TRANSACTIONS ON DATA PRIVACY 4 (2011) 55–72

Third party geolocation services in LBS: privacy requirements and research issues

Maria Luisa Damiani

Universita degli Studi di Milano (I), E-mail: damiani@dico.unimi.it

Abstract. The advances in positioning technologies and the emergence of geolocation standards opens up to the development of innovative location-based services (LBS), e.g., *web-based LBS*. These services challenge existing privacy protection solutions. For example, the position information is provided by a third party, the location provider, and this party may be not fully trusted. In this paper, we analyze the web-based LBS model. Then we outline the *privacy-aware geolocation* strategy which minimizes the interaction with the untrusted location provider by caching the information that is useful to determine the position in proximity of the private positions, e.g., home, which have been already visited. The deployment of this strategy requires investigating several issues and novel tools. The objective of this paper is to discuss the technical challenges and suggest directions of research towards a comprehensive privacy-preserving framework. To our knowledge, this is the first work on privacy protection against untrusted location providers.

Keywords. Location-based services, location privacy, privacy standards, wifi

1 Introduction

Location-based services (LBS) are gaining increasing attention due to their potential of providing highly-personalized services. Typical services allow to search for points of interest in the vicinity (i.e., which is the closest petrol station) and share the position with the members of a social network (i.e., where are my friends). The conventional client-server architecture of an LBS is illustrated in Figure 1: the user requests the service through the interface of an application running on a mobile device (*client*); the client transfers the position, acquired for example through the GPS or some other trusted location source, to the *LBS provider* which uses such information to customize the service.

The protection of the position information (*location privacy*) is a major issue in LBSs [22]. This concern is due to the fact that LBS providers may not have the ability or interest to protect the position information conveyed by the requesters of the service, while position is a kind of personal information that, as such, requires special care. For example, LBS providers can use position data for purposes other than providing the service, i.e., advertising; or even though there is no malicious intent, data can be stolen or accessed improperly by unauthorized subjects. Because of this, LBS providers are generally considered untrusted.

The lack of user's trust hinders the diffusion of LBSs. For example, according to a recent market research, more than half of consumers in UK are not comfortable with businesses using location-based technology, even if that would improve the customer service [7].

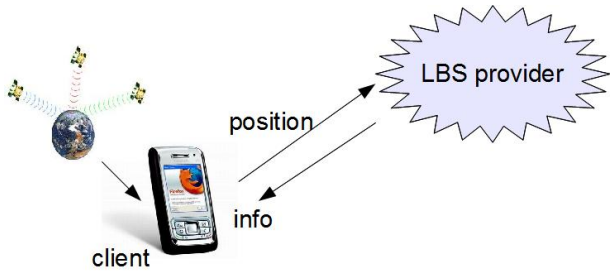

Figure 1: Conventional LBS model: the LBS provider is untrusted while the location source is trusted

In several countries, the collection and use of personal information, including position, is regulated by privacy norms. Privacy norms state that personal information can only be collected respecting a set of principles which can be summarized as: *transparency*, *legitimate purpose* and *proportionality*[1]. Regulations, however, cannot protect position data against malicious parties which deliberately try to violate privacy. That is instead the goal of privacy-enhancing technologies (PET).

In recent years, a broad range of PETs have been developed to protect location information against untrusted providers. LBSs are now rapidly evolving under the pressure of various factors such as advances in positioning systems and business models [22, 11], while new privacy requirements are emerging which challenge existing PETs. In this paper we focus on privacy issues raised by a novel class of services that we term *web-based LBS*. Web-based LBS are services which can be accessed through a geo-enabled web browser (i.e., web browser that can locate the hosting device). Examples of web-based LBSs are Google Latitude[2] and Foursquare[3]. Because these services are representative of the next generation LBSs we believe that this discussion is of general interest.

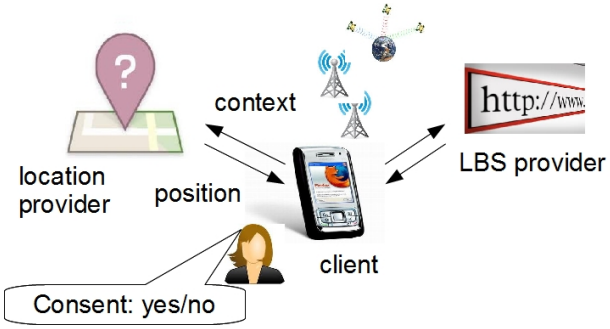

Figure 2: Web-based LBS model: the location source is a third party location provider

[1]Transparency means that "an individual or consumer has the right to be informed when their personal data is being processed; and, has the right to access all data processed about him". Legitimate purpose means that "personal data can only be processed for specified explicit and legitimate purposes and may not be processed further in a way incompatible with those purposes". Proportionality means that "personal data may be processed only insofar as it is adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed" [33]

[2]http://www.google.com/intl/en_us/latitude/intro.html

[3]https://addons.mozilla.org/en-US/firefox/tag/foursquare

1.1 The web-based LBS model

Figure 2 shows the reference architecture of a web-based LBS. The service is accessed through a website which has the capability of locating the client. We refer to the website as LBS provider.

This architecture presents a major novelty with respect to the conventional LBS model, namely the position is computed by a third party, the *location provider*. Examples of location providers are Google Location Service[4] and Skyhook Wireless[5]. Location providers use *hybrid* positioning (or geolocation) systems to provide high quality geolocation services, i.e., accurate and pervasive geolocation, across heterogeneous environments such as indoor/outdoor, and urban/rural settings.

