Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6691975, 12 pages
https://doi.org/10.1155/2021/6691975

Research Article

Protecting the Moving User's Locations by Combining Differential Privacy and k-Anonymity under Temporal Correlations in Wireless Networks

Weiqi Zhang,[1] **Guisheng Yin**,[1] **Yuhai Sha**,[1] **and Jishen Yang**[2]

[1]*College of Computer Science and Technology, Harbin Engineering University, Heilongjiang, China*
[2]*Department of Computer Science, Georgia State University, Georgia, USA*

Correspondence should be addressed to Weiqi Zhang; zhangweiqi@hrbeu.edu.cn

Received 11 November 2020; Revised 21 December 2020; Accepted 12 January 2021; Published 2 February 2021

Academic Editor: Xiao Zhang

The rapid development of the Global Positioning System (GPS) devices and location-based services (LBSs) facilitates the collection of huge amounts of personal information for the untrusted/unknown LBS providers. This phenomenon raises serious privacy concerns. However, most of the existing solutions aim at locating interference in the static scenes or in a single timestamp without considering the correlation between location transfer and time of moving users. In this way, the solutions are vulnerable to various inference attacks. Traditional privacy protection methods rely on trusted third-party service providers, but in reality, we are not sure whether the third party is trustable. In this paper, we propose a systematic solution to preserve location information. The protection provides a rigorous privacy guarantee without the assumption of the credibility of the third parties. The user's historical trajectory information is used as the basis of the hidden Markov model prediction, and the user's possible prospective location is used as the model output result to protect the user's trajectory privacy. To formalize the privacy-protecting guarantee, we propose a new definition, L&A-location region, based on k-anonymity and differential privacy. Based on the proposed privacy definition, we design a novel mechanism to provide a privacy protection guarantee for the users' identity trajectory. We simulate the proposed mechanism based on a dataset collected in real practice. The result of the simulation shows that the proposed algorithm can provide privacy protection to a high standard.

1. Introduction

In recent years, the booming amount of personal mobile devices with location services has promoted the development of location-based systems in wireless networks [1]. The widespread use of mobile smart devices has laid the foundation for massive data collection based on mobile perception. In these data collection systems, location-based services (LBSs) provide real-time services related to the user's current location information. Various useful applications depend on LBSs. For example, Google Maps provides navigation services such as route suggestions and road traffic condition notifications. Groupon and Yelp [2] provide business service information based on the distances from the users' location. Although LBSs (as shown in Figure 1) are very useful and convenient for users, these conveniences are at the expense of users' private information. The service providers can infer an individual's residence information, work location, and other private information by observing the user's temporal correlated data [3–6].

Many methods for personal information protection are proposed as follows. One of the solutions is Private Information Retrieval (PIR) [7]. In modern cryptography, the main purpose of PIR is to allow a user to retrieve items from a server without disclosing any private information. In other words, the server does not know the user's specific query information and retrieved data in the process. However, one major disadvantage of this technique is the enormous amount of calculation for redesigning different queries according to different query types. Most of the methods are

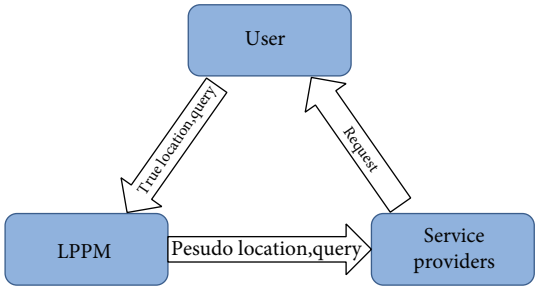


Figure 1: Location-based services.

developed based on location obfuscation, which uses a cloaking area or a perturbed location. These solutions rely on syntactic privacy models, which cannot provide a strict privacy guarantee. Unfortunately, most of the solutions only consider the stationary scenario and perturb the location at a single timestamp while neglecting the temporal correlation of the movement of the user's locations. Hence, the adversaries can effortlessly access more private information by linking inference attacks. Most typical methods to protect users' private information use differential privacy and k-anonymity. k-Anonymity as one of the principal approaches [8–10] ensures the probability of success to any linking attack to be lower than $1/k$. However, it provides a lower privacy guarantee and data utility.

Differential privacy [11] was originally proposed by Dwork in 2006. Later, the idea is regarded as a standard for private information preservation. Although the applications of differential privacy in protecting private information have gradually become applicable in practice, some challenges still exist in the problem of continuous location sharing. First of all, in the standard privacy protection settings, only user-level privacy (whether a user appears in a dataset or not) is protected. In our setting, the trajectories of a single user are protected for a period of time. Second, the released trajectory can be identified based on road networks without temporal correlation. Furthermore, the adversary can identify the user captured by moving patterns. Finally, none of the effective released trajectory mechanism utilizes the combination of k-anonymity and differential privacy.

In this paper, we propose an all-new solution to preserve the user's trajectory privacy with k-anonymity and differential privacy. As shown in Figure 2, a moving user needs to continuously share locations with untrusted service providers or other third parties in a period of time. In other words, in our solution, a user's accurate location information is only known by him. We regard all service providers and third parties as adversaries. The adversaries have side knowledge as much as they can obtain. We propose a new privacy protection system that enables private location sharing without disclosing users' accurate locations to these adversaries and protects users' trajectories in a continuous time period.

