Trajectory Privacy in Location-based Services and Data Publication

Chi-Yin Chow
Department of Computer Science
City University of Hong Kong
Kowloon, Hong Kong
chiychow@cityu.edu.hk

Mohamed F. Mokbel
Department of Computer Science and Engineering
University of Minnesota
Minneapolis, MN, USA
mokbel@cs.umn.edu

ABSTRACT

The ubiquity of mobile devices with global positioning functionality (e.g., GPS and AGPS) and Internet connectivity (e.g., 3G and Wi-Fi) has resulted in widespread development of location-based services (LBS). Typical examples of LBS include local business search, e-marketing, social networking, and automotive traffic monitoring. Although LBS provide valuable services for mobile users, revealing their private locations to potentially untrusted LBS service providers pose privacy concerns. In general, there are two types of LBS, namely, snapshot and continuous LBS. For snapshot LBS, a mobile user only needs to report its current location to a service provider once to get its desired information. On the other hand, a mobile user has to report its location to a service provider in a periodic or on-demand manner to obtain its desired continuous LBS. Protecting user location privacy for continuous LBS is more challenging than snapshot LBS because adversaries may use the spatial and temporal correlations in the user's location samples to infer the user's location information with higher certainty. Such user location trajectories are also very important for many applications, e.g., business analysis, city planning, and intelligent transportation. However, publishing such location trajectories to the public or a third party for data analysis could pose serious privacy concerns. Privacy protection in continuous LBS and trajectory data publication has increasingly drawn attention from the research community and industry. In this survey, we give an overview of the state-of-the-art privacy-preserving techniques in these two problems.

1. INTRODUCTION

With the advanced location-detection technologies, e.g., global positioning system (GPS), cellular networks, Wi-Fi, and radio frequency identification (RFID), location-based services (LBS) have become ubiquitous [6; 30; 41]. Examples of LBS include local business search (e.g., searching for restaurants within a user-specified range distance from a user), e-marketing (e.g., sending e-coupons to nearby potential customers), social networking (e.g., a group of friends sharing their geo-tagged messages), automotive traffic monitoring (e.g., inferring traffic congestion from position and speed information periodically reported from probe vehicles), and route finder applications (e.g., finding a route with the shortest driving time between two locations). There are two types of LBS, namely, *snapshot* and *continuous* LBS. For snapshot LBS, a mobile user only needs to report its current location to a service provider once to get its desired information. On the other hand, a mobile user has to report its location to a service provider in a periodic or on-demand manner to obtain its desired continuous LBS.

Although LBS provide many valuable and important services for end users, revealing personal location data to potentially untrustworthy service providers could pose privacy concerns. Two surveys reported in July 2010 found that more than half (55%) of LBS users show concern about their loss of location privacy [54] and 50% of U.S. residents who have a profile on a social networking site are concerned about their privacy [39]. The results of these surveys confirm that location privacy is one of the key obstacles for the success of location-dependent services. In fact, there are many real-life scenarios where perpetrators abuse location-detection technologies to gain access to private location information about victims [14; 16; 51; 52].

Privacy in continuous LBS is more challenging than snapshot LBS because adversaries could use the spatial and temporal correlations in the user's location samples to infer the user's location information. Such user location trajectories are also very important for many real-life applications, e.g., business analysis, city planning, and intelligent transportation. However, publishing such location trajectories to the public or a third party for data analysis could pose serious privacy concerns. Privacy protection in continuous LBS and trajectory data publication has increasingly drawn attention from the industry and academia. In this survey, we give an overview of the existing techniques in these two problems.

The rest of this paper is organized as follows. Section 2 presents the derivation of location trajectory privacy. Section 3 discusses the state-of-the-art privacy-preserving techniques in continuous LBS. Section 4 gives existing privacy protection techniques for user location trajectory publication. Finally, Section 5 concludes this survey with research directions in privacy-preserving continuous LBS and trajectory data publication.

2. THE DERIVATION OF LOCATION TRAJECTORY PRIVACY

This section gives the derivation of location trajectory privacy from data privacy and location privacy.

2.1 Data Privacy

Many agencies and other organizations often need to publish microdata, i.e., tables that contain unaggregated information about individuals, (e.g., medical, voter registration, census, and customer data) for many practical purposes such as demographic and public health research. In general, microdata is stored in a table where each row corresponds to one individual. In order to avoid the identification of records in microdata, known identifiers (e.g., name and social security number) must be removed. However, joining such "de-identified" microdata with other released microdata may still pose *data privacy* issues for individuals [48]. A study estimated that 87% of the population of the United States can be uniquely identified using the collection of non-identity attributes, i.e., gender, date of birth, and 5-digit zip code [50]. In fact, those three attributes were used to link Massachusetts, USA voter registration records including name, gender, zip code and date of birth to "de-identified" medical data from Group Insurance Company including gender, zip code, date of birth and diagnosis to identify the medical records of the governor of Massachusetts in the medical data [50]. Terminologically, attributes whose values taken together can potentially identify an individual record are referred to as "quasi-identifiers" and a set of records that have the same values for the quasi-identifiers in a released microdata is defined as an "equivalence class".

Data privacy-preserving techniques have been developed to anonymize microdata. Several privacy-preserving properties are proposed to limit disclosure of anonymized microdata. For example, **k-anonymity** requires each record to be indistinguishable with at least other $k-1$ records with respect to the quasi-identifier, i.e., each equivalence class contains at least k records [35; 48; 50; 49]. However, a k-anonymized equivalence class suffers from a homogeneity attack if all records in the class have less than k values for the sensitive attribute (e.g., disease and salary). To this end, **l-diversity** property is proposed to ensure that an equivalence class must have at least l values for the sensitive attribute [38; 55]. To further strengthen data privacy protection, **t-closeness** principle is defined that an equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the entire data set is no more than a threshold parameter t [36]. For the details of these and other data privacy principles for data publishing, we refer the reader to the recent survey paper [19].

2.2 Location Privacy

In LBS, mobile users issue location-based queries to LBS service providers to obtain information based on their physical locations. LBS pose new challenges to traditional data privacy-preserving techniques due to two main reasons [40]. (1) These techniques preserve data privacy, but not the location-based queries issued by mobile users. (2) They ensure desired privacy guarantees for a snapshot of the database. In LBS, queries and data are continuously updated with high rates. Such highly dynamic behaviors need continuous maintenance of anonymized user and object sets.

Privacy-preserving techniques for LBS can be classified into three categories: (1) **False locations** [28; 33; 58]**.** The basic idea is to send either one or more fake locations that are related to the user location. (2) **Space transformation** [21; 32]**.** The techniques in this category transform the location information into another space where the spatial relationships among queries and data are encoded. (3) **Spatial cloaking** [2; 5; 7; 11; 12; 15; 20; 22; 23; 25; 31; 40; 60]**.** The main idea is to blur users' locations into cloaked spatial regions that are guaranteed to satisfy the k-anonymity [50] (i.e., the cloaked spatial region contains at least k users) and/or minimum region area privacy requirements [5; 15; 40] (i.e., the spatial region size is larger than a threshold). Spatial cloaking techniques have been extended to support road networks where a user's location is cloaked into a set of connected road segments so that the cloaked road segment set satisfies the privacy requirements of k-anonymity and/or minimum total road segment length [10; 34; 42; 53].