These systems estimate the position based on context information sent by the client (*context* in the figure), e.g., the wifi infrastructure in the vicinity. Notably, the geolocation service is offered on a free basis provided that the user agrees to disclose his/her position.

The information flow is as follows: the user requesting the LBS is asked to give consent for his/her position to be monitored. In case of affirmative consent, the client requests its own position to the location provider. Based on the contextual information, the location provider estimates the position and returns the geographical coordinates, at the finest granularity obtainable, together with the estimated accuracy of the position. The position can then be forwarded to the LBS provider.

1.2 Privacy-aware geolocation

In this scenario, users may not trust location providers for the same reasons they may not trust LBS providers. Even though location providers comply with privacy regulations, it is clear that collected location data are exposed to a variety of risks, such as fraudulent access.

Therefore, if high quality geolocation services are provided by third parties, and those parties are not fully trusted, the high quality of position is paid in terms of privacy loss. The core question is how to balance these conflicting requirements, i.e., if and how users can benefit from high quality geolocation services without disclosing their detailed whereabouts to third party location providers.

Ensuring the protection of the position against location providers is an intriguing issue because it means to protect the position against the party which computes such position. A definitive solution would be to run the geolocation system on the client. This is also the solution of Intel's Place Lab [26], the system precursor of the commercial hybrid positioning systems. The drawback of this solution is that it is not appropriate for the business model under consideration because it is unlikely that geolocation systems comparable in terms of quality of service to those used by third parties can be provided to users for free.

We present a different viewpoint. We advocate that the amount of information that the user conveys to the location provider exceeds what is really necessary to determine the user's position. Accordingly, a more appropriate approach is to minimize the interaction with the location provider. One approach is to record locally selected positions that have already been visited and associate each of those positions with an environmental fingerprint, i.e., a signature based on the value of environmental variables. That way, the client can recognize known positions and interact with the location provider only in those cases in which the current position cannot be found locally. We refer to this form of geolocation

[4] http://www.google.com/intl/zh-CN/events/facultysummit/2010/files/mobile_location.pdf
[5] http://www.skyhookwireless.com/

as *privacy-aware geolocation*. This strategy has the potential of balancing the interests of the third party with the privacy demands of users. A nice feature of privacy-aware geolocation is that it is in line with the data protection principle of *proportionality*.

1.3 Contributions and paper organization

The contribution of this paper is twofold:

- We analyze the components of the web-based LBS architecture. We consider in some detail wifi-based positioning (WPS) because it is becoming increasingly pervasive. Moreover, it can represent one of the enabling factors of privacy-aware geolocation. This first contribution provides the background knowledge and the motivation of the work.
- We outline the features of a comprehensive architectural and modeling framework providing the privacy-aware geolocation service. We define the key components of the framework, and for each of them we describe research issues, the current state-of-the-art, and propose directions of research.

The paper is organized as follows: Section 2 overviews main lines of research on privacy protection in on-line applications. Section 3 presents the key features of the web-based LBS architecture. In Section 4 we introduce the proposed privacy-preserving framework. The discussion on the research issues is presented in Section 5. The final section reports future plans and some conclusions.

2 Related work

There has been considerable work in defining methods for the safeguard of privacy in on-line applications. Closest related work is about the specification of *privacy policies* and related standards, and the development of PETs, respectively. These two paradigms are complementary in that privacy policies emphasize the importance of privacy personalization while PETs emphasize the effectiveness of privacy protection. Both aspects are relevant for the design of a comprehensive privacy-preserving framework. This section briefly overviews the main lines of research within each of these streams.

Policy-based approaches. A privacy policy specifies the privacy practices of an organization, basically what kind of personal information is collected, the purpose and how the information will be used. Privacy policies are defined in compliance with the existing privacy regulations. In simplest form, privacy policies are encoded in natural language and are directly enforced by users. Manual enforcement is however expensive and thus policies are often ignored by users. This problem is particularly relevant when users visit websites. Automating privacy policy enforcement has been the main objective of early research on privacy protection on the web. The most popular privacy framework is P3P [8] which allows websites to publish their policies in XML. While the platform has become a standard supported by major browsers, P3P is not greatly used in practice. A major criticism of P3P is the poor usability of the platform [28].

A different approach, focused on the protection of location information is IETF Geopriv. Specifically, Geopriv is a standard for the transmission of location information over the

Internet [3]. In this case, the policy is specified by the user and not by the organization. Geopriv has not found yet a broad consensus.

Further approaches investigate the usability of formal policy specification languages. For example, the study conducted in [32] analyzes the feedback provided by the users of a mobile location sharing application using solely time-based privacy rules such as: *Show my location between 9 am and 6 pm on Mondays and Wednesdays*.

In summary, privacy policies are flexible and support the personalization of privacy preferences. However, privacy polices only provide a deterrent against privacy violations. Therefore, if a third party decides to violate those norms, in spite of the risk of penalties, the user's position cannot be protected.

Privacy-enhancing techniques. Since the pioneering work of Gruteser et al. [16] and Beresford et al. [4], a large number of solutions have been developed to prevent undesirable inferences that can defeat privacy protection in LBS. In parallel, different taxonomies have been proposed which classify protection techniques [20, 5, 14, 25]. One of the fundamental classification criteria used across the various taxonomies emphasizes the distinction between techniques which target *identity privacy* as opposed to *location privacy* [20].

