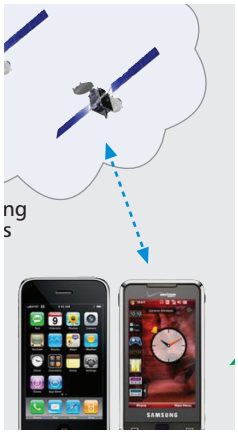
PRIVACY PROTECTION FOR USERS OF LOCATION-BASED SERVICES

KANG G. SHIN, XIAOEN JU, ZHIGANG CHEN, AND XIN HU, THE UNIVERSITY OF MICHIGAN



The authors present a comprehensive overview of the existing schemes for protecting LBS users' privacy. They introduce potential privacy threats to LBS users, and discuss privacy metrics. They classify the protection schemes according to their architectural properties.

ABSTRACT

With the proliferation of mobile devices such as smartphones and tablets, location-based services are becoming increasingly popular. LBSs, albeit useful and convenient, pose a serious threat to users' privacy as they are enticed to reveal their locations to LBS providers via their queries for location-based information. How to protect users' privacy against potentially compromised LBS providers is of vital importance to the well being of the LBS ecosystem, given that the LBS market can expand and prosper only if users feel comfortable about using LBSs. This article presents a comprehensive overview of the existing schemes for protecting LBS users' privacy. We first introduce potential privacy threats to LBS users, followed by a discussion on privacy metrics. We then classify the protection schemes according to their architectural properties (i.e., server-based or mobile-device-based) and privacy metrics (e.g., *k*-anonymity or location entropy). Finally, we discuss several promising directions for future research into LBS users' privacy protection.

INTRODUCTION

Location-based services (LBSs) provide personalized service to smartphone/tablet users by exploiting their location information. As smartphones become increasingly popular and resource-rich, LBSs have become more feature-rich and versatile, improving users' daily lives [1] by, for example, finding restaurants with their favorite menus, obtaining just-in-time coupons from nearby shopping centers, and tracking their physical fitness.

However, LBSs also pose a serious threat to users' privacy. By collecting the location information embedded in the LBS queries, an adversary who has compromised the LBS server can infer sensitive privacy information about service recipients, such as their home locations, life styles, political/religious associations, and health conditions. For example, Hoh *et al.* [2] and Krumm [3] showed that a driver's home location can be inferred from GPS data collected on his/her vehicle even if the location data were pseudonymized or anonymized. In another study, Matsuo [4] exploited a user's indoor location data to infer a variety of personal information, such as work role, smoker or not, coffee drinker or not, and even age. Moreover, Gruteser and Hoh [5, 6] showed that individuals' tracks can be reassembled from completely anonymized GPS data from three or even five users by using multiple hypotheses tracking (MHT) [7].

This problem has received considerable attention from users/consumers, service providers, and government organizations. From the consumers' side, according to a recent survey commissioned by Microsoft [8], LBS users are concerned about the use of their location information and would like to have control over such information. As for service providers, they also have strong incentives to eliminate or mitigate users' privacy concerns because LBSs cannot be successfully marketed unless users are comfortable about using them. Finally, the U.S. Department of Commerce recommended, in December 2010, the inclusion of privacy protection associated with LBSs in electronic privacy laws in order to provide stronger protection enforcement with legislation support [9].

Researchers have long been aware of the potential privacy risks associated with LBSs, and have proposed a number of promising schemes that can help users protect their privacy. In this article, we provide a comprehensive review of the existing techniques and also suggest future research directions to enhance LBS users' privacy.

The remainder of this article is organized as follows. We first present a brief overview of LBSs and provide a number of representative real-world LBS applications. We then discuss the privacy threat of LBSs, as well as the common threat model used in the LBS privacy protection research. We also show existing privacy metrics and protection schemes, focusing on *k*-anonymity and location entropy metrics for the former, and on location obfuscation schemes for the latter. Finally, we conclude the article with a brief look at future research directions.

OVERVIEW OF LBS

In this section, we first briefly review the development of LBS and its diverse applications. We then present a common LBS architecture, followed by a generic threat model.

1536-1284/12/$25.00 © 2012 IEEE

Figure 1. *LBS application examples.*

Today's LBS combines the results of location-related research with other attractive features into complex location-aware applications. A main use of LBS is to provide an easy means of location information sharing and location-aware information retrieval.

Development and Applications of LBSs

LBS was pioneered by early location research efforts [10], including the Active Badge system [11], Microsoft's RADAR system [12], MIT's Cricket project [13], and Intel's Place Lab project [14], each providing location awareness with different localization methods suitable for indoor scenarios, outdoor scenarios, or both.

Today's LBSs combine the results of location-related research with other attractive features into complex location-aware applications. The main use of LBSs is to provide an easy means of location information sharing and location-aware information retrieval. For example, foursquare [15] is a popular location-based social networking application that enables users to interact with their circumstances via mobile devices. Registered users can share their location information with their friends, and receive to-do lists based on their current locations, which facilitates their exploration of exciting ongoing events in their vicinity. Another important category of applications is advertisements. For example, ShopAlerts [16], AT&T's newly launched LBS, is the first large-scale location-aware mobile marketing program in the United States, delivering coupons and special deals to registered consumers via their mobile devices when they are near a participating retailer or brand. The third commonly used LBS application is related to navigation and tracking. For example, the OnStar service provided by General Motors [17] is a comprehensive LBS system that supports a number of location-aware features such as turn-by-turn navigation and stolen vehicle tracking. Other LBS applications include location-aware weather reports and location-aware emergency medical services. Figure 1 shows some common popular LBS applications.