Research efforts have also dedicated to dealing with privacy-preserving location-based queries, i.e., getting anonymous services from LBS service providers (e.g., [5; 21; 29; 31; 32; 40; 58]). These query processing frameworks can be divided into three main categories. (1) **Location obstruction** [58]**.** The basic idea is that a querying user first sends a query along with a false location as an anchor to a database server. The database server keeps sending the list of nearest objects to the anchor to the user until the list of received objects satisfies the user's privacy and quality requirements. (2) **Space transformation** [21; 32]**.** This approach converts the original location of data and queries into another space through a trusted third party. The space transformation maintains the spatial relationship among the data and query, in order to provide accurate query answers. (3) **Cloaked query area processing** [5; 9; 13; 29; 31; 40]**.** In this framework, a privacy-aware query processor is embedded in the database server to deal with the cloaked spatial area received either from a querying user [5; 29] or from a trusted third party [9; 31; 40]. For spatial cloaking in road networks, an efficient and query-aware algorithm is proposed to process privacy-aware location-based queries [3].

2.3 Trajectory Privacy

A location trajectory is a moving path or trace reported by a moving object in the geographical space. A location trajectory Tr is represented by a set of n time-ordered points, $Tr : p_1 \rightarrow p_2 \rightarrow \ldots \rightarrow p_n$, where each point p_i consists of a geospatial coordinate set (x_i, y_i) (which can be determined by a GPS-like device) and a timestamp t_i, i.e., $p_i = (x_i, y_i, t_i)$, where $1 \leq i \leq n$. Such spatial and temporal attributes of a location trajectory can be considered as powerful quasi-identifiers that can be linked to various other kinds of physical data objects [19; 43]. For example, a hospital releases a trajectory data set of its patients to a third-party research institute for analysis, as shown in Table 1. The released trajectory data set does not contain any explicit identifiers, such as patient name, but it contains a sensitive attribute (i.e., disease). Each record with a unique random ID, RID, corresponds to an individual, e.g., the record with $RID = 1$ means a patient visited locations $(1, 5)$, $(6, 7)$, $(8, 10)$, and $(11, 8)$ at timestamps 2, 4, 5, and 8, respectively. Suppose that an adversary knows that a patient of the hospital, Alice, visited locations $(1, 5)$ and $(8, 10)$ at timestamps 2 and 8, respectively. Since only the trajectory record with $RID = 1$ satisfies such spatial and temporal attributes, the adversary can infer that Alice has HIV with 100% confidence. This example shows that publishing "de-identified" trajectory data can still cause serious privacy threats.

Table 1: Patient trajectory data.

RID	Trajectory	Disease	...
1	$(1,5,2) \rightarrow (6,7,4) \rightarrow (8,10,5) \rightarrow (11,8,8)$	HIV	...
2	$(5,6,1) \rightarrow (3,7,2) \rightarrow (1,5,6) \rightarrow (7,8,7) \rightarrow (1,11,8) \rightarrow (6,5,10)$	Flu	...
3	$(4,7,2) \rightarrow (4,6,3) \rightarrow (5,1,6) \rightarrow (11,8,8) \rightarrow (5,8,9)$	Flu	...
4	$(10,3,5) \rightarrow (7,3,7) \rightarrow (4,6,10)$	HIV	...
5	$(7,6,3) \rightarrow (6,7,4) \rightarrow (6,10,6) \rightarrow (4,6,9)$	Fever	...

In LBS, when a mobile user issues a continuous location-based query to a database server (e.g., "continuously send me the traffic condition within 1 mile from my vehicle"), the user has to report his/her new location to the database server in a periodic or on-demand manner. Similarly, intelligent transportation systems require their users (e.g., probe vehicles) to periodically report their location and speed information to the system for analysis. Although such location-based queries and reports can be made anonymous by replacing the identifiers of users with random identifiers, in order to achieve pseudonymity [46], the users may still suffer from privacy threats. This is because movements of whereabouts of users in public spaces can be openly observed by others through chance or engineered meetings [37]. In the worst case, if the starting location point of a trajectory is home, an adversary uses reverse geocoding[1] [24] to translate a location point into a home address, and then uses a people-search-by-address engine (e.g., `http://www.intelius.com` and `http://www.peoplefinders.com`) to find the residents of the home address. Even though users generate a random identity for each of their location samples, multi-target tracking techniques (e.g., the multiple hypothesis tracking algorithm [47]) can be used to link anonymous location samples to construct target trajectories [26]. To this end, new techniques are developed to protect user location trajectory. The key difference between continuous LBS and trajectory data publication with respect to challenges in privacy protection is twofold: (1) The scalability requirement of the privacy-preserving techniques for continuous LBS is much more important than that for trajectory data publication. This is because continuous LBS require the anonymization module to deal with a large number of real-time location updates at high rates while the anonymization process for trajectory data publication can be performed offline. (2) Global optimization can be applied to trajectory data publication because the anonymization process is able to analyze the entire (static) trajectory data to optimize its privacy protection or usability. However, global optimization is very difficult for continuous LBS, due to highly dynamic, uncertain user movements. Sections 3 and 4 present the state-of-the-art privacy-preserving techniques for continuous LBS and trajectory publication, respectively.

3. PROTECTING TRAJECTORY PRIVACY IN LOCATION-BASED SERVICES

In general, there are two categories of LBS based on whether they need consistent user identities. A consistent user identity is not necessarily a user's actual identity or name because it can be an internal pseudonym. **Category-I LBS:** Some LBS require consistent user identities. For example, "Q1: let me find out where my friends are if they are within 2km from my location", "Q2: recommend 10 nearby restaurants to me based on my profile", and "Q3: continuously tell me the nearest shopping mall to my location". Q1 and Q2 require consistent user identities to let applications to find out their friends and profiles. Although Q3 does not need any consistent user identity, the query with its parameters can be considered as a virtual user identity that remains active until the query expires. **Category-II LBS:** Other LBS do not require consistent user identities, or even any user identities, such as "Q4: send e-coupons to users within 1km from my coffee shop". In this section, we discuss five privacy-preserving techniques for continuous LBS, namely, spatial cloaking, mix-zones, vehicular mix-zones, path confusion and dummies, and indicate whether each of them supports Category-I and/or II LBS from Sections 3.1 to 3.5, as summarized in Table 2.

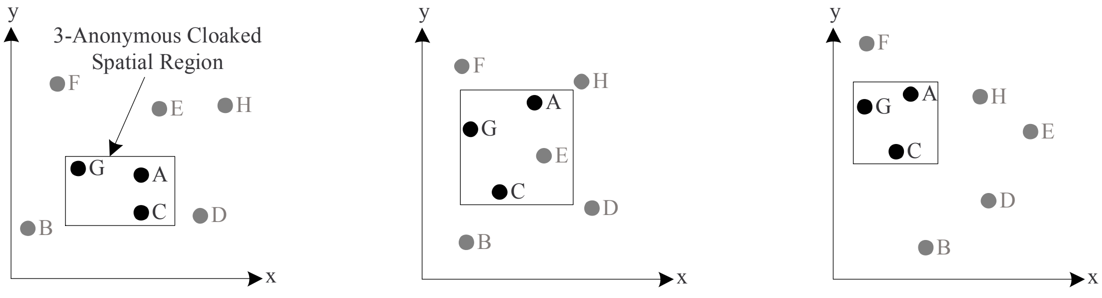
Table 2: Privacy-preserving techniques for continuous LBS.