Identity privacy refers to the protection of the user's anonymity against inferences which relate the user's position to his/her identity. The most popular paradigm for the protection of identity privacy in LBS is *location k-anonymity* [16], which translates into the spatial domain the paradigm of k-anonymity originally developed in the context of relational data publishing [31]. The key idea of location k-anonymity is to degrade the quality of the user's position to generate coarse regions which contain at least k individuals (k-anonymous regions). This way the position cannot be univocally associated with an individual. Methods relying on location k-anonymity, such as [6] assume the existence of a trusted third party, which generates the k-anonymous regions based on the knowledge of all of the users' position.

The second stream of research on location privacy focuses on protection of the user's position, where the position itself represents a type of personal information. As such, position must be protected. Popular methods include techniques which degrade the quality of the position information (*cloaking methods*) such as [13], and techniques based on cryptographic transformations, such as [15]. In all of these approaches, the semantic dimension of positions, i.e., what the position represents for the user, say a hospital or a point along a street, is ignored. A different viewpoint is presented by PROBE [9, 10]. PROBE introduces two novel ideas: i) the user can easily express privacy preferences by selecting the *semantic locations* to protect from a pre-defined list of categories, e.g. hospital, religious buildings, along with the degree of privacy; ii) the cloaking methods generate coarse regions which cover the spatial extent of sensitive locations independently from the user's position. Thus, at runtime if the user's position falls inside one of these regions, that region is returned in place of the exact position. A major limitation of the approach is that the location sources must be trusted. Note that this assumption is common to all the PETs developed for the protection of privacy in LBS. Accordingly PROBE as well as the other solutions cannot be used within the framework of a web-based LBS, whenever location providers are untrusted.

3 The web-based LBS architecture

This section examines relevant features of the web-based LBS architecture and in particular the role of the location provider and how the geolocation functionalities can be requested through geo-enabled browsers. The theme of this analysis is twofold: to demonstrate the need for PETs protecting against untrusted location providers; to provide background knowledge on wifi-based positioning to facilitate the reading of the next sections.

3.1 The location provider

The location providers provide geolocation services. In order to serve clients which can be variously equipped and which can be located in different settings, i.e., urban vs. rural settings, location providers use different types of techniques, integrated in the so-called hybrid positioning systems. These systems encompass one or more cell-based positioning technologies, for example WPS. WPS offers exciting opportunities because the position can be determined both indoors and outdoors, with an accuracy [6] that is sufficient for most LBS. Moreover, WPS does not require a dedicated infrastructure, is cheap, and is commonly available on a variety of devices, e.g., smartphones, laptops. For these reasons wifi geolocation can be also relevant for the deployment of the privacy-aware geolocation service.

3.1.1 Wifi-based positioning system

The wireless LAN (WLAN or wifi) technology is a family of protocols for data transmission based on the IEEE 802.11 specification [37]. The most popular among those standards, i.e, 802.11b/g operate in the 2.4GHz band. At that frequency, the propagating signals are affected by noise and lose strength when encountering physical obstacles, such as walls and human bodies. The WLAN infrastructure consists of a set of wireless access points (APs) connected to a cabled LAN. Each AP identifies a cell. As the wifi-enabled mobile device enters a cell, the device can hear the AP. i.e., receive the beacon frame containing among the others the ID of the AP (i.e., MAC address).

WPS exploits the knowledge of the wifi infrastructure to estimate the position of the mobile device. WPS is used in two different settings, for the localization in indoor spaces as in [2] and for the localization in metropolitan environments, respectively. The Intel's Place Lab system [26] falls into this second stream and is the precursor of the hybrid positioning systems in use today. Place Lab stores in a database (*beacon database*) the association between the IDs of the beacons, which can consist not only of wifi AP, but also of cell phone towers, and fixed Bluetooth devices, and their positions. Clients compute their own location by hearing one or more IDs, looking up the associated beacons positions in a local database, and estimating their own position referenced to the beacons positions [26]. Place Lab is designed to run on the client, i.e., system does not rely on any external structure and processing that could reveal user location and that for privacy safeguard[18]. The commercial solutions, which derive from Place Lab, e.g., Skyhook Wireless[7] and Google Location Service rely on a central database handled by the third party, i.e., the location provider. Since every geolocation request is forwarded to the location provider, the location provider is necessarily aware of the clients locations.

[6] Skyhook's wifi positioning system is supposed to determine the position of a mobile device within 20 to 30 meters (http://en.wikipedia.org/wiki/Skyhook_Wireless)

[7] http://www.skyhookwireless.com/

3.1.2 Communicating with the location provider

Figure 3 shows the architecture representing the client and the location provider. The interaction between the wifi-enabled client and the location provider is as follows: first the client scans the WLANs and detects the APs in the vicinity along with the signal strength for each of them. We refer to this information as *context*. Thus the context is transmitted

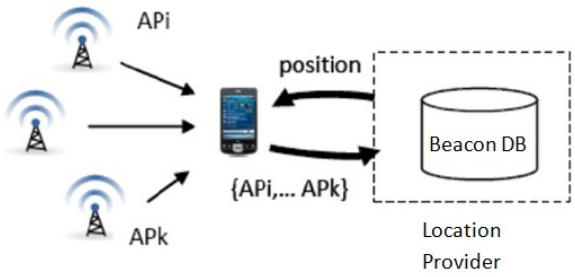

Figure 3: Reference WPS architecture

to the location provider. The location provider estimates the position based on the context and the beacon database. The position is finally returned to the client.