Common LBS Architecture

A common LBS architecture, as illustrated in Fig. 2, consists of four major entities: mobile devices, positioning systems, communication networks, and service providers. Users utilize their *mobile devices* (e.g., smartphones) to send queries to LBS servers. The locations in the queries are obtained via *positioning systems* such as the Global Positioning System (GPS). The queries, as well as their responses from the LBS servers, are transmitted via *communication networks*, such as third-generation (3G) networks. LBS servers are *service providers*, replying to

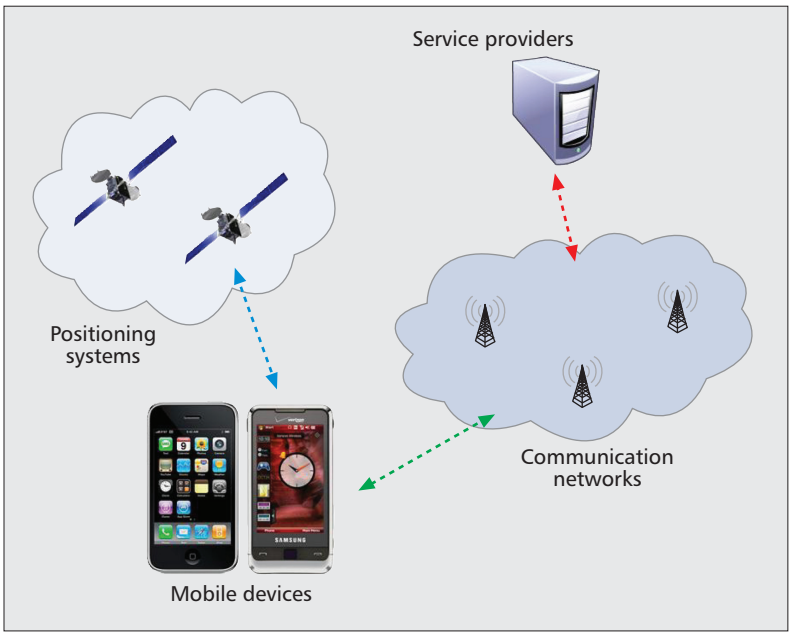


Figure 2. *A common LBS architecture.*

queries with well tailored responses based on the location information in the queries.

Common Threat Model

In the threat model commonly used in LBS privacy protection studies, LBS servers are regarded as malicious observers, and all the other components in the LBS architecture can be considered benign. An adversary can be the owner/maintainer of an LBS server, or able to compromise and then seize control of an LBS server. In both cases, the adversary has the ability to access all the information stored on the servers, such as the IP address and location information associated with each query. Although we embrace a generic model without imposing additional requirements on the usage of LBSs (e.g., we do not assume that users need to login to use LBS), an adversary can still make use of side channels (e.g., the IP address of each query), as well as sophisticated object tracking algorithms (e.g., MHT mentioned above) to relate consecutive anonymous LBS queries to the same user.

Note that this widely accepted threat model intentionally simplifies the privacy protection problem by concentrating only on how to regulate the location information contained inside LBS queries. In reality, the location privacy issue is a complicated and multifaceted problem that needs to be handled with care from several perspectives. For example, users' mobile devices can be compromised, thus becoming malicious and actively revealing users' privacy information (including, but not limited to, location information). How to secure users' mobile devices is an active research topic by itself, and interested readers are referred to the related literature, such as [18].

In addition, while LBS queries and responses are transmitted via communication networks, they need to be protected against eavesdroppers and man-in-the-middle attacks. Conventional solutions, such as cryptography and hashing, can be used to protect the secrecy, integrity, and freshness of the queries and responses transferred through networks.

For the rest of this article, we discuss the existing location privacy protection schemes for LBSs, all of which assume a generic threat model, without considering the issues discussed in this section.

Privacy Issues

Albeit valuable and promising, LBSs pose a serious threat to users' privacy. As mentioned earlier, the privacy concern has received significant attention from the research, government, and industry communities, and is believed to be pivotal to the overall LBS industry.

There are two major types of LBS-related privacy: *query privacy* (e.g., [19–22]) and *location privacy* (e.g., [6, 23, 24]). Query privacy refers to users' private information related to LBS query attributes. Example issues of query privacy include:

- Whether a user can be identified (i.e., de-anonymized) by a malicious LBS provider
- Whether a user's interest and/or habit can be inferred from LBS query contents, among others

Location privacy refers to users' private information directly related to their locations, as well as other private information that can be inferred from the location information. Example issues of location privacy include

- Whether a user can be accurately located
- Whether a user's interest and/or habit can be inferred from the location information contained in LBS queries, among others

These two types are closely related. On one hand, if a user can be accurately located and tracked (i.e., his/her location privacy is compromised), then s/he may be relatively easily de-anonymized (i.e., his/her query privacy is compromised). On the other hand, if a user can be identified, then his/her location privacy can be relatively easily compromised. It is because in this way, an adversary has more information available (e.g., an identified user's historical trace record) for achieving location-related inference attacks (e.g., movement tracking).

Nevertheless, query privacy and location privacy are two different aspects of general privacy. It is possible that one of them can be compromised while the other is preserved. We use the following two examples to illustrate the difference between these two aspects.

Example 1: When sending a query to an LBS provider, the location information contained in that query covers an area with k users of that LBS. In this way, the LBS provider cannot distinguish the query sender from the other $k - 1$ users in that area at that time. According to k-anonymity — a query anonymity metric commonly used in LBS-related privacy protection — the user's query privacy is properly protected. However, all the k users may happen to be within a restricted area, which in turn causes their location privacy to be compromised. In other words, although the LBS provider cannot identify the query sender, s/he may still obtain rela-

tively accurate location information related to that query sender.