Techniques	Category-I LBS	Category-II LBS
Spatial cloaking	✓	✓
Mix-zones	×	✓
Vehicular mix-zones	×	✓
Path confusion	×	✓
Dummies	✓	✓

3.1 Spatial Cloaking

Mobile users have to reveal their locations to database servers in a periodic or on-demand manner to obtain continuous LBS. Simply applying a snapshot spatial cloaking technique (e.g., [2; 5; 15; 20; 22; 23; 25; 31; 40; 60]) to each user location independently cannot ensure k-anonymity for a user location trajectory. Thus, new spatial cloaking techniques based on either *real-time* or *historical* user trajectories are designed to protect user location trajectories. Similar to snapshot spatial cloaking techniques, a fully-trusted third party, usually termed *location anonymizer*, is placed between mobile users and database servers. The location anonymizer is responsible for collecting users' locations and blurring their locations into cloaked spatial regions that satisfy the user-specified k-anonymity level and/or minimum spatial region area. Since spatial cloaking techniques do not change user identities, they can support both Category I and II LBS. In the following sections, we will discuss three main kinds of spatial cloaking techniques over user trajectories, namely, group-based, distortion-based, and prediction-based approaches, from Sections *3.1.1* to *3.1.3*. The first two approaches are designed for real-time user trajectories, while the last one is for historical trajectory data.

[1] Reverse geocoding is the process of translating a human-readable address, such as a street address, from geographic coordinates.



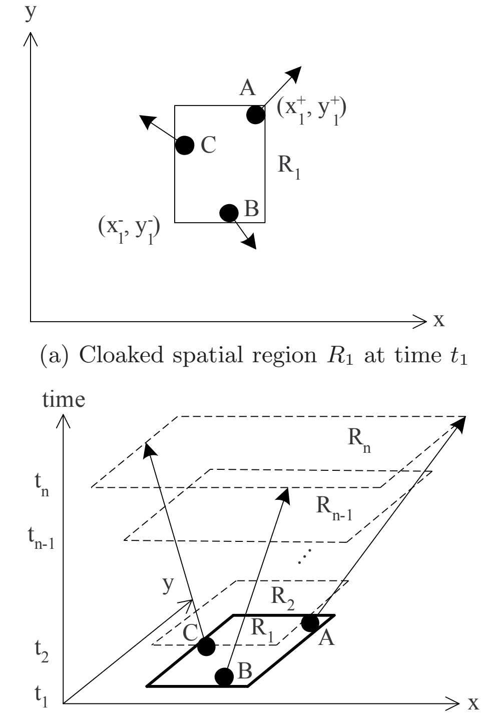
(a) At time t_1 (b) At time t_2 (c) At time t_3

Figure 1: Group-based spatial cloaking over real-time location trajectory data.

3.1.1 Group-based Approach for Real-time Trajectory Data

The group-based algorithm is proposed to use real-time location trajectory data to protect trajectory privacy for continuous location-based queries [8]. The basic idea is that a querying user u forms a group with other $k-1$ nearby peers. Before the algorithm issues u's location-based query or reports u's location to the database server, it blurs u's location into a spatial area that contains all the group members as a cloaked spatial area. Figure 1 depicts an example of continuous spatial cloaking over real-time user location trajectories. In this example, user A that issues a continuous location-based query at time t_i requires its location to be k-anonymized, where $k = 3$. At time t_1, a location anonymizer forms a group of users A, C, and G, so that A's cloaked spatial region contains all these group members, as represented by a rectangle in Figure 1a. The location anonymizer sends A's query with its cloaked spatial region to a database server. At later times t_2 and t_3, when A reports its new location to the location anonymizer, a new cloaked spatial region that contains the group members is formed, as shown in Figures 1b and 1c. The drawbacks of this approach are that users not issuing any query have to report their locations to the location anonymizer and the cloaked spatial area would become very large after a long time period. Such a large cloaked spatial area may incur high computational overhead at the database server and results in many candidate answer objects returned from the database server to the location anonymizer.

In theory, let R_i be the cloaked spatial region for a querying user u at time t_i and $S(R_i)$ be a set of users located in R_i. Suppose u's query is first successfully cloaked at time t_1, it expires at time t_n, $u \in S(R_1)$ and $|S(R_1)| \geq k$. Without any additional information, the value of R_1's entropy, $H(R_1)$, is at least $\log_2 |S(R_1)|$ which means that every user in R_1 has an equal chance of $1/|S(R_1)|$ to be u [56], i.e., R_1 is a k-anonymous region for u. For u's cloaked spatial regions R_{i-1} and R_i generated at two consecutive times t_{i-1} and t_i ($1 < i \leq n$), respectively, if R_{i-1} is a k-anonymous region and $S(R_{i-1}) \subseteq S(R_i)$, R_i is also a k-anonymous region [56]. Thus, the group-based approach can ensure k-anonymity for the entire life span of a continuous location-based query.



(a) Cloaked spatial region R_1 at time t_1

(b) Cloaked spatial regions $R_2, \ldots, R_n$ at times $t_2, \ldots, t_n$, respectively

Figure 2: Query distortion for continuous spatial cloaking.

3.1.2 Distortion-based Approach for Real-time Trajectory Data

The distortion-based approach aims to overcome the drawbacks of the group-based approach. It not only requires querying users to report their locations to the location anonymizer, but it also considers their movement directions and velocities to minimize cloaked spatial regions [45]. A distortion function is defined to measure the temporal query distortion of a cluster of continuous queries. Figure 2 gives an example of how to determine query distortion. In this example, three users A, B and C that issue their continuous location-based queries at time t_1 constitute a cloaking set and their queries expire at time t_n. Their cloaked spatial region R_1 at time t_1 is a minimum bounding rectangle of the cloaking set, as represented by a rectangle (Figure 2a). Let (x_i^-, y_i^-) and (x_i^+, y_i^+) be the left-bottom and right-top vertices of a cloaked spatial region R_i at time t_i, respectively. The distortion for their queries with a cloaked spatial region R_i at time t_i is defined as:

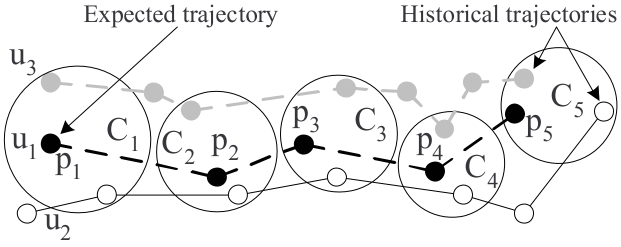


Figure 3: Continuous spatial cloaking over historical trajectories.

$$\Delta(R_i) = \frac{(x_i^+ - x_i^-) + (y_i^+ - y_i^-)}{A_{height} + A_{width}}, \qquad (1)$$

where A_{height} and A_{width} are the height and width of the minimum bounding rectangle of the entire system space, respectively. Based on their movement directions and velocities (represented by arrows in Figure 2b), their subsequent cloaked spatial regions $R_2, R_3, \ldots, R_n$ at times $t_2, t_3, \ldots, t_n$ can be predicted, respectively. The distortion for their queries with respect to the time period from t_1 to t_n is defined as:

$$\int_{t_1}^{t_n} \Delta(R_i) = \frac{1}{P}\left\{\int_{t_1}^{t_2}\Delta(R_1)dt + \int_{t_2}^{t_3}\Delta(R_2)dt + \ldots + \int_{t_{n-1}}^{t_n}\Delta(R_n)dt\right\}, \qquad (2)$$

where $P = A_{height} + A_{width}$. Given a new query Q, greedy cloaking and bottom-up cloaking algorithms are designed to cluster Q with other $k-1$ outstanding queries into a group such that the group satisfies k-anonymity and their query distortion is minimized.

