check-in-based services, social navigation services, and user reviews. Offering location privacy in such services via spatial cloaking is problematic because the generalized locations either still identify the places to be hidden (if a finite set of possible places exists) or will compromise the services' utility. Such services can also use encryption so that only trusted recipients can decrypt the location information. On the other hand, in services that don't require using exact locations — such as microblogging, photo sharing, or proximity-notification services — spatial cloaking is applicable.

"Real-time" publication. Some services' utility depends on publishing information in real time because it's inherent to the services' purpose. Examples of strict real-time requirements include dating and proximity services, microblogging, and social navigation services. Such services must apply temporal cloaking sparingly to preserve real-time functionality. In contrast, other GeoSNs, such as user review and photo-sharing services, can tolerate some temporal uncertainty without it substantially affecting the quality of service.

User tagging. The examples we presented highlight how an adversary can violate users' privacy by analyzing and linking resources others have published. GeoSNs generally afford their users little support for setting privacy preferences with regard to information other users upload about them.

Regardless of the strategy adopted to achieve privacy, when content is tagged with multiple users (for example, a photo tagged with everyone who appears in it), the content should be altered so that all tagged users' privacy requirements are satisfied, meaning that an adversary can't violate privacy by observing multiple published resources. Removing a user tag can also be effective when it isn't possible for an adversary to identify that user from the content.

User identity. Some GeoSNs allow pseudonyms — or even anonymous use — instead of real identities. A user participating in a GeoSN anonymously or through a pseudonym should be concerned about re-identification (recall Example 5). The actual risk of re-identification through location depends on the external information that an adversary can acquire to match the anonymous users in that location at a given time with their identities.

Spatial cloaking[2,5] can mitigate this risk, transforming the exact user location into a spatial region that contains k other users, so that an adversary wouldn't be able to associate the request (or content) with a specific user among those k candidates. PIR techniques are also effective against this threat.

Applicability to Existing GeoSNs

Table 1 gives an overview of popular commercial GeoSNs and their relation to the features described in the previous section.

In some GeoSNs, either the temporal or the spatial dimension is less crucial and thus can be generalized if a privacy concern exists. For instance, Twitter doesn't require that users publish an exact location — rather, they can omit the geotag or generalize it to a coarser location (such as a neighborhood or city). In contrast, a tweet's utility generally relies heavily on its publication in real time. Review services such as Yelp or Qype require using an exact location, but reviews need not be published instantly — also, these services let users employ pseudonyms.

In contrast, some services (for example, Whrrl and Waze) require both high spatial and temporal accuracy. According to our analysis, applying any cloaking techniques wouldn't be possible. For these services, encryption is instead a potentially effective solution. Pseudonymous or anonymous use is also possible in these services. To avoid re-identification, anonymization strategies not based on cloaking (such as cooperating with nearby devices to proxy a request so that the issuer stays anonymous) are applicable.

Even when GeoSNs don't strictly require exact locations and times, maintaining a certain degree of accuracy might be important to overall quality of service. For example, if a location reported to Google Latitude is too large (as with a country), other users could perceive this information as irrelevant. Analogously, statistics on Facebook Places about current users

Table 1. Features of existing services.

	Multiple user tagging/check-in	Exact location required	Real-time publication	User identity*
Facebook Places	✓	✓		R
Foursquare		✓		P
Twitter			✓	P
Google Latitude			✓	R
Gowalla		✓		P
MyTown		✓		P
SCVNGR	✓	✓		P
Whrrl		✓	✓	P
MeetMoi		✓	✓	P
Flickr	✓			P
Picasa	✓			P
Brightkite				P
Google Buzz			✓	R
Yelp		✓		A
Qype		✓		P
Grindr			✓	P
Loopt			✓	R
Gbanga		✓		P
Geocaching		✓		P
Waze		✓	✓	A
Trapster		✓	✓	A

*R = real, P = pseudonymous, A = anonymous

located in a certain place are affected if check-ins are delayed due to temporal cloaking.

Emerging Techniques and Challenges

GeoSNs call for new approaches to privacy preservation, and efforts in this direction are emerging. A recent study proposes two solutions based on spatial and temporal cloaking to preserve location and absence privacy in GeoSNs that support user tagging.[11] These solutions use reasoning on the distance and time between resources, user tags, and geotags to detect resources that can cause privacy violations like the ones Examples 1 and 3 illustrate.

This approach achieves location privacy by using *minimum uncertainty regions*.[10] For absence privacy, the privacy preferences users express are called *absence privacy regions* (APRs). The significance of an APR for a user is that no information must be disclosed that makes it possible for an adversary to exclude any location in the APR as the user's possible current location. The proposed technique determines a publication delay for each submitted resource, after which it's safe to publish.

Many open problems remain. Here, we briefly illustrate key challenges.

Threat Formalization

The formalization of the privacy threats described previously in a GeoSN setting is still an open task. It should start from a solid definition of what users can express as privacy preferences.

Although researchers have conducted some initial work for specific combinations of privacy threats and service categories (such as location privacy in proximity services[6,10] or location and absence privacy in GeoSNs with user tagging[11]), a need still exists for a more expressive formal model that's applicable to all of the scenarios we address.

For the absence privacy threat, recent work[11] targets only absence at the current time

(Example 3). An important extension is the generalization to a time interval, as Example 2 presents. Preferences of this type could take the form "Don't exclude that I am in the office during working hours."

The problem of co-location privacy, which we believe hasn't been investigated, is particularly challenging. For instance, a possible co-location privacy preference for Alice in Example 4 could be "Don't reveal my co-location with Bob in less than 50 meters and during the evenings, unless Bob's wife is there as well."