Example 2: A user resides in a rural area where s/he happens to be the only active user of a certain LBS. When sending a query to the LBS provider, the user only discloses his/her location information as a relatively large area covering the accurate location. This way, the user's location privacy can be considered preserved because the LBS provider cannot obtain his/her accurate location. His query privacy, however, is still compromised because the LBS provider knows that all the queries can be related to the same user. In other words, the user's queries are de-anonymized.

These two aspects of LBS privacy — query privacy and location privacy — naturally lead to two categories of privacy metrics.

Privacy Metrics

Numerous privacy metrics have been proposed in the LBS privacy protection community for the purpose of quantifying users' privacy. Both query and location privacy metrics have been thoroughly studied.

A broad range of terminology has been proposed to define the privacy metrics. A thorough study of the terminology used in privacy research in a general setting can be found in [25].

Query Privacy Metrics

k-anonymity is the most popular metric used for LBS query privacy protection. The concept of *k*-anonymity was first proposed in the database research community [26] and then quickly became a popular privacy metric among LBS privacy protection researchers. The original form of this metric specifies that "a release provides *k*-anonymity protection if the information for each person contained in the release cannot be distinguished from at least $k - 1$ individuals whose information also appears in the release." In the context of LBSs, this metric refers to the situation in which the location information in an LBS query corresponds to an area where the query sender is indistinguishable from at least $k - 1$ other users also present in that area. This popular metric has been adopted in a number of different approaches, including the generation of spatial cloaking boxes [20] and spatio-temporal cloaking boxes [19]. We discuss and compare the details of the various approaches.

Location entropy, stemming from Shannon's entropy [27] in information theory, is another widely used metric for measuring the uncertainty associated with location information in LBS queries. This metric quantifies the information an adversary can obtain from one (or a series) of location update(s). Note that this metric can be used as both a query privacy metric and a location privacy metric, depending on how the entropy is defined. Here we focus on the query-privacy related usage of this metric, and discuss its usage for location privacy in the next section.

Despite the wide application of location entropy, researchers have not reached consensus on the selection of random variables whose probability mass function should be used in the entropy calculation. Several candidate probability mass functions have been studied, including the probability that:

- A user enters a predefined area from a certain preceding area and then leaves for a certain subsequent area (the "area" may refer to a mix zone [28] or a geographic map grid [24]).
- A location sample belongs to a certain vehicle [29].
- A user's historical trajectories are inside a certain area [30].

Several other approaches have extended the concept of location entropy and proposed new privacy metrics for better quantification of location privacy. For example, Xu and Cai [30] define the popularity of a spatial region as 2^E, where E is the location entropy. The authors then compute the popularity on a per-user basis and can thus further define the *P*-popular trajectory (PPT) for each user according to the popularity configuration (i.e., the *P* value in PPT). Hoh *et al.* [29] defines *time to confusion* with the observation that the degree of privacy risk strongly depends on the duration for which the user can be tracked by an adversary (with high confidence). They first generate traceable paths based on the linkability of a set of location samples with respect to a predefined location entropy threshold (called an *uncertainty threshold*) and then compute the time to confusion as the difference between the timestamp of the first location sample and that of the last one.

Besides *k*-anonymity and location entropy, a number of other query privacy metrics have also been proposed. For example, ubiquity, congestion, and uniformity have been used in [31] to generate an enhanced anonymity set, which in turn serves as the location anonymity measurement for LBS queries. There, ubiquity is present when users exist in an entire area, thus increasing the location anonymity for that entire area. Congestion means that there is a concentration of users in a local region, increasing the anonymity of that specific region. Finally, uniformity requires that each distributed region contains the same number of users. The authors mentioned that a high-quality anonymity set cannot be generated without taking into consideration all three metrics.

Query attributes can also be employed by an adversary to infer LBS users' private information. For example, if consecutive queries from one user are related to bars, an adversary could infer that the user is likely to be alcoholic. Thus, how diverse LBS query attributes are from an adversary's perspective is another interesting LBS query privacy metric [22].

Query attributes can also be employed by an adversary to infer LBS users' private information. For example, if consecutive queries from one user are related to bars, then an adversary could infer that the user is likely to be alcoholic.

Location Privacy Metrics

Location entropy can also be used as a location privacy metric. For example, Chen [23] applies the concept of "unobservability" [25] to LBS privacy protection. Unobservability represents location entropy in an intuitive way by allowing users to specify the number of points of interest (POIs) that can be considered equally likely to be visited by users from the locations they revealed to the LBS servers.

Expected distance error was used in [6] to measure how accurately an adversary can estimate a user's position. The authors justify their

Although the policy-based schemes have advanced the state of the art of the privacy protection technology, stronger enforcement approaches are preferred to achieve better protection and stronger guarantees.

choice by indicating that the location entropy calculated by using the probabilities for different assignments of user identities to the observed locations (in the queries sent to LBS) is insufficient in that the metric does not take into consideration the difference among those locations reported to the adversary. A similar but more general use of the adversary's estimation error for quantifying location privacy can be found in [32]. This work formalized the concepts of accuracy, certainty, and correctness related to the adversary's inference attacks, and pointed out that the expected estimation error of the adversary (i.e., incorrectness) is the appropriate location privacy metric. The expected estimation error in [32] is a generic location privacy metric in that it can be used in any adversary model where the distance (between the adversary's estimation and the true value) can be properly defined.