3.1.3 Predication-based Approach for Historical Trajectory Data

Another way to ensure k-anonymity is to use individuals' historical footprints, instead of their real-time locations [57]. A footprint is defined as a user's location collected at some point of time. Similar to the previous two approaches, a fully-trusted location anonymizer is placed between users and LBS service providers to collect users' footprints. Given a user's predicted trajectory (i.e., a sequence of expected footprints), the location anonymizer cloaks it with $k-1$ historical trajectories collected from other users. Figure 3 gives an example for continuous spatial cloaking over historical trajectories, where a user u_1 wants to subscribe continuous LBS from a service provider. u_1's predicted time-ordered footprints are represented by black circles. If u's desired anonymity level is $k = 3$, the location anonymizer finds historical trajectories from two users, u_2 and u_3. Then, each u's expected footprint p_i $(1 \leq i \leq 5)$ is cloaked with at least one unique footprint of each of u_2's and u_3's trajectories to form a cloaked spatial region C_i. The sequence of such cloaked spatial regions constitute the k-anonymized trajectory for u_1.

Given a k-anonymized trajectory $T = \{C_1, C_2, \ldots, C_n\}$, its resolution is defined as:

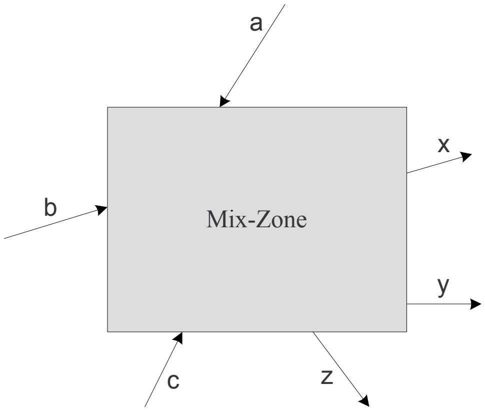


Figure 4: A mix-zone with three users.

$$|T| = \frac{\sum_{i=1}^{n} Area(C_i)}{n}, \qquad (3)$$

where $Area(C_i)$ is the area of cloaked spatial region C_i. For quality of services, $|T|$ should be minimized. Since the computation of an optimal T would be expensive, heuristic approaches are designed to find T. Although using historical trajectory data gives better resolutions for k-anonymized trajectories, it would suffer from an observation attack. This is because an attacker may only see a querying user or less than k users located in a cloaked spatial region at its associated timestamp.

3.2 Mix-Zones

The concept of "mix" has been applied to anonymous communication in a network. A mix-network consists of normal message routers and mix-routers. The basic idea is that a mix-router collects k equal-length packets as input and reorders them randomly before forwarding them, thus ensuring unlinkability between incoming and outgoing messages. This concept has been extended to LBS, namely, *mix-zones* [4]. When users enter a mix-zone, they change to a new, unused pseudonym. In addition, they do not send their location information to any location-based application when they are in the mix-zone. When an adversary that sees a user u exits from the mix-zone cannot distinguish u from any other user who was in the mix-zone with u at the same time. The adversary is also unable to link people entering the mix-zone with those coming out of it. A set of users S is said to be k-anonymized in a mix-zone Z if all following conditions are met [44]:

1. The user set S contains at least k users, i.e., $|S| \geq k$.
2. All users in S are in Z at a point in time, i.e., all users in S must enter Z before any user in S exits.
3. Each user in S spends a completely random duration of time inside Z.
4. The probability of every user in S entering through an entry point is equally likely to exit in any of the exit points.

Table 3 gives an example of 3-anonymity for the mix-zone depicted in Figure 4, where three users with real identities,

Table 3: An example of 3-anonymized mix-zone.

User ID	P_{old}	P_{new}	ts_{enter}	ts_{exit}	t_{inside}
α	a	y	2	9	7
β	c	x	5	8	3
γ	b	z	1	11	10

α, β, and γ enter the mix-zone with old pseudonyms (P_{old}) a, c, and b at timestamps (ts_{enter}) 2, 5, and 1, respectively. Users α, β, and γ exit the mix-zone with new pseudonyms (P_{new}) y, x, and z at timestamps (ts_{exit}) 9, 8, and 11, respectively. Thus, they all are in the mix-zone during the time period from 5 to 8. Since they stay inside the mix-zone with random time periods (i.e., t_{inside}), there is a strong unlinkability between their entry order ($\gamma \rightarrow \alpha \rightarrow \beta$) and exit order ($\beta \rightarrow \alpha \rightarrow \gamma$).

We can see that mix-zones require pseudonym change to protect user location privacy, so this technique can only support Category-II LBS. Mix-zones also impose limits on the services available to mobile users inside a mix-zone because they cannot update their locations until exiting the mix-zone. To minimize disruptions caused to users, the placement of mix-zones in the system should be optimized to limit the total number of mix-zones required to achieve a certain degree of anonymity [18].

3.3 Vehicular Mix-Zones

In a road network, vehicle movements are constrained by many spatial and temporal factors, such as physical roads, directions, speed limits, traffic conditions, and road conditions. Mix-zones designed for the Euclidean space are not secure enough to protect trajectory privacy in road networks [17; 44]. This is because an adversary can gain more background information from physical road constraints and delay characteristics to link entering events and exiting events of a mix-zone with high certainty. For example, a mix-zone (represented by a shaded area) is placed on an intersection of three road segments $Seg1$, $Seg2$, and $Seg3$, as depicted in Figure 5. If u-turn is not allowed in the intersection, an adversary knows that a vehicle with pseudonym c enters the mix-zone from either $Seg1_{in}$ or $Seg2_{in}$. Since a vehicle turning from $Seg1_{in}$ to $Seg3_{out}$ normally takes a longer time than turning from $Seg2_{in}$ to $Seg3_{out}$, the adversary would use this delay characteristic to link an exiting event at $Seg3_{out}$ to an entering event at $Seg1_{in}$ or $Seg2_{in}$. In addition, every vehicle may spend almost the same time during a short time period for a specific direction, e.g., u-turn, left, straight, or right. This temporal characteristic may violate the third necessary condition for mix-zones listed in Section 3.2.