As an example, Figure 4 reports a fragment of the data transmitted by the client requesting the position to the location provider through Google Gears, an early plug-in providing geolocation capabilities to web browsers [8]. In this example, the information which is transferred to the location provider (e.g., Google Location Service) includes, besides the IP address and additional information, the field "wifi_towers" specifying the list of observed APs, each identified by the six-byte MAC Address, along with the signal strength and an additional attribute.

```
{ "version": "1.1.0",
  "host": "maps.google.com",
.......................
  "wifi_towers": [  {
      "mac_address": "01-23-45-67-89-ab",
      "signal_strength": 8,
      "age": 0},
    { "mac_address": "01-23-45-67-89-ac",
      "signal_strength": 4,
      "age": 0} ]}
```

Figure 4: Example of transmitted data

The geolocation service returns the geographical coordinates of the estimated position along with a measure of accuracy that varies depending on both the positioning technique and the context information. The context however fluctuates with time, because the signal is very sensitive to noise and subject to attenuation [23, 38]. Also the position estimated at a point may vary with time.

[8] http://code.google.com/apis/gears/geolocation_network_protocol.html

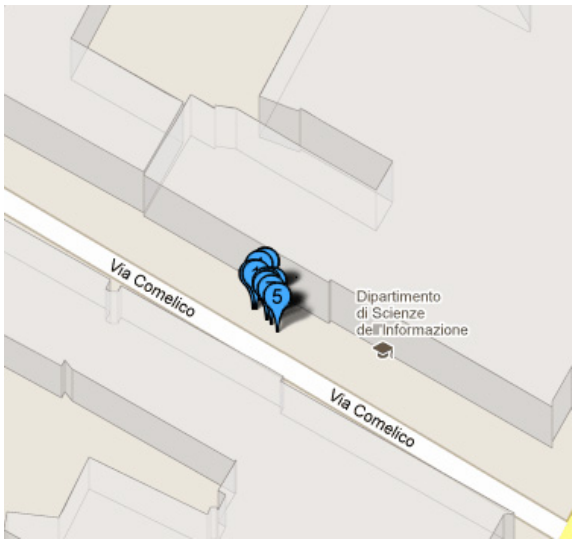

Figure 5: Position distribution at a fixed point

As an example, Figure 5 shows the distribution of the positions at a fixed point inside the University building. In this case the accuracy of the position varies between 20 and 45 meters.

The beacon database is evidently the key resource for this kind of geolocation service. These databases contain millions of records. Various strategies are currently adopted to populate the database, e.g., *wardriving*, i.e., the scanning of wifi networks while driving and *crowdsourcing* in which the database is populated by users. Different techniques are also used to automatically update the database based on the contextual information sent by the clients. The quality of the beacon database definitely affects the quality of the geolocation service.

3.2 Web interface

A simple way to request the geolocation services is through the web. The geo-localization service can be invoked through a standard programming interface (W3C geolocation API, [34]), consisting of a set of functions embedded into scripts. This geolocation standard has found large consensus and all major browsers support it. Moreover, the programming interface is fairly simple, thus embedding geolocation functionalities in web pages is straightforward. A key feature of this interface is that it is agnostic of the underlying location information sources as well as of the geo-locating process, i.e., the way the position is obtained and by whom is completely transparent to the programmer. Another key aspect is that the interface is privacy-aware, in that the specification prescribes that the user agent must, in most cases, get the user's consent to send the device's location to a particular website before initiating a process to obtain a cached or new location [12]. Figure 6 shows the request of consent that is presented to the visitor of the webpage, embedding geolocation functions, using a geo-enabled browser (in italian). The query specifies that the url, in this case the url of the University, wants to monitor the physical position of the user. Moreover, the user is explicitly asked to grant or deny the consent.

Figure 6: The request of user's consent

4 Privacy-aware geolocation: research issues

The development of a comprehensive and coherent framework which minimizes the interaction of the client with the location provider requires investigation of novel approaches and specialized tools. To put these research issues into the right perspective, it is necessary to make assumptions about the work-flow of the privacy-aware geolocation process.

Work-flow. The client maintains a set of positions that have been visited in a local memory, the *cache* (Figure 7). The positions represented in the cache are those which are to be protected. Positions can also represent regions of space. Each position is recorded along with an environmental fingerprint (simply fingerprint) and a representative point. The representative point is acquired from the location provider. Note that this position is communicated to the location provider only the first time the position is visited. To estimate the current position, the client gets in input the fingerprint of the current position and matches it against the fingerprints in the cache. If the matching is successful, then the representative point of the position in the cache is returned; otherwise, the geolocation request is forwarded to the location provider. The output of the geolocation process is thus a point (x,y) that can be either determined locally or obtained from the location provider.

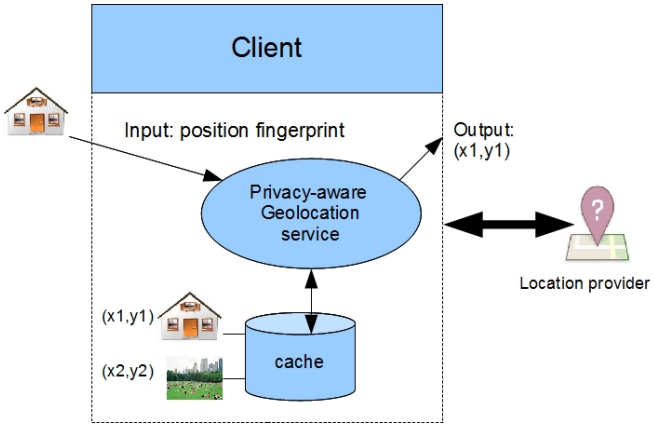

Figure 7: Privacy-aware geolocation workflow

The geolocation service is privacy-aware because it avoids the communication of repeated visits to the same position. Note that the frequency with which a position is visited can reveal a lot about an individual. For example, detecting that an individual is at the hospital at a certain time is an information which may be not sufficient to infer that the user has health concerns. Evidently if *any* position in a time interval falls in the hospital, the information

has a different value.