Recently, researchers introduced the theory of private information retrieval (PIR) [33, 34] to the LBS privacy protection community and proposed solutions in a theoretical setting by transforming the location privacy protection problem into the nearest neighbor (NN) problem [35]. In this approach, privacy protection is achieved via cryptographic techniques. No privacy metric is needed because the location information is not revealed to the adversary, and it is computationally intractable for the adversary to find a user's location based on the data communicated between the user and the LBS server.

Privacy Protection Schemes

In this section, we classify the existing privacy-protection schemes in three categories: policy, location perturbation and obfuscation, and PIR-based approaches. We first briefly discuss policy-based protection approaches, and then focus on the various location perturbation techniques which are currently the main research direction for LBS privacy protection. We also briefly discuss some other techniques not belonging to the two categories at the end of this section.

Policy-Based Schemes

The Platform for Privacy Preferences Project (P3P) [36] is a policy-based approach for protection of users' data privacy. Being developed by the World Wide Web Consortium (W3C), P3P helps websites express their data usage and management practices; also, users understand the server-provided policies, thus facilitating privacy-related decisions. This approach is applicable to LBSs. For common web browser users, the application of P3P in LBSs is believed to enhance the understanding of privacy policies as well as trust in LBS providers. Despite these potential benefits, the lack of policy enforcement specified by service providers is widely considered a significant drawback, rendering the approach an incomplete solution. To rely on P3P, the users have to trust service providers to faithfully follow P3P.

Unlike P3P's broad goal of protecting data privacy in general, the Internet Engineering Task Force (IETF) Geographic Location/Privacy (GEOPRIV) Working Group [37] focuses on how to properly represent location information in the Internet protocols, and investigates the privacy issues related to the location information when it is created, stored, and used. This working group aims to deliver specifications that are widely applicable to location-aware applications.

Depending on how policies are designed, approaches in this category can be used to protect both query and location privacy.

Location Perturbation and Obfuscation Schemes

Although the policy-based schemes have advanced the state of the art of the privacy protection technology, stronger enforcement approaches are preferred to achieve better protection and stronger guarantees. This is because the more control users have over their location data, the more secure they feel, and the more likely the global LBS ecosystem can prosper.

Over the past decade, location perturbation and obfuscation has been the most active research direction in the LBS privacy protection community. A large number of such protection systems have been designed, each with its own strengths and shortcomings. We categorize this related work according to the overall architecture and privacy metrics. We first discuss a popular architecture in which a centralized trusted anonymization server functions as the safeguard of users' location information. We then discuss mobile-device-based protection approaches at the client side.

Trusted Anonymization Server-Based Schemes — In their pioneering work [19], Gruteser and Grunwald introduced the concept of k-anonymity to the LBS privacy protection research community. They represented the location information as a tuple consisting of three intervals ($[x_1, x_2]$, $[y_1, y_2]$, $[t_1, t_2]$), with the first two representing a spatial area where a user is located and the third the time interval for which the user is in the area. Using this representation, they designed an adaptive interval cloaking algorithm that generates spatio-temporal cloaking boxes containing at least $k_{\min}$ users and use the boxes as location information sent to the LBS server. The global parameter, $k_{\min}$, is the size of the minimum acceptable anonymity set, indicating how well a user's location is camouflaged. The drawbacks of this solution include susceptibility to single-point failures and the trustworthiness related to the centralized anonymization server; the inflexibility in tuning the protection level due to its global setting of $k_{\min}$; the lack of guarantee for the resolution of the location information sent to the LBS server; and the inability to provide satisfactory protection in sparse areas.

Recently, significant efforts have been made to overcome these drawbacks. For instance, several efforts have tackled how to enable LBS users to personalize their privacy requirements instead of the global setting in the anonymization server. For example, CliqueCloak [21] is a personalized k-anonymity model in which users can adjust their minimum level of anonymity, and the maximum temporal and spatial resolutions they can tolerate. This work modeled the anonymization constraints as a constraint graph,

and thus transformed the problem of finding cloaking boxes into that of finding cliques that satisfy certain conditions in the constraint graph.

Casper [20] is another personalizable *k*-anonymity-based LBS-related privacy protection framework. A location anonymizer resides on a trusted server, and can use a per-user privacy profile to satisfy each user's privacy protection requirements. A user privacy profile combines the user's privacy protection target (i.e., the k value in *k*-anonymity) and the minimal acceptable location resolution (below which the user's privacy is considered compromised even if the *k*-anonymity condition is satisfied). Moreover, Casper proposed the use of incomplete pyramid structure [38]-a data structure commonly used in LBS privacy protection schemes-in order to adaptively maintain the users' location information over an area in the anonymization server, thus lowering both location update and cloaking costs.

The authors of [39] pointed out that historical locations of different mobile devices (called *footprint*) should be considered in addition to their current locations by the trusted server to enhance the protection of users' location privacy. The authors of [40] proposed "policy-aware" *k*-anonymity, defending against more realistic adversaries who are aware of the policy for cloaking box generation.

k-anonymity-based approaches aim to protect LBS query privacy because the very nature of this metric is anonymity — that is, how to make query senders indistinguishable.

Besides *k*-anonymity, several other metrics have also been applied to the centralized trusted server architecture. Noticeably, Beresford and Stajano [28] introduced the concept of mix zones. A mix zone is an area satisfying the following two conditions:

- No user updates his/her location information to LBSs inside the area.
- Each user is assigned a new pseudonym when leaving the area.

Mix zones weaken the adversary's ability to relate a new pseudonym with the old one. Both the size of an anonymity set and a location entropy-based metric are used to evaluate the effectiveness of mix zones. Results show that the floor plan layout has a significant impact on the protection effect of mix zones. Furthermore, the effectiveness of protection decreases significantly if an adversary takes into consideration the movement patterns of users, indicating that the location entropy indeed generates a more accurate estimate of the available uncertainty, and should thus be considered a useful metric for designing privacy-protection systems. Mix zones are for protecting LBS query privacy, because this approach makes it difficult to link different pseudonyms outside mix zones back to one user, which in turn helps keep queries quasi-anonymous.