An effective solution for vehicular mix-zones is to construct non-rectangular, adaptive mix-zones that start from the center of an road segment intersection on its outgoing road segments [44], as depicted in Figure 6. The length of each mix-zone on an outgoing segment is determined based on the average speed of the road segment, the time window, and the minimum pairwise entropy threshold. The dark shaded area should also be included in the mix-zone to ensure that an adversary cannot infer the vehicle movement direction (e.g., turn left or go straight in this example). The pairwise entropy is computed for every pair of users a and b in an anonymity set S by considering a and b to be the only members in S and determining the linkability between their old and new pseudonyms. Similar to mix-zones, vehicular mix-zones require a pseudonym change, so they can only support Category-II LBS.

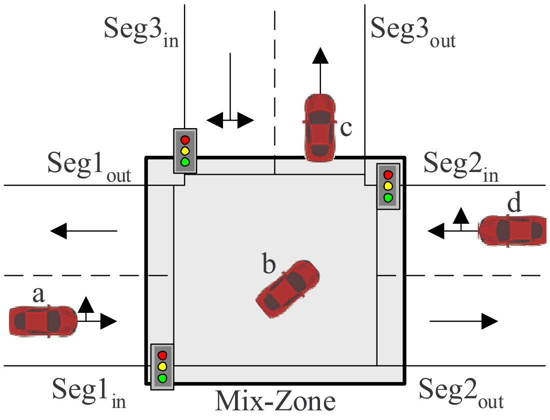


Figure 5: A vehicular mix-zone.

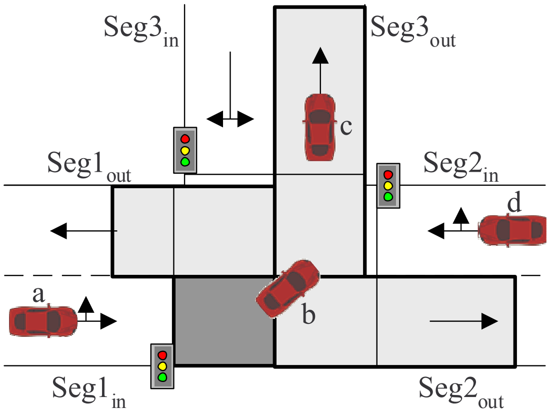


Figure 6: Non-rectangular, adaptive vehicular mix-zones.

3.4 Path Confusion

Since consecutive location samples from a vehicle are temporally and spatially correlated, trajectories of individual vehicles can be constructed from a set of location samples with anonymized pseudonyms reported from several vehicles through target tracking algorithms [26]. The general idea of these algorithms is to predict the position of a target vehicle based on the last known speed and direction information and then decide which next location sample (or the one with the highest probability if there are multiple candidate location samples) to link to the same vehicle through Maximum Likelihood Detection [26].

The main goal of the path confusion technique is to avoid linking consecutive location samples to individual vehicles through target tracking algorithms with high certainty [27]. The degree of privacy of the path confusion technique is defined as the "time-to-confusion", i.e., the tracking time between two location samples where an adversary could not determine the next sample with sufficient tracking certainty. Tracking uncertainty is computed by $H = -\sum p_i \log p_i$, where p_i is the probability that location sample i belongs

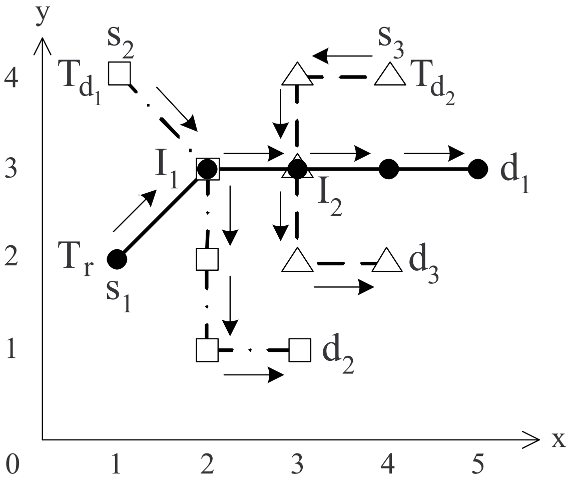


Figure 7: One real trajectory T_r and two dummies T_{d_1} and T_{d_2}.

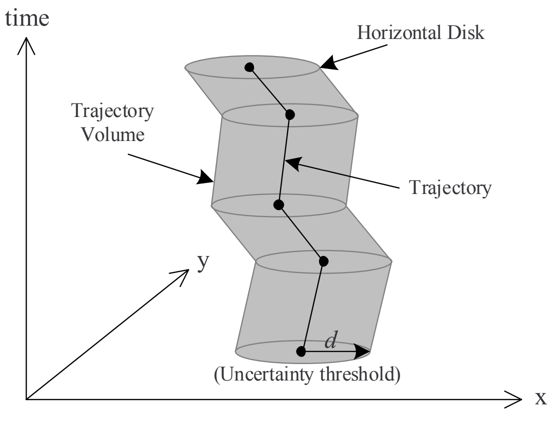


Figure 8: The trajectory uncertainty model.

Table 4: Privacy measures of the example in Figure 7.

Time (i)	1	2	3	4	5
Real trajectory (T_r)	(1,2)	(2,3)	(3,3)	(4,3)	(5,3)
Dummy (T_{d_1})	(1,4)	(2,3)	(2,2)	(2,1)	(3,1)
Dummy (T_{d_2})	(4,4)	(3,4)	(3,3)	(3,2)	(4,2)
$\|S_i\|$	3	2	2	3	3
$\frac{1}{n}\sum_{j=1}^{n} dist(T_r^i, T_{d_j}^i)$	2.80	0.71	0.71	2.12	2.12

to a target vehicle. Smaller values of H means higher certainty or lower privacy. Given a maximum allowable time to confusion, *ConfusionTime*, and an associated uncertainty threshold, *ConfusionLevel*, a vehicle's location sample can be safely revealed if the time between the current time t and the last point of its confusion is less than *ConfusionTime* and tracking uncertainty of its sample with all location samples revealed at time t is higher than *ConfusionLevel*. To reduce computational overhead, the computation of tracking uncertainty can only consider the k-nearest location samples to a predicted location point (calculated by the target tracking algorithm), rather than all location samples reported at time t.

3.5 Dummy Trajectories

Without relying on a trusted third party to perform anonymization, a mobile user can generate fake location trajectories, called *dummies*, to protect trajectory privacy [33; 59]. Given a real user location trajectory T_r and a set of user-generated dummies T_d, the degree of privacy protection for the real trajectory is measured by the following metrics [59]:

1. **Snapshot disclosure (SD).** Let m be the number of location samples in T_r, S_i be the set of location samples in T_r and any T_d at time t_i, and $|S_i|$ be the size of S_i. SD is defined as the average probability of successfully inferring each true location sample in T_r, i.e., $SD = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{|S_i|}$. Figure 7 gives a running example of $n = 3$ trajectories and $m = 5$ where T_r is from location s_1 to location d_1 (i.e., $s_1 \rightarrow d_1$), T_{d_1} is $s_2 \rightarrow d_2$, and T_{d_2} is $s_3 \rightarrow d_3$. There are two intersections I_1 and I_2. At time $i = 1$, since there are three different locations, i.e., $(1, 2)$, $(1, 4)$ and $(4, 4)$, $|S_1| = 3$. Thus, $SD = \frac{1}{5}(\frac{1}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3}) = \frac{2}{5}$.
2. **Trajectory disclosure (TD).** Given n trajectories, where k trajectories have intersection with at least one other trajectory and $n - k$ trajectories do not intersect any other trajectory, let N_k be the number of possible trajectories among the k trajectories. TD is defined as the probability of successfully identifying the true trajectory among all possible trajectories is $\frac{1}{N_k+(n-k)}$. In the running example, $N_k = 3$ and there are eight possible trajectories, i.e., $s_1 \rightarrow I_1 \rightarrow d_2$, $s_1 \rightarrow I_1 \rightarrow I_2 \rightarrow d_1$, $s_1 \rightarrow I_1 \rightarrow I_2 \rightarrow d_3$, $s_2 \rightarrow I_1 \rightarrow d_2$, $s_2 \rightarrow I_1 \rightarrow I_2 \rightarrow d_1$, $s_2 \rightarrow I_1 \rightarrow I_2 \rightarrow d_3$, $s_3 \rightarrow I_2 \rightarrow d_1$, and $s_3 \rightarrow I_2 \rightarrow d_3$. Hence, $TD = \frac{1}{8+(3-3)} = \frac{1}{8}$.
3. **Distance deviation (DD).** DD is defined as the average distance between the i-th location samples of T_r and each T_{d_j}, i.e., $DD = \frac{1}{m}\sum_{i=1}^{m}(\frac{1}{n}\sum_{j=1}^{n} dist(T_r^i, T_{d_j}^i))$, where $dist(p, q)$ denotes the Euclidean distance between two point locations p and q. In the running example, $DD = \frac{1}{5} \times (2.80 + 0.71 + 0.71 + 2.12 + 2.12) = 1.69$.

Given a real trajectory T_r and the three user-specified parameters SD, TD, and DD in a privacy profile, the dummy-based anonymization algorithm incrementally uses DD to find a set of candidate dummies and selects one with the best matching to SD and TD until it finds a set of trajectories (including T_r and selected dummies) that satisfies all the parameters [59]. Since a user can use an consistent identity for its actual trajectory and other dummies, the dummy-based approach can support both Category I and II LBS, as depicted in Table 2.

4. PROTECTING PRIVACY IN TRAJECTORY PUBLICATION

In this section, we discuss anonymization techniques for trajectory data publication. The anonymized trajectory data can be released to the public or third parties for answering

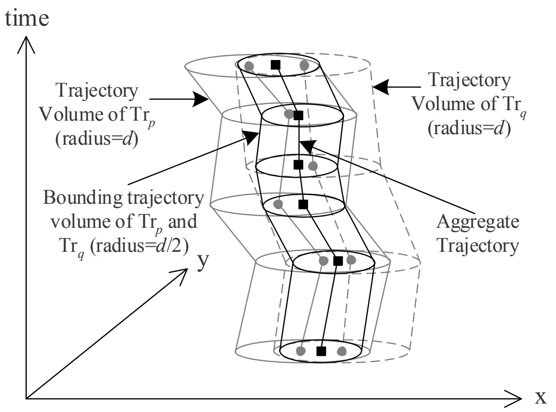

time
Trajectory Volume of Tr_p (radius=d)
Trajectory Volume of Tr_q (radius=d)
Bounding trajectory volume of Tr_p and Tr_q (radius=$d/2$)
Aggregate Trajectory
y
x

Figure 9: 2-anonymized co-localized trajectories.

spatio-temporal range queries [1; 43] and data mining [43]. In the following sections, we present two well-studied trajectory anonymization techniques, namely, clustering-based approach [1] (Section 4.1) and generalization-based approach [43] (Section 4.2).

4.1 Clustering-based Approach

The clustering-based approach [1] utilizes the uncertainty of trajectory data to group k co-localized trajectories within the same time period to form a k-anonymized aggregate trajectory. Given a trajectory Tr between times t_1 and t_n, i.e., $[t_1, t_n]$, and an uncertainty threshold d, each location sample in Tr, $p_i = (x_i, y_i, t_i)$, is modeled by a horizontal disk with radius d centered at (x_i, y_i). The union of all such disks constitute the trajectory volume of Tr, as shown in Figure 8. Two trajectories Tr_p and Tr_q defined in $[t_1, t_n]$ are said to be co-localized with respect to d, if the Euclidean distance between each pair of points in Tr_p and Tr_q at time $t \in [t_1, t_n]$ is less than or equal to d. An anonymity set of k trajectories is defined as a set of at least k co-localized trajectories. The cluster of k co-localized trajectories is then transformed into an aggregate trajectory where each of its location points is computed by the arithmetic mean of the location samples at the same time. Figure 9 gives the trajectory volumes of Tr_p and Tr_q that are represented by grey and dotted lines, respectively. The trajectory volume with black lines is a bounding trajectory volume for Tr_p and Tr_q. The bounding trajectory volume is then transformed into an aggregate trajectory which is represented by a sequence of square markers.

The clustering-based anonymization algorithm consists of three main phases [1]:

1. **Pre-processing phase.** The main task of this phase is to group all trajectories that have the same starting and ending times, i.e., they are in the same equivalence class with respect to time span. To increase the number of trajectories in an equivalence class, given an integer parameter π, all trajectories are trimmed if necessary such that only one timestamp every π can be the starting or ending point of a trajectory.

2. **Clustering phase.** This phase clusters trajectories based on a greedy clustering scheme. For each equivalence class, a set of appropriate pivot trajectories are selected as cluster centers. For each cluster center, its nearest $k-1$ trajectories are assigned to the cluster, such that the radius of the bounding trajectory volume of the cluster is not larger than a certain threshold (e.g., $d/2$).

3. **Space transformation phase.** Each cluster is transformed into a k-anonymized aggregate trajectory by moving its every location sample horizontally to the center of its bonding trajectory volume.

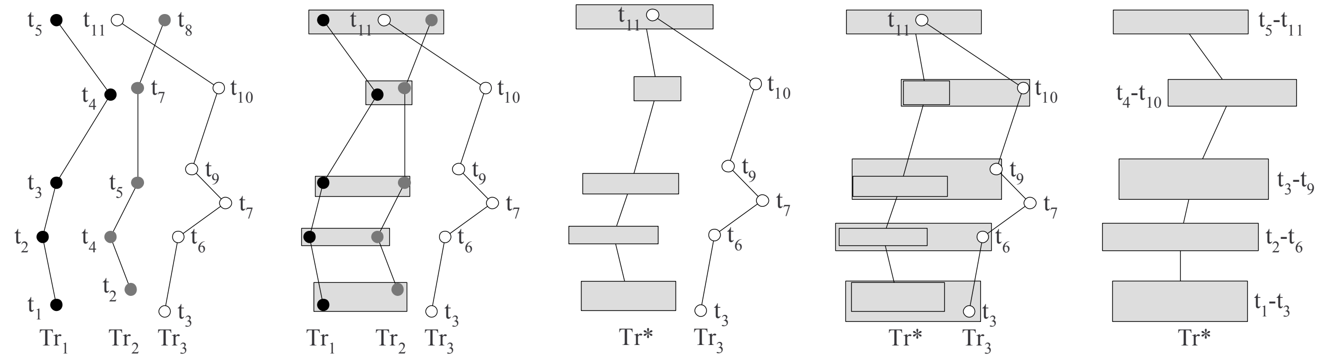
4.2 Generalization-based Approach

Since most data mining and statistical applications work on atomic trajectories, they are needed to be modified to work on aggregate trajectories generated by an anonymization algorithm (e.g., the clustering approach). To address this limitation, the generalization-based algorithm first generalizes a trajectory data set into a set of k-anonymized trajectories, i.e., each one is a sequence of k-anonymized regions. Then, for each k-anonymized trajectory, the algorithm uniformly selects k atomic points from each anonymized region and links a unique atomic point from each anonymized region to reconstruct k trajectories. More details about these two main steps are given below [43]:

1. **Anonymization step.** Given a trajectory data set $\mathcal{T}$, each iteration of this step creates an empty anonymity group G and randomly samples one trajectory $Tr \in \mathcal{T}$. Tr is put into G as the group representative $Rep_G = Tr$. Then, the closest trajectory $Tr' \in \mathcal{T} - G$ to Rep_G is inserted into G and Rep_G is updated as the anonymization of Rep_G and Tr'. This anonymization process continues until G contains k trajectories. At the end of the iteration, the trajectories in G are removed from $\mathcal{T}$. This step finishes when there are less than k remaining trajectories in $\mathcal{T}$.

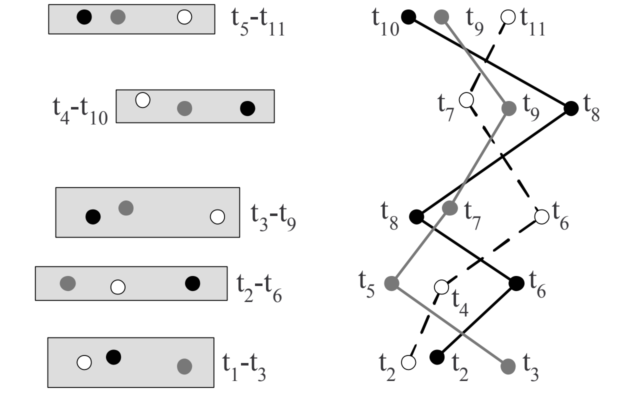
   Figure 10 gives an example of generalizing three trajectories Tr_1, Tr_2 and Tr_3 into a 3-anonymized trajectory, where the timestamp of each location sample is shown beside its location. In this example, Tr_2 is first added into an empty group G as its representative Rep_G. Next Tr_1 is added to G and the location samples of Tr_1 and Tr_2 are generalized into a sequence of anonymized regions (represented by shaded rectangles), as depicted in Figure 10b. Rep_G is updated as the anonymization of Tr_1 and Tr_2, $Tr*$ (Figure 10c). Tr_3 is also added into G and a sequence of new anonymized regions are formed for G (Figure 10d). The time span of an anonymized region is the range from the smallest and largest timestamps of the location samples included in the region. Note that unmatched points (e.g., the location sample of Tr_3 at timestamp t_7) are suppressed in this step. Since G already contains $k = 3$ trajectories, the anonymization process for G is done (Figure 10e).

2. **Reconstruction step.** Given a k-anonymized trajectory, k locations are uniformly selected in each of its anonymized region, as illustrated in Figure 11a. Next, for each selected location, a timestamp is also uniformly selected from its associated time span. k trajectories are reconstructed by linking a unique location sample in each monitored region (Figure 11b).



(a) Original trajectories (b) Anonymity group $G = \{Tr_1, Tr_2\}$ (c) Anonymization of G (d) Add Tr_3 into G (e) Anonymization of G

Figure 10: Generalization-based approach: Anonymization step.



(a) Location samples selection (b) Trajectory reconstruction

Figure 11: Generalization-based approach: Reconstruction step.

The reconstructed trajectory data set can be released to the public or third parties for answering spatio-temporal queries and data analysis (e.g., data mining).

5. CONCLUSION

Location privacy protection in continuous location-based services (LBS) and trajectory data publication has drawn a lot of attention from the industry and academia. It is expected that more effective and efficient privacy preserving technologies will be developed in the near future. We want to provide some future directions in these two problems as the conclusion of this survey. For continuous LBS, new privacy-preserving techniques are needed to protect personalized LBS. This is because personalized LBS require more user semantics, e.g., user preferences and background information, such as salary and occupation, rather than just some simple query parameters, such as a distance range and an object type of interest. An adversary could use such user semantics to infer the user location with higher certainty. For example, suppose that an adversary knows that a target user Alice usually has dinner from 6pm to 7pm during weekdays and she does not like Japanese and Thailand food. Given a cloaked spatial region of Alice's location at 6:30pm on Monday and the region contains two Japanese restaurants, one Thailand restaurant and one Chinese restaurant, the adversary can infer that Alice in the Chinese restaurant with very high certainty. Existing privacy-preserving techniques for location trajectory publication only support simple aggregate analysis, such as range queries and clustering. Researchers should develop new trajectory anonymization techniques that support more useful and complex spatio-temporal queries (e.g., how many vehicles travel from a shopping mall to a cinema from 1pm to 2pm during weekends, the most popular path, and their average travel time) and data analysis (e.g., pattern recognition and association rules).

6. REFERENCES

[1] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *Proceedings of the IEEE International Conference on Data Engineering*, 2008.

[2] B. Bamba, L. Liu, P. Pesti, and T. Wang. Supporting anonymous location queries in mobile environments with PrivacyGrid. In *Proceedings of the International Conference on World Wide Web*, 2008.

[3] J. Bao, C.-Y. Chow, M. F. Mokbel, and W.-S. Ku. Efficient evaluation of k-range nearest neighbor queries in road networks. In *Proceedings of the International Conference on Mobile Data Management*, 2010.

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

[5] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar. Preserving user location privacy in mobile data management infrastructures. In *Proceedings of International Privacy Enhancing Technologies Symposium*, 2006.

[6] C.-Y. Chow, J. Bao, and M. F. Mokbel. Towards location-based social networking services. In *Proceed-*

ings of the ACM SIGSPATIAL International Workshop on Location Based Social Networks, 2010.

[7] C.-Y. Chow, M. Mokbel, and T. He. A privacy-preserving location monitoring system for wireless sensor networks. *IEEE Transactions on Mobile Computing*, 10(1):94–107, 2011.

[8] C.-Y. Chow and M. F. Mokbel. Enabling private continuous queries for revealed user locations. In *Proceedings of the International Symposium on Spatial and Temporal Databases*, 2007.

[9] C.-Y. Chow, M. F. Mokbel, and W. G. Aref. Casper*: Query processing for location services without compromising privacy. *ACM Transactions on Database Systems*, 34(4):24:1–24:48, 2009.

[10] C.-Y. Chow, M. F. Mokbel, J. Bao, and X. Liu. Query-aware location anonymization in road networks. *GeoInformatica*, In press, `http://dx.doi.org/10.1007/s10707-010-0117-0`.

[11] C.-Y. Chow, M. F. Mokbel, and X. Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based services. In *Proceedings of the ACM Symposium on Advances in Geographic Information Systems*, 2006.

[12] C.-Y. Chow, M. F. Mokbel, and X. Liu. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. *GeoInformatica*, 15(2):351–380, 2011.