Research issues. Defining a privacy-preserving framework which supports the above workflow raises several issues:

(1) To define a model of the protected positions. Different criteria can be adopted for selecting the positions to keep in the cache. For example, a trivial criteria is to keep only the most recent positions. The drawback of this approach is that users do not have any form of control. A more appropriate solution is the let the user select the positions to protect based on personal preferences. We refer to these positions as *private positions*. The *model of private position* is the user-oriented model which defines the meaning of private positions.

(2) To define a model for the representation of the private positions in the cache and their recognition. Each private position should be associated with an environmental fingerprint. The *privacy-aware geolocation model* defines the fingerprint model as well as the computation model for the recognition of the private positions in the cache.

(3) To measure the effectiveness of privacy-aware geolocation. Privacy metrics are needed to evaluate the effectiveness of the protection. Those metrics should also allow the users to tune their privacy preferences.

(4) Users can establish different relationships of trust with third parties (i.e., LBS provider and location provider). We distinguish two cases:

   (i) The user does not trust the location provider but trusts the LBS provider. Moreover the two parties are known not to collude. Note that this case is not so unusual as could appear at first sight. For example, a user might want to share the position with the website of the cycling club the user belongs to without letting the location provider know that he/she is at home.

   (ii) Parties are both untrusted. For example, the user wants to share the position with the members of a popular geo-social network without letting the parties know his/her the exact position. The two parties may represent related or even the same organization. The requirement is to extend the protection to LBS providers by integrating existing PETs. We refer to this problem as *PETs interoperability*.

Figure 8 illustrates the dependencies among the above models. Each model is represented by a rectangle. The direction of the arrow has the meaning *depends-from*. The model of the private position is the basic building blocks. Privacy metrics are defined based on the position model. The privacy-aware geolocation model depends on both the position model and the privacy metrics. Finally, PETs interoperability involves all the components.

5 Research directions

We now discuss in more detail the research issues. For each of the previous models, we outline some of the requirements, current state-of-the-art approaches and their limitations, and a possible approach.

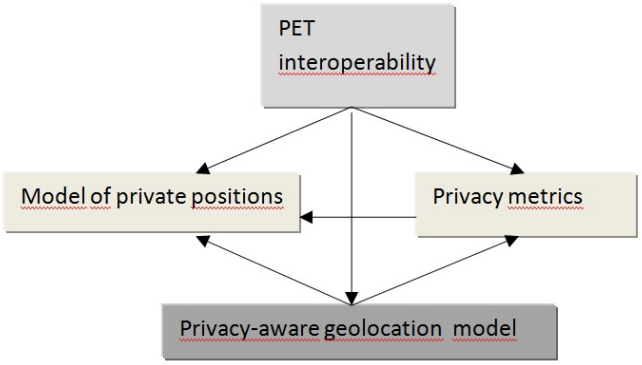

Figure 8: Dependencies among models

5.1 The model of private positions

Requirements. The privacy-preserving framework should not only provide a comprehensive protection of privacy against untrusted providers, but also be usable. The lack of privacy usability is one of the factors that hinder the adoption of PETs. " Usability relates not only to understanding what taking a particular action means in the context of a particular interaction, but also to whether the user understands the implications of his or her choices in a broader context " [30]. Defining which positions the user can protect through high level concepts and metaphors can improve system usability of the system and enhance user experience.

Current approaches and limitations. The usability aspects of privacy have been extensively investigated within the human-computer interaction community (HCI). Research in HCI investigates, among the others, how people perceive privacy threats and "how well a given system facilitates (or inhibits) desired privacy practices" [19]. Research, however, mainly focuses on the usability of privacy polices. The usability of PETs in the LBS realm is an open issue [19]. A PET which is brought as example of poor usability is location k-anonymity in LBS [35]. In fact, although users can personalize the level of privacy by choosing the value of k, it is difficult to figure out which value is most appropriate in a given context.

The position model, commonly adopted by PETs, is generally very basic: the implicit assumption is that the private position is a point in the target space. Moreover any point is potentially private, in that users cannot express preferences on the positions to protect. A different viewpoint is presented in [9]. Positions are naturally aggregated in *places*, and places may require different levels of protection depending on the privacy requirements.

The notion of place is a powerful modeling spatial concept which is increasingly used in a variety of contexts. A place is defined in the area of pervasive computing as "an important locale where the user spends a substantial amount of time and/or visits frequently" [24]. In environmental psychology, *sense of place* is used to mean the"meaning attached to a spatial setting by a person or group" [21]. The notion of places also supports a variety of applications, including personalized mobile searches based on place preference, the analysis of human spatial and temporal behavior [24] as well as spatio-temporal data modeling [29]. Finally, the notion of place is relevant in a variety of commercial applications. For example,

the users of Facebook Place can tag their current position with the name of the place to let their friends know they are or have been in that place. Yet, the full potential of the concept of place in location privacy has not been fully explored.