Hoh and Gruteser [6] made use of the expected distance error to quantify the accuracy with which an adversary can estimate a user's location, and achieved privacy protection through path confusion. The key idea is to have paths crossed with each other; that is, at least two users' paths intersect via a perturbation algorithm in the trusted anonymization server so as to increase the chance that an adversary would be confused about the paths of different users. However, the authors pointed out that the protection effectiveness depends on the characteristics of users' traces. Moreover, whether the perturbation algorithm could preserve its effectiveness when confronted with an adversary with a priori knowledge of geographic maps and common movement patterns, remains an open question. This approach aims to protect both query and location privacy because it reduces the adversary's ability to track LBS users, and also increases the distance error between an adversary's location estimations and users' accurate locations.

Another trusted anonymization server-based scheme is CacheCloak [24]. It caches LBS responses along the predicted paths and uses them to reply to users' queries. Hence, it achieves real-time location privacy protection without loss of location accuracy from the LBS providers' perspective. But LBS providers can only observe intersecting paths predicted by CacheCloak, and have difficulty in inferring the real locations of the query senders. A key contribution of CacheCloak lies in the fact that it resolves the unfortunate trade-off between privacy protection and usefulness of LBSs, which was considered the zero game by most previous researchers. However, since this approach uses cached information to reply to users' queries, the scalability of the system may become questionable with the proliferation of the LBS market. In the worst case, if each user requires a different kind of LBS, the trusted server may have to cache a myriad of information from various LBS servers, and thus become a scalability bottleneck for the system. This approach aims to protect both query and location privacy due to the increased location anonymity and the untrackability of users' movement.

Xu and Cai [30] proposed a feeling-based location privacy model, used the entropy to measure the popularity of a region, and took a quadtree-style approach [41] to prevent an adversary from relating LBS queries to specific users. Extending their previous findings in [39], this work facilitated the expression of users' intended protection level by a feeling-based privacy modeling-users can indicate their desired protection level by circling a region where they feel comfortable using this region as a location in the LBS queries sent within that region. Specifically, the authors define the popularity of a spatial region as 2^E, where E is the entropy calculated using visitors' footprints inside the region. The authors then compute the popularity on a per-user basis and can thus further define the *P*-popular trajectory (PPT) for each user according to the popularity configuration (i.e., the *P* value in PPT). This approach is a query privacy protection technique because it prevents an adversary from identifying the query sender apart from the other users in a selected cloaking set.

Mobile Device-Based Schemes — The intrinsic drawbacks related to the centralized trusted anonymization server schemes — especially the

Over the past decade, location perturbation and obfuscation has been the most active research direction in the LBS privacy protection community. A large number of such protection systems have been designed, each with its own strengths and shortcomings.

Are there any satisfactory thresholds of such metrics above which a user would feel comfortable using a LBS system? Such issues should be investigated in a broader context, including analyses of extensive user location traces, mobile user surveys, and psychological evaluations.

susceptibility to single-point failures and the trustworthiness of the server — have led researchers to explore different protection schemes that are applicable to users' mobile devices. In terms of practicality and ease of deployment, mobile device-based approaches have advantages over trusted server-based approaches, provided the two can achieve a similar level of protection. This is because in the real world, it is difficult, if not impossible, to deploy a trusted anonymization server with a large user base. In contrast, by using mobile device-based schemes, users can protect their location privacy with their own mobile devices, thus greatly reducing the trusted computing base (TCB) of the global LBS architecture.

Although *k*-anonymity seems most suitable for server-based protection approaches given the need for global knowledge about the locations of a large number of mobile devices in order to achieve *k*-anonymous cloaking, researchers have already proposed ways to employ *k*-anonymity in mobile device-based protection schemes. For example, CAP [42] used a quad tree to maintain road-density information and conducted the Various-grid-length Hilbert Curve (VHC) mapping and perturbation to achieve *k*-anonymity. Since enforcing *k*-anonymity on mobile devices requires the knowledge of a geographic map and a simple solution to the problem of storing a whole map on the mobile device inevitably incurs high computational and storage overheads, CAP proposed the use of VHC mapping. This mapping projects the two-dimensional map into one dimension and has two attractive properties:

- Every point in the projected space has a homogeneous context.
- Adjacent locations in the original space remain close in the projected space.

Thus, CAP reduces the computational and storage overhead, and renders the protection scheme practical for mobile device usage.

Another mobile-device-based scheme is SMILE [43]. SMILE applied *k*-anonymity to measure and configure the users' privacy level for the use of encounter-based LBSs. This method protects users' privacy by selecting the prefix length of the location hash values so as not to reveal encounter involvements to untrusted servers. The intelligent use of cryptographic tools in SMILE's messaging protocol excludes the trusted anonymization server from the LBS architecture. However, this approach may not be applied easily to a broader range of LBSs, such as those in which the notion of an encounter does not exist.

In [44], *k*-anonymous cloaking boxes are generated by employing the mobile devices to build the proximity information among the users via the received signal strength or the time difference of arrivals. However, this approach may become ineffective if the query sender cannot detect a sufficient number of users in the vicinity.

Non-*k*-anonymity schemes have also been studied. For example, Kido *et al.* [31] investigated ways to hide real user movement with dummies. By generating dummies with movement patterns similar to those of real users, and providing LBS servers with mixed locations of real users and dummies, they successfully protected users' location privacy without any trusted server. Despite this success, a malicious LBS server may be able to differentiate the real user from those dummies after long-term movement tracking.