[13] C.-Y. Chow, M. F. Mokbel, J. Nap, and S. Nath. Evaluation of range nearest-neighbor queries with quality guarantee. In *Proceedings of the International Symposium on Spatial and Temporal Databases*, 2009.

[14] Dateline NBC. Tracing a stalker. `http://www.msnbc.msn.com/id/19253352`, June 2007.

[15] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In *Proceedings of International Conference on Pervasive Computing*, 2005.

[16] FoxNews. Man accused of stalking ex-girlfriend with GPS. `http://www.foxnews.com/story/0,2933,131487,00.html`, September 2004.

[17] J. Freudiger, M. Raya, M. Felegyhazi, P. Papadimitratos, and J.-P. Hubaux. Mix-zones for location privacy in vehicular networks. In *Proceedings of the International Workshop on Wireless Networking for Intelligent Transportation Systems*, 2007.

[18] J. Freudiger, R. Shokri, and J.-P. Hubaux. On the optimal placement of mix zones. In *Proceedings of International Privacy Enhancing Technologies Symposium*, 2009.

[19] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4):14:1–14:53, 2010.

[20] B. Gedik and L. Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 7(1):1–18, 2008.

[21] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan. Private queries in location based services: Anonymizers are not necessary. In *Proceedings of the ACM Conference on Management of Data*, 2008.

[22] G. Ghinita, P. Kalnis, and S. Skiadopoulos. PRIVÉ: Anonymous location-based queries in distributed mobile systems. In *Proceedings of the International Conference on World Wide Web*, 2007.

[23] G. Ghinita1, P. Kalnis, and S. Skiadopoulos. MobiHide: A mobile peer-to-peer system for anonymous location-based queries. In *Proceedings of the International Symposium on Spatial and Temporal Databases*, 2007.

[24] Google Geocoding API. `http://code.google.com/apis/maps/documentation/geocoding/`.

[25] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*, 2003.

[26] M. Gruteser and B. Hoh. On the anonymity of periodic location samples. In *Proceedings of the International Conference on Security in Pervasive Computing*, 2005.

[27] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Achieving guaranteed anonymity in GPS traces via uncertainty-aware path cloaking. *IEEE Transactions on Mobile Computing*, 9(8):1089–1107, 2010.

[28] J. I. Hong and J. A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*, 2004.

[29] H. Hu and D. L. Lee. Range nearest-neighbor query. *IEEE Transactions on Knowledge and Data Engineering*, 18(1):78–91, 2006.

[30] S. Ilarri, E. Mena, and A. Illarramendi. Location-dependent query processing: Where we are and where we are heading. *ACM Computing Surveys*, 42(3):12:1–12:73, 2010.

[31] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias. Preventing location-based identity inference in anonymous spatial queries. *IEEE Transactions on Knowledge and Data Engineering*, 19(12):1719–1733, 2007.

[32] A. Khoshgozaran and C. Shahabi. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy. In *Proceedings of the International Symposium on Spatial and Temporal Databases*, 2007.

[33] H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. In *Proceedings of IEEE International Conference on Pervasive Services*, 2005.

[34] W.-S. Ku, R. Zimmermann, W.-C. Peng, and S. Shroff. Privacy protected query processing on spatial networks. In *Proceedings of the International Workshop on Privacy Data Management*, 2007.

[35] K. LeFevre, D. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In *Proceedings of the IEEE International Conference on Data Engineering*, 2006.

[36] N. Li, T. Li, and S. Venkatasubramanian. Closeness: A new privacy measure for data publishing. *IEEE Transactions on Knowledge and Data Engineering*, 22(7):943–956, 2010.

[37] C. Y. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao. Privacy vulnerability of published anonymous mobility traces. In *Proceedings of the ACM International Conference on Mobile Computing and Networking*, 2010.

[38] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1):3:1–3:52, 2007.

[39] Marist Institute for Public Opinion (MIPO). Half of Social Networkers Online Concerned about Privacy. `http://maristpoll.marist.edu/714-half-of-social-networkers-online-%concerned-about-privacy/`. July 14, 2010.

[40] M. F. Mokbel, C.-Y. Chow, and W. G. Aref. The new casper: Query procesing for location services without compromising privacy. In *Proceedings of the International Conference on Very Large Data Bases*, 2006.

[41] M. F. Mokbel and J. Levandoski. Towards context and preference-aware location-based database systems. In *Proceedings of the ACM International Workshop on Data Engineering for Wireless and Mobile Access*, 2009.

[42] K. Mouratidis and M. L. Yiu. Anonymous query processing in road networks. *IEEE Transactions on Knowledge and Data Engineering*, 22(1):2–15, 2010.

[43] M. E. Nergiz, M. Atzori, Y. Saygin, and B. Güç. Towards trajectory anonymization: A generalization-based approach. *Transactions on Data Privacy*, 2(1):47–75, 2009.

[44] B. Palanisamy and L. Liu. Mobimix: Protecting location privacy with mix zones over road networks. In *Proceedings of the IEEE International Conference on Data Engineering*, 2011.

[45] X. Pan, X. Meng, and J. Xu. Distortion-based anonymity for continuous queries in location-based mobile services. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009.

[46] A. Pfitzmann and M. Kohntopp. Anonymity, unobservability, and pseudonymity - a proposal for terminology. In *Proceedings of the Workshop on Design Issues in Anonymity and Unobservability*, 2000.

[47] D. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24(6):843–854, 1979.

[48] P. Samarati. Protecting respondents identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.

[49] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):571–588, 2002.

[50] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.

[51] USAToday. Authorities: GPS system used to stalk woman. `http://www.usatoday.com/tech/news/2002-12-30-gps-stalker_x.htm`, December 2002.

[52] J. Voelcker. Stalked by satellite: An alarming rise in gps-enabled harassment. *IEEE Spectrum*, 47(7):15–16, 2006.

[53] T. Wang and L. Liu. Privacy-aware mobile services over road networks. In *Proceedings of the International Conference on Very Large Data Bases*, 2009.

[54] Webroot Software, Inc. Webroot survey finds geolocation apps prevalent amongst mobile device users, but 55% concerned about loss of privacy. `http://pr.webroot.com/threat-research/cons/social-networks-mobile-security-071310.html`. July 13, 2010.

[55] X. Xiao, K. Yi, and Y. Tao. The hardness and approximation algorithms for l-diversity. In *Proceedings of the International Conference on Extending Database Technology*, 2010.

[56] T. Xu and Y. Cai. Location anonymity in continuous location-based services. In *Proceedings of the ACM Symposium on Advances in Geographic Information Systems*, 2007.

[57] T. Xu and Y. Cai. Exploring historical location data for anonymity preservation in location-based services. In *Proceedings of IEEE INFOCOM*, 2008.

[58] M. L. Yiu, C. Jensen, X. Huang, and H. Lu. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In *Proceedings of the IEEE International Conference on Data Engineering*, 2008.

[59] T.-H. You, W.-C. Peng, and W.-C. Lee. Protecting moving trajectories with dummies. In *Proceedings of the International Workshop on Privacy-Aware Location-Based Mobile Services*, 2007.

[60] C. Zhang and Y. Huang. Cloaking locations for anonymous location based services: A hybrid approach. *GeoInformatica*, 13(2):159–182, 2009.