Approach. The approach is to specialize the concept of place to that of *private place*.
A private place conceptualizes the intuition that there are some regions of space that belong to the personal sphere. Users' positions inside private places should not be disclosed to untrusted parties. Relevant features of the place model are:

- A private place is a place that the user frequents. Whether a place is private or not depends exclusively on the individual's perception and not on physical constraints. Thus a place can be indoor or outdoor. Private places can be tagged with a name, yet those names are not part of any shared geographical ontology, i.e., they can be arbitrarily defined by users. While *home* is perhaps the most popular example of private space, we can imagine private places like: *my boyfriend's home, my department*.

- A private place denotes a space which may not have well-defined boundaries. It is often the case that users are not aware of the geographical extent of a place. For example, an user may have a very approximate idea of the spatial extent of a hospital, yet the hospital can be perceived as private. Of course, users could locate themselves over maps and then identify the private places of concern, i.e., the hospital. This is also the approach in [9]. Yet, a private place may not have any direct relationship with generally recognized places. Moreover, places on maps are generally identified by points, and thus do not have clear boundaries. Further, the operation can be costly for the user.

  A private place thus should reflect the (possibly partial) knowledge that the user has of a certain region of space. A private place can thus be created out of some basic information, for example a single position inside a place, and later the knowledge of that place can expand through a learning process.

The concept of private place is useful also for two other reasons: it naturally supports privacy personalization. Users can choose the private places based on the individual preferences and in this way tune the degree of privacy. Moreover it provides a rationale for the limited protection achievable through the privacy-aware geolocation (i.e., not all positions can be protected).

5.2 Privacy metrics

Requirements. Privacy metrics quantify the extent of privacy problems and the effectiveness of protection. Privacy measures can be also useful to increase the user's awareness on the level of privacy. The application context under consideration is however peculiar. Therefore privacy metrics should be properly tailored.

Current approaches and limitations. In the area of privacy-preserving data publishing, the effectiveness of PETs, in particular of anonymization methods, is measured in terms of privacy loss and utility loss, where: the privacy loss is quantified by the adversary's knowledge gain about the sensitive values on an individual; the utility loss by the information loss about sensitive values of large populations [27]. The fundamental problem is thus to balance the individual interest, i.e., privacy, against the collective interest, i.e., utility. Those

metrics still make sense when the privacy can be personalized, as for example in LBS. In that case the idea of collective interest should be replaced by the notion of individual benefit. e.g., the quality of the service. For example in [9], the protection mechanism discloses coarse locations in place of exact positions. The privacy loss is quantified by the probability that the user in a coarse region, is inside a sensitive place. Instead the utility is measured in terms of positional accuracy.

Privacy metrics originated in data publishing and then translated into the LBS area, however, cannot be directly transposed into the scenario under consideration. The main reason is that, in our case, the privacy-preserving strategy is significantly different: the strategy is not to degrade the quality of the sensitive information, i.e., the location, but simply not to disclose it. It is difficult to say therefore what "degree of privacy" means in this case. Suitable privacy metrics are required.

Approach. The approach is to re-define the notion of privacy and utility. The idea is to consider defining private places as the maximal form of protection achievable in this context. Ideally the amount of protection depends only on the number of private places. In practice the geolocation methods, for the inherent complexity of the problem, are not able to ensure maximum protection. The privacy and utility measures thus express the deviation from the *optimal protection*. We say that the protection is optimal when both these conditions are verified:

(i) No position falling inside private places is communicated to the location provider. We call *false negatives* the private positions which are disclosed to the location provider.

(ii) The only positions which are not requested from the location provider are those which are inside private places. We call *false positives*, the positions which are erroneously considered private.

In other words, the protection is optimal when the protection mechanism does not deviate from the expected behavior. False positives determine positional inaccuracy, because the exact position is approximated by the representative position of a private place. False negatives determine a loss of privacy because private positions are disclosed to the location provider. A simple approach is to take the number of false positives as measure of utility loss; and the number of false negatives as measure of privacy loss.

5.3 Privacy-aware geolocation model

Requirements. Privacy-aware geolocation assumes that private positions can be associated with environmental fingerprints and that those fingerprints can be recognized in real time. Although one can envisage using different kinds of environmental conditions, a reasonable choice is to exploit the wifi infrastructure. Accordingly each private position should be associated with a set of wifi APs.

This is similar to what a WPS does. Yet there is a substantial difference in that in our case the position is estimated by the client and the client is not aware of the physical position of the wifi APs. In other words, clients cannot rely on a beacon database (see Section 3.1.1). At most, clients can detect the wifi context and match it against possible patterns. The ultimate requirement is that the client should be able to recognize private positions based on limited knowledge of the wifi infrastructure. Another requirement is that these techniques should try to minimize privacy and utility loss, in accordance with the selected privacy metrics.

Current approaches and limitations. Approaches which address some of the above requirements are developed within the recent *place discovery* research stream. Place discovery methods attempt to find places that are important to an individual user [24]. These techniques comprise two main classes of solutions, called *geometry-based* and *fingerprint-based* methods, respectively [24]. Of these, the latter class of methods are definitely significant for our problem.

- Geometry-based methods are applied to historical positions (e.g., temporal ordered sequences of GPS readings) to find places where the individual stays for significant periods of time. Approaches are generally based on clustering as in[39], and statistical techniques, as in [36]. These methods are used off-line to discover places that are not known in advance. This case does not match our scenario.