Chen *et al.* [23] designed LISA, a new location privacy protection scheme that prevents an adversary from relating any particular POI to the user's current location based on the *m*-unobservability metric. LISA guarantees, with high probability, that the uncertainty related to each location revealed to the LBS servers is at least as high as the uncertainty corresponding to the situation in which *m* POIs are equally likely to be visited by a user from that location. In addition, LISA employs an extended Kalman filter to predict users' movement patterns, thus improving overall privacy protection effectiveness and also conserving resources. This is a location privacy protection technique due to the fact that it weakens the adversary's ability to infer users' destinations based on query location information.

Other Schemes

Some recent approaches have made use of private information retrieval (PIR) for achieving stronger and provable location privacy. For example, Ghinita *et al.* [35] demonstrated the practicality of applying computational PIR to the protection of LBS users' privacy by solving the nearest neighbor (NN) search in a theoretical setting. The advantage of this approach is that it does not disclose any spatial information, and thus prevents any type of location-based attacks, including correlation attacks. However, this approach incurs significant computational overhead on the server side, imposing stringent requirements on the deployment of LBS servers. Papadopoulos *et al.* [45] employed secure hardware-aided PIR and achieved *strong location privacy* for *k* nearest neighbor (*k*NN) queries; that is, the adversary cannot distinguish the query location from any other locations in the data space.

Pingley *et al.* [22] designed DUMMY-Q, increasing query-attribute uncertainty observed by an adversary by hiding the real query with a number of dummy queries of different attributes from the same location. Two aspects are considered for the selection of dummy query attribute values: the query context and the motion model. By taking into consideration the plausible query attributes for possible locations along a user's current movement, [22] protected users' query privacy against potential inference attacks based on query attributes.

Summary

Table 1 summarizes the existing protection approaches discussed so far.

Future Research Directions

Privacy Metrics

As discussed earlier, a number of privacy metrics have been proposed and studied. Despite these efforts, however, it remains unclear which of them is the best fit for a given protection sce-

	Category			Architecture		Privacy aspect	
Approaches	Policy	Obfuscation	Others	Central server	Mobile device	Query privacy	Location privacy
P3P	*					*	*
GEOPRIV	*					*	*
K-anonymity pioneer work		*		*	*	*	
CliqueCloak		*		*		*	
Casper		*		*		*	
Footprints *k*-anonymity		*		*		*	
Policy-aware *k*-anonymity		*		*		*	
Mix zones		*		*		*	
Path confusion		*		*		*	*
CacheCloak		*		*	*	*	*
Feeling-based pyramid		*		*		*	
CAP		*			*	*	
SMILE		*			*	*	
Proximity		*			*	*	
Dummy positions		*			*		*
LISA		*			*		*
PIR			*		*		*
DUMMY-Q			*		*	*	

Table 1. *Protection scheme summary.*

nario. A thorough comparison of the existing approaches would be beneficial to the privacy protection research community and industry to gain a better understanding of the state of the-art. Several pioneering efforts (e.g., [32]) serve as good examples in this direction.

Moreover, there have been few studies on how quantitative privacy metrics translate to privacy protection in the real world. For example, would $k = 8$ in k-anonymity, or 4 at privacy entropy be good enough? Are there any satisfactory thresholds of such metrics above which a user would feel comfortable using an LBS system? Such issues should be investigated in a broader context, including analyses of extensive user location traces, mobile user surveys, and psychological evaluations.

Protection Schemes

Mobile device-based protection schemes may become more intriguing as they avoid the single-point failure and trustworthiness issues related to the centralized trusted anonymization server.

Although mobile devices are becoming more powerful, there is still a large gap between their computing power and storage space and those of commodity servers. Thus, future mobile-device-based approaches favor achieving the same protection effect with reduced computation and limited local information. In addition, mobile devices are battery-powered, so energy efficiency is crucial to the success of their privacy protection solutions. Therefore, we may witness hybrid approaches that combine the benefits of both: storage and preprocessing of a large amount of public data or crude user data is distributed across untrusted yet unrelated servers, and highly private information is distilled locally on mobile devices.

Compatibility

An important question that must be addressed is how to design protection approaches that are compatible with the existing infrastructure and popular services in the LBS industry. While most existing work focuses on how to minimize the sizes of cloaking boxes, the relation between the box sizes and the quality of LBSs remains unclear. The practicality of integrating the proposed approaches into existing popular LBS use cases should thus be studied in the future.

Balancing Conflicting Interests

Today, large advertisement and analytic firms have been harvesting users' private information — including, but not limited to, location information — without proper notification or permission. The tight connection between these firms and LBS providers, as well as the strong economic incentives behind this connection, induces LBS providers to abuse users' information. While the protection schemes discussed in this article can effectively serve as the first line of defense, achieving balance among the interests of mobile users, LBS providers, and advertising

While most existing work focuses on how to minimize the sizes of cloaking boxes, the relation between the box sizes and the quality of LBSs remains unclear. The practicality of integrating the proposed approaches into existing popular LBS use cases should thus be studied in the future.

While the protection schemes discussed in this article can effectively serve as the first line of defense, achieving balance among the interests of mobile users, LBS providers and Ads firms is vital to the prosperity of LBS eco-system.

firms is vital to the prosperity of the LBS ecosystem. It requires significant effort and cooperation among experts from computer science and economics, and government regulation agencies. Research is necessary in this important direction.