Technique Design

Some GeoSNs have requirements that current privacy-preservation techniques fail to address, so researchers are encouraged to devise new methods for detecting privacy violations and providing protection.

A possibility for protecting co-location privacy is to apply cloaking to one or more of the reported locations so that co-location involves sufficiently many people, and the generalized locations can't be viewed as close with enough confidence. For instance, Alice's co-location privacy preference just described might be satisfied if she reports a large location (or time interval) so that no one can conclude that she and Bob are together.

As for absence privacy, we can apply both spatial and temporal cloaking: spatial cloaking can ensure that the sensitive location is a possible user location, whereas temporal cloaking can introduce uncertainty about the publication time. Recall the absence privacy preference from the previous section: if a resource Alice posts is generalized to the entire day or to a region that contains Alice's workplace, we can't exclude that Alice has been at work.

Managing Historical Data

Given that GeoSNs have emerged only recently, the importance of aspects related to historical geo-referenced data that service providers and users continuously acquire hasn't been fully recognized. Historical data can enable very interesting services — for example, tracing when two users happened to be in the same place or attended the same event in the past, or helping predict future user locations. However, such data can also expose users to new threats; inferences on historical location data, for instance, can lead to loss of anonymity.[3]

Several sites and mainstream articles are alerting users about the risks of oversharing in GeoSNs without appropriate privacy controls. Indeed, users are, in many cases, unaware that they're publishing location information, as when they upload a picture from a smart phone that automatically geotags photos. The lack of privacy controls in GeoSNs makes it difficult for users to understand the risks and take appropriate action to protect their privacy.

In order to address this problem, we must fully understand the privacy issues involved in participating in GeoSNs. We hope that by explaining the setting and identifying technical challenges, our work will increase awareness and stimulate further research.

References

1. G. Friedland and R. Sommer, "Cybercasing the Joint: On the Privacy Implications of Geo-Tagging," *Proc. 5th Usenix Workshop on Hot Topics in Security* (HotSec 10), Usenix Assoc., 2010, pp. 1–8.
2. M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking," *Proc. 1st Int'l Conf. Mobile Systems, Applications, and Services*, Usenix Assoc., 2003, pp. 31–42.
3. C. Bettini, X.S. Wang, and S. Jajodia, "Protecting Privacy against Location-Based Personal Identification," *Proc. 2nd Very Large Databases Workshop on Secure Data Management*, LNCS 3674, Springer, 2005, pp. 185–199.
4. H. Kido, Y. Yanagisawa, and T. Satoh, "An Anonymous Communication Technique Using Dummies for Location-Based Services," *Proc. Int'l Conf. Pervasive Services*, IEEE CS Press, 2005, pp. 88–97.
5. P. Kalnis et al., "Preventing Location-Based Identity Inference in Anonymous Spatial Queries," *IEEE Trans. Knowledge & Data Eng.*, vol. 19, no. 12, 2007, pp. 1719–1733.
6. G. Zhong, I. Goldberg, and U. Hengartner, "Louis, Lester, and Pierre: Three Protocols for Location Privacy," *Privacy Enhancing Technologies*, LNCS 4776, Springer, 2007, pp. 62–76.
7. G. Ghinita et al., "Private Queries in Location-Based Services: Anonymizers Are Not Necessary," *Proc. SIGMOD*, ACM Press, 2008, pp. 121–132.
8. M.L. Yiu et al., "SpaceTwist: Managing the Trade-Offs among Location Privacy, Query Performance, and Query Accuracy in Mobile Services," *Proc. Int'l Conf. Data Eng.*, IEEE CS Press, 2008, pp. 366–375.
9. C.S. Jensen, H. Lu, and M. Lung Yiu, "Location Privacy Techniques in Client-Server Architectures," *Privacy*

in Location-Based Applications, LNCS 5599, Springer, 2009, pp. 31–58.
10. S. Mascetti et al., "Privacy in Geo-Social Networks: Proximity Notification with Untrusted Service Providers and Curious Buddies," *Very Large Databases J.*, 2011; www.springerlink.com/content/y6m385570364876j/.
11. D. Freni et al., "Preserving Location and Absence Privacy in Geo-Social Networks," *Proc. Conf. Information and Knowledge Management*, ACM Press, 2010, pp. 309–318.

Carmen Ruiz Vicente is a PhD student in the Department of Computer Science at Aalborg University, Denmark. Her research interests include location privacy in location-based services and social networks. Ruiz Vicente has a computer engineering degree from the Universidad de Valladolid, Spain. Contact her at carmrui@cs.aau.dk.

Dario Freni is a PhD student in the Dipartimento di Informatica e Comunicazione at the Università degli Studi di Milano, Italy. His research interests include privacy-preserving data management for location-based services and social networks. Freni has an MSc in computer science from the Università degli Studi di Milano. Contact him at dario.freni@unimi.it.

Claudio Bettini is a professor of computer science at the Università degli Studi di Milano, Italy, where he leads the EveryWare Laboratory. He's also a member of the Center for Secure Information Systems at George Mason University. His main research interests are temporal and spatial data management, mobile and pervasive computing, security, and privacy. Bettini has a PhD in computer science from the Università degli Studi di Milano. He's a member of ACM SIGMOD. Contact him at claudio.bettini@unimi.it.

Christian S. Jensen is a professor of computer science at Aarhus University, Denmark. His research concerns data management and spans semantics, modeling, indexing, and query and update processing. Jensen has a PhD in computer science from Aalborg University. He's a member of the Royal Danish Academy of Sciences and Letters, the Danish Academy of Technical Sciences, and the Extending Database Technology Endowment. He's the vice president of ACM SIGMOD and an editor-in-chief of the *International Journal on Very Large Databases*. Contact him at csj@cs.au.dk.