- Location fingerprinting associates location dependent characteristics of radio signals (the *fingerprint*) with a location and uses these characteristics to infer the location [23]. Commonly, fingerprint-based techniques are used for localization purposes, e.g., WPS. Place discovery is different from fingerprint-based localization, not only because of the different granularity of the location information, but also for the objectives. An example of place discovery problem [24] is to recognize that an individual enters and exits from a place which is not known in advance and in which the user has spent significant amount of time. Another type of problem is place learning. The problem is to build in real time and interactively [1, 17] the wifi-based fingerprints of places. For example the approach in [1] uses a learning algorithm to identify a representative set of APs for a place (called *semantically meaningful area*).

  To our knowledge, place discovery techniques have not been used for privacy-preserving purposes.

Approach. The design of the privacy-aware geolocation process can rely on two kinds of methods: i) place learning methods to interactively create wifi-based fingerprints of private places: architectural terms, the effect of this operation is to insert a private place into the cache. ii) Place learning (or place classification) methods to possibly assign a place label to the current position of the client (i.e., its wifi context).
Since the results of these operation may be not always correct, the geolocation operation should return privacy measures, i.e. false positives and false negatives. These measures can then be used to improve the learning process.

5.4 PETs interoperability

Requirements. The problem is to provide a comprehensive framework which not only protects the position from the location provider, but also from untrusted LBS providers. As an example, consider the case of an individual sharing his/her position with the members of a geo-social network. The user does not want to reveal his/her exact position whenever such position is inside a sensitive place, say hospital. Although the user activates the protection against the location provider and creates, for example, the private place "hospital", the representative position of "hospital" is disclosed to the LBS provider every time the service is accessed from that place.
Existing PETs provide various forms of protection against LBS providers. Commonly these techniques map exact positions onto locations providing privacy guarantees. The integration of an existing PET into our framework is seemingly straightforward, because

the PET can simply refer to the privacy-aware geolocation process as location source. This approach, however, presents important limitations. For example, whenever different privacy metrics are used, it is not clear how to measure the effectiveness of the protection as a whole. All this calls for criteria and methods addressing the issue of PETs interoperability.

Current approaches and limitations. We introduce PETs interoperability issues through a few examples. A popular privacy-preserving strategy is to replace the actual position with a coarse location before the location information is transmitted to the LBS provider. The coarse location can be generated in various ways, for example by a trusted third party, which knows the position of individuals and can generate k-anonymous regions. The problem of integrating k-anonymity based PETs is twofold: these techniques require a trusted third party while in our framework the only third party aware of the position of users is the location provider which is untrusted; the k-anonymous region may be contained in some place that the user wants to hide, i.e., the hospital. The degree of personalization is thus limited, i.e., the users cannot select the places to protect. In the ultimate analysis, there is a significant mismatch between the two models of private positions and that seems to preclude an effective integration.

The personalization of the private positions is possible in PROBE [9]. The private positions are places and user can select the places to protect. Even in this case the integration is not straightforward because the two notions of place are different: in PROBE the places are geometric entities with precise boundaries, while in our scenario the places do not have a geometric characterization, i.e., places are represented through radio signal-based fingerprints. The problem is thus how to reconcile the two views in a coherent framework.

Approach. An approach is to experiment with the integration with an existing model to hopefully come up with a more general methodology. The existing model in this case is PROBE. To overcome the position models mismatch, two alternative approaches can be devised: the first approach is to force the user to create places with a geometric extent, for example through the use of maps. In this way the private position model is unique in the system. At run-time, the position returned by the privacy-aware geolocation process is matched against the coarse regions generated off-line by PROBE to cover the private places. If the position falls into one of these regions, such region is transmitted to the LBS provider in place of the exact position. The drawback of this solution is that it requires the user's manual intervention.

The alternative approach is to represent the geometry of private places in approximated manner. For example, the boundaries of private places can be automatically generated from their representative points, e.g. as a circle. The drawback is that the lack of geometric accuracy makes it difficult to measure the degree of privacy and utility of the whole solution.

6 Future plans and conclusion

Protecting privacy in LBS is an evolving research area. Web-based LBS will become increasingly popular in the near future, following the diffusion of geolocation standards, the advances in wifi-based geo-localization and novel business models. Those services offer incredible opportunities to third parties to collect a huge amount of position data in a simple way. Protecting individuals' positions in web-based LBS is thus important. The development of a comprehensive privacy preserving framework poses several technical challenges

because the position information can be provided by an untrusted third party. In this paper, we have outlined major technical issues, and suggested possible directions of research toward the deployment of the concept of privacy-aware geolocation.

We conclude suggesting some priorities and directions for the future: the first direction concerns the development and experiment with fingerprint-based place learning methods to support privacy-aware geolocation. A second line of activities concerns the engineering aspects, i.e., integrating the system into existing open sources geo-enabled web browsers. The third line of activity aims at extending the framework through the integration of existing PETs. Finally, experimenting with the users of web-based LBSs is definitely important to validate the effective usability of the solution.

References

[1] U. Ahmad, B. J. d'Auriol, Y. Lee, and S. Lee. The election algorithm for semantically meaningful location-awareness. In *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia*, MUM '07, 2007.

[2] P. Bahl and V.N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. In *Proc. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, 2000.

[3] R. Barnes, M. Lepinski, A. Cooper, J. Morris, H. Tschofenig, and H. Schulzrinne. An Architecture for Location and Location Privacy in Internet Applications. Technical report, IETF Geopriv Internet Drafts, 2009.

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

[5] C. Bettini, S. Jajodia, P. Samarati, and X. S. Wang (Eds.), editors. *Privacy in Location-Based Applications, State of the Art Survey*. Springer, 2009.

[6] C. Chow, M. F. Mokbel, and W. G. Aref. Casper*: Query Processing for Location Services without Compromising Privacy. *ACM Transactions on Database Systems*, 34(4):1–48, 2009.