Conclusion

The threat posed by the emerging LBSs to users' privacy has received considerable attention from consumers, service providers, and government administrations, and has also drawn significant interest from both academia and industry. In this article, we have discussed the common threat models used in LBS privacy protection, summarized the existing privacy metrics, and also presented privacy protection solutions with a focus on location perturbation and obfuscation schemes. We have also suggested future research directions.

Acknowledgment

The work reported in this article is supported in part by the National Science Foundation under Grant CNS-1138200. The authors would like to express their thanks to the anonymous reviewers for their valuable comments.

References

[1] B. Harrison and A. Dey, "What Have You Done with Location-Based Services Lately?," *IEEE Pervasive Computing*, vol. 8, Oct.–Dec. 2009, pp. 66–70.
[2] B. Hoh *et al.*, "Enhancing Security and Privacy in Traffic-Monitoring Systems," *IEEE Pervasive Computing*, vol. 5, no. 4, 2006, pp. 38–46.
[3] J. Krumm, "Inference Attacks on Location Tracks," *PERVASIVE'07, Proc. 5th Int'l. Conf. Pervasive Computing*, Springer-Verlag, 2007, pp. 301–09.
[4] Y. Matsuo *et al.*, "Inferring Long-Term User Properties Based on Users Location History," *IJCAI '07 Proc. 20th Int'l. Joint Conf. Artificial intelligence*, Morgan Kaufmann Publishers Inc., 2007.
[5] M. Gruteser and B. Hoh, "On the Anonymity of Periodic Location Samples," *Security in Pervasive Computing*, vol. 3450/2005, Mar. 2005, pp. 179–92.
[6] B. Hoh and M. Gruteser, "Protecting Location Privacy Through Path Confusion," *IEEE SecureComm*, 2005, pp. 194–205.
[7] D. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Trans. Automatic Control*, vol. 24, Dec. 1979, pp. 843–54.
[8] Microsoft, "Location Based Services Usage and Perceptions Survey," http://www.microsoft.com/downloads/en/details.aspx?FamilyID=0e52758c-3ab8-49b6-9d84-20cc53c2c308, Retrieved Apr. 2011.
[9] C. Kang, "Obama Administration Calls for 'Privacy Bill of Rights'," http://voices.washingtonpost.com/posttech/2010/12/obama_administration_calls_for.html, Retrieved Apr. 2011.
[10] A. Dey *et al.*, "Location-Based Services," *IEEE Pervasive Computing*, vol. 9, Jan.–Mar. 2010, pp. 11–12.
[11] R. Want *et al.* "The Active Badge Location Systems," *ACM Trans. Information Systems*, vol. 10, Jan. 1992.
[12] P. Bahl and V. Padmanabhan, "Radar: An In-Building RF-Based User Location And Tracking System," *IEEE INFOCOM*, Mar. 2000, pp. 775–84.
[13] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-Support System," *MOBICOM*, Aug. 2000.
[14] A. LaMarca *et al.*, "Place lab: Device Positioning Using Radio Beacons in the Wild," *PERVASIVE 2005, LNCS*, vol. 3468, 2005, pp. 116–33.
[15] foursquare, http://foursquare.com/, retrieved Apr. 2011.
[16] AT&T, "Shopalerts." http://shopalerts.att.com/sho/att/index.html, Retrieved Apr. 2011.
[17] General Motors, "Onstar." http://www.onstar.com/web/portal/home, Retrieved Apr. 2011.
[18] A. Bose, X. Hu, K. Shin, and T. Park, "Behavioral Detection of Malware on Mobile Handsets," *MobiSys, ACM*, 2008, pp. 225–38.
[19] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking," *MobiSys, ACM*, 2003.
[20] M. F. Mokbel, C.-Y. Chow, and W. G. Aref, "The New Casper: Query Processing for Location Services Without Compromising Privacy," *VLDB, ACM*, Sept. 2006, pp. 763–74.
[21] B. Gedik and L. Liu, "Protecting Location Privacy With Personalized *k*-Anonymity: Architecture and Algorithms," *IEEE Trans. Mobile Computing*, vol. 7, Jan. 2008, pp. 1–18.
[22] A. Pingley *et al.*, "Protection of Query Privacy for Continuous Location Based Services," *IEEE INFOCOM'11*, Apr. 2011.
[23] Z. Chen, "Energy-Efficient Information Collection and Dissemination in Wireless Sensor Networks," Ph.D. Thesis, University of Michigan, 2009.
[24] J. Meyerowitz and R. R. Choudhury, "Hiding Stars with Fireworks: Location Privacy Through Camouflage," *MobiCom, ACM*, 2009.
[25] A. Pfitzmann and M. Hansen, "A Terminology for Talking About Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management (v0.34)," http://dud.inf.tu-dresden.de/Anon_Terminology.shtml, Aug. 2010.
[26] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," *Int'l. J. Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, 2002, pp. 557–70.
[27] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical J.*, vol. 27, July, Oct. 1948, pp. 379–423, 623–56.
[28] A. Beresford and F. Stajano, "Location Privacy in Pervasive Computing," *IEEE Pervasive Computing*, vol. 2, no. 1, 2003, pp. 46–55.
[29] B. Hoh *et al.*, "Achieving Guaranteed Anonymity in GPS Traces Via Uncertainty-Aware Path Cloaking," *IEEE Trans.Mobile Computing*, vol. 9, pp. 1089–107, Aug. 2010.
[30] T. Xu and Y. Cai, "Feeling-Based Location Privacy Protection for Location-Based Services," *CCS, ACM*, Nov. 2009, pp. 348–57.
[31] H. Kido, Y. Yanagisawa, and T. Satoh, "An Anonymous Communication Technique Using Dummies for Location-based Services," *IEEE Proc. Int'l. Conf. Pervasive Services*, ICPS '05, July 2005.
[32] R. Shokri *et al.*, "Quantifying Location Privacy," *IEEE Symp. Security and Privacy*, May 2011.
[33] B. Chor *et al.*, "Private Information Retrieval," *IEEE Symp. Foundations of Comp. Sci.*, 1995, pp. 41–50.
[34] E. Kushilevitz and R. Ostrovsky, "Replication is Not Needed: Single Database, Computationally-Private Information Retrieval," *IEEE Symp. Foundations of Comp. Sci.*, 1997, pp. 364–73.
[35] G. Ghinita *et al.*, "Private Queries in Location Based Services: Anonymizers Are Not Necessary," *SIGMOD, ACM*, 2008, pp. 121–32.
[36] W3C, "Platform for Privacy Preferences (P3P) Project." http://www.w3.org/P3P, retrieved Apr. 2011.
[37] IETF, "Geographic Location/Privacy Working Group." http://datatracker.ietf.org/wg/geopriv/charter/, retrieved April 2011.
[38] W. G. Aref and H. Samet, "Efficient Processing of Window Queries in the Pyramid Data Structure," *PODS '90, Proc. 9th ACM SIGACT-SIGMOD-SIGART Symp. Principles of Database Sys.*, 1991.
[39] T. Xu and Y. Cai, "Exploring Historical Location Data for Anonymity Preservation in Location-Based Services," *IEEE INFOCOM '08*, Apr. 2008, pp. 547–55.
[40] A. Deutsch *et al.*, "Policy-Aware Sender Anonymity in Location Based Services," *IEEE ICDE*, Mar. 2010, pp. 133–44.
[41] R. Finkel and J. L. Bentley, "Quad Trees: A Data Structure for Retrieval on Composite Keys," *Acta Informatica*, vol. 4, no. 1, 1974, pp. 1–9.
[42] A. Pingley *et al.*, "Cap: A context-Aware Privacy Protection System for Location-Based Services," *IEEE ICDCS*, June 2009, pp. 49–57.
[43] J. Manweiler, R. Scudellari, and L. P. Cox, "Smile: Encounter-Based Trust for Mobile Social Services," *CCS '09, ACM*, 2009, pp. 246–55.
[44] H. Hu and J. Xu, "Non-Exposure Location Anonymity," *IEEE ICDE*, 2009, pp. 1120–31.

[45] S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest Neighbor Search with Strong Location Privacy," *VLDB*, Sept. 2010.

BIOGRAPHIES

KANG G. SHIN [F] (kgshin@eecs.umich.edu) is the Kevin & Nancy O'Connor Professor of Computer Science in the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. He has supervised the completion of 70 Ph.D.s, and authored/coauthored more than 770 technical articles (more than 270 of these are in archival journals), one textbook, and more than 20 patents or invention disclosures, and received numerous best paper awards, including the Best Paper Awards from the 2011 ACM International Conference on Mobile Computing and Networking (MobiCom '11), the 2011 IEEE International Conference on Autonomic Computing, the 2010 and 2000 USENIX Annual Technical Conferences, as well as the 2003 IEEE Communications Society William R. Bennett Prize Paper Award and the 1987 Outstanding *IEEE Transactions on Automatic Control* Paper Award. He has also received several institutional awards, including the Research Excellence Award in 1989, Outstanding Achievement Award in 1999, Distinguished Faculty Achievement Award in 2001, and Stephen Attwood Award in 2004 from the University of Michigan (the highest honor bestowed to Michigan Engineering faculty); a Distinguished Alumni Award of the College of Engineering, Seoul National University, in 2002; 2003 IEEE RTC Technical Achievement Award; and 2006 Ho-Am Prize in Engineering (the highest honor bestowed to Korean-origin engineers). He has chaired several major conferences, including 2009 ACM MobiCom, 2008 IEEE SECON, 2005 ACM/USENIX MobiSys, 2000 IEEE RTAS, and 1987 IEEE RTSS. He is a Fellow of ACM, and HAS served on editorial boards, including *IEEE TPDS* and *ACM Transactions on Embedded Systems*. He has also served or is serving on numerous government committees, such as the U.S. NSF Cyber-Physical Systems Executive Committee and the Korean Government R&D Strategy Advisory Committee. He has also co-founded a couple of startups.

XIAOEN JU is a Ph.D. student in the Electrical Engineering and Computer Science Department at the University of Michigan. He received his B.E. and M.E. degrees from Shanghai Jiao Tong University in 2006 and 2009, respectively, and his engineer's degree from Ecole Centrale Paris in 2009 (Promo 2007). His current research focuses on operating systems, virtual machines, system security, and privacy issues in location-based services.

ZHIGANG CHEN received his Ph.D. degree in computer science from the University of Michigan, Ann Arbor in 2009. Before that, he received his B.S. from Wuhan University and M.S from Peking University in China. His research interests at University of Michigan was focused on distributed embedded system, including energy-efficiency in wireless sensor networks and location privacy in smartphones. He is currently working in mobile Internet industry, affiliated with Moovweb Corp.

XIN HU received his Ph.D. from the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, in 2011 and then joined the IBM T. J. Watson Research Center as a member of technical staff. He received his B.S. degree from Zhejiang University, China in 2005 and M.S degree from the University of Michigan in 2007, respectively. His research interests lie primarily in the area of network and system security. His current research focuses on monitoring and detecting botnets, reverse engineering malware, and developing systems to facilitate malware analysis and automatic signature generation. The goal of his research is to discover the underlying principles of real-world security problems, and design effective and scalable solutions to enhance the security, availability, and integrity of computer systems.