[7] computing.co.uk. Most people wary of location-based services, says Ovum. http://www.computing.co.uk/ctg/news/2074892/people-wary-location-services-ovum, 31 May 2011.

[8] L.F. Cranor. P3P: Making Privacy Policies More Useful. *IEEE Security and Privacy*, 1(6):50–55, 2003.

[9] M.L. Damiani, E. Bertino, and C. Silvestri. The PROBE Framework for the Personalized Cloaking of Private Locations. *Transactions on Data Privacy*, 3(2):123–148, 2010.

[10] M.L. Damiani, C. Silvestri, and E. Bertino. Fine-grained cloaking of sensitive positions in location sharing applications. IEEE Pervasive Computing (accepted for publication).

[11] S. Dhar and U. Varshney. Challenges and business models for mobile location-based services and advertising. *Commun. ACM*, 54:121–128, 2011.

[12] N. Doty, D. Mulligan, and E. Wilde. Privacy issues of the W3C Geolocation API. Technical report, UC Berkeley, School of Information, 2010.

[13] M. Duckham and L. Kulik. A Formal Model of Obfuscation and Negotiation for Location Privacy. In *Pervasive Computing*, pages 152–170. Springer, 2005.

[14] M Duckham and L. Kulik. Location privacy and location aware computing. In *Drummond J (ed) Dynamic & mobile GIS: investigating change in space and time. Boca Raton. CRC Press*, 2006.

[15] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K. Tan. Private queries in location based services: anonymizers are not necessary. In *SIGMOD '08*, New York, NY, USA, 2008. ACM.

[16] M. Gruteser and D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proc. of the 1st International Conference on Mobile systems, Applications and Services*. ACM Press, 2003.

[17] J. Hightower, S. Consolvo, A. Lamarca, I. Smith, and J. Hughes. Learning and Recognizing the Places We Go. In *UbiComp*. Springer, 2005.

[18] J. Hightower, A. LaMarca, and I. E. Smith. Practical Lessons from Place Lab. *IEEE Pervasive Computing*, 5(3):32–39, 2006.

[19] G. Iachello and J. Hong. End-User Privacy in Human-Computer Interaction. *Foundations and Trends in Human-Computer Interaction* , 1(1):1–137, 2007.

[20] C. S. Jensen, H. Lu, and M.L. Yiu. Location Privacy Techniques in Client-Server Architectures. In *Privacy in Location-Based Applications: Research Issues and Emerging Trends*. Springer-Verlag, 2009.

[21] B. S. Jorgensen and R. C. Stedman. Sense of place as an attitude: lakeshore owners attitudes towards their properties. *Journal of Environmental Psychology*, 21(3):233 – 248, 2001.

[22] I. Junglas and R. T. Watson. Location-based services. *Commun. ACM*, 51:65–69, March 2008.

[23] K. Kaemarungsi and P. Krishnamurthy. Properties of indoor received signal strength for wlan location fingerprinting. In *MobiQuitous'04*, 2004.

[24] D. H. Kim, J. Hightower, R. Govindan, and D. Estrin. Discovering semantically meaningful places from pervasive rf-beacons. In *Proceedings of the 11th international conference on Ubiquitous computing*, Ubicomp '09, 2009.

[25] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

[26] A. Lamarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F.Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit. Place Lab: Device Positioning Using Radio Beacons in the Wild. In *Proceedings of the Third International Conference on Pervasive Computing*, 2005.

[27] T. Li and N. Li. On the tradeoff between privacy and utility in data publishing. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, 2009.

[28] I. K. Reay, P. Beatty, S. Dick, and J. Miller. A survey and analysis of the p3p protocol's agents, adoption, maintenance, and future. *IEEE Trans. Dependable Secur. Comput.*, 4(2):151–164, 2007.

[29] S. Spaccapietra, C. Parent, M.L. Damiani, J. de Macedo, F. Porto, and C. Vangenot. A conceptual view on trajectories. *Data & Knowledge Engineering*, 65(1):126–146, 2008.

[30] Security Steering Committee on the Usability and Privacy of Computer Systems; National Research Council. Overview of Security, Privacy, and Usability. In *In: Toward Better Usability, Security, and Privacy of Information Technology: Report of a Workshop*, 2010.

[31] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *Int. Journal on Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):571–588, 2002.

[32] J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh. Who's viewed you?: the impact of feedback in a mobile location-sharing application. In *Proc. of the 27th International Conference on Human Factors in Computing Systems(CHI '09)*, 2009.

[33] European Union. European Union Directive 95/46/EC.

[34] W3C. Geolocation api specification. http://dev.w3.org/geo/api/spec-source.html, 2010.

[35] T. Xu and Y. Cai. Feeling-based location privacy protection for location-based services. In *Proc. of the 16th ACM Conference on Computer and Communications Security*, 2009.

[36] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer. Semitri: a framework for semantic annotation of heterogeneous trajectories. In *Proc. of the 14th International Conference on Extending Database Technology*, EDBT/ICDT '11, 2011.

[37] K. Yu, I. Sharp, and J. Guo. *Ground-based wireless positioning*. Wiley, 209.

[38] C. Yu-Chung, Y. Chawathe, A. LaMarca, and J. Krumm. Accuracy characterization for metropolitan-scale wi-fi localization. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, MobiSys '05, 2005.

[39] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen. Discovering personally meaningful places: An interactive clustering approach. *ACM Trans. Inf. Syst.*, 25(3), 2007.