The proposed system is noted as UGIS (User and Geographic Space-Indistinguishable System), and this system consists of two parts. One part is the KD-location region (KD is the k-anonymity and differential privacy for short), referred to as a special region in the context. Another part

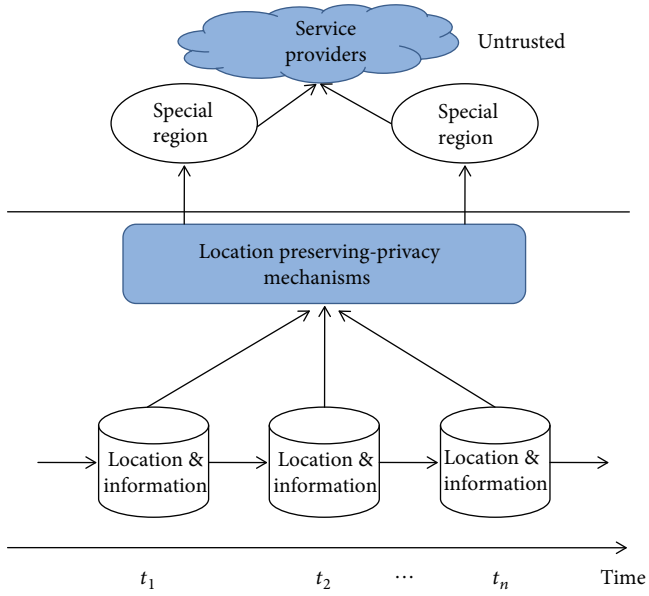


Figure 2: Processing mechanism.

is the users' accurate trajectory processing mechanism. In the KD-location region, adversaries cannot recognize the target user. We then move on to the trajectory processing mechanism that makes a good performance to protect users' trajectories. To our knowledge, UGIS is a better private processing mechanism that combines k-anonymity and differential privacy methods to protect the location and trajectory information of users. The following is a summary of this paper's contribution:

(i) To protect the user's accurate location, we only need to "hide" it in a special region set in which the adversaries cannot distinguish the locations or users. Accordingly, we propose a special region set based on k-anonymity and differential privacy to protect the accurate location of each timestamp

(ii) To show that the user's movement is associative and temporally correlated. In our problem, the user's location transfer is time-related. We use the Markov model to represent the user's location change between continuous timestamps [12]. Hence, from the perspective of adversaries, the user's location transfer model is a hidden Markov model (HMM)

(iii) We focus on the transfer mechanism in a Markov model. We utilize the concept of differential privacy to add noise to this transfer mechanism to make the users' trajectories indistinguishable

The rest of this paper is organized as follows. Section 2 is our related works. In Section 3, we discuss the notions of location privacy from the literature and analyze the weaknesses and strengths of the state-of-the-art algorithms. Also, in Section 3, we introduce the coordinate system and location transition model. We then describe several components and definitions of our UG-indistinguishable system in Section 4.

Section 5 contains the illustration of the framework and the implementation of our location release algorithm. The experiment and evaluation are presented in Section 6.

2. Related Works

In this part, we mainly make some generalizations and summaries of the previous literature. A few recent works [3–15] provide an overview of location privacy protection mechanisms (LPPMs) and methods. These location privacy protection mechanisms mainly use obfuscation technology to achieve the anonymity area. The most widely used approach to construct the LPPMs is k-anonymity. The notion of k -anonymity is commonly used to protect privacy for location-based systems in most of the works in the field. These systems mainly focus on protecting the users' identities and preventing the adversary from inferring accurate information among k users from the published user datasets. One way to implement this method is to use dummy locations mentioned in [16, 17]. However, since the output of the dummy location is controlled by the server, the adversaries can easily find out where the dummy location is not logically generated. Another method to achieve k-anonymity is through the cloaking region [18–20]. The disadvantage of the method is the high risk of having a too-large cloaking area to satisfy the$k$value in the scene with few users. A different approach is to add certain quality constraints to provide better privacy protection [21], while [22] additionally using bandwidth constraints. Literature [7] also proposed a location privacy mechanism focusing on the evaluation based on location-based range queries. This method evaluates the degree of privacy according to the size of the cloaking area and the coverage of the sensitive area. Two methods have been proposed to deal with the adversary's background knowledge, by expanding the anonymous area or delaying the sending of requests. Both solutions may lead to a decline in service quality. The methods based on k-anonymity are improved, but the definition of differential privacy provides a more rigorous guarantee.

Several privacy-protecting methods use the differential privacy approach in recent works [23, 24]. For instance, [25] presents a way to statistically simulate the location data from a database while providing privacy guarantees. They designed an information perturbation mechanism to generate aggregated information from a large amount of locations, trajectories, and spatiotemporal data [26–29]. [30] proposed a differential privacy data mining algorithm that uses a spatial quadtree decomposition technique to preprocess the locations. The work closest to ours is [31]. A large part of the research is based on the use of cloaking areas to enforce location confusion mechanisms. This method leads to a reduction in the utility of published data. [32] proposed a data sanitization method collectively manipulating users' profiles and friend relationships. This method is not suitable for our framework setting and further research. However, the method does not solve the users' movement trajectory problem. In this paper, our system protects the users' accurate location with a rigorous privacy guarantee and makes the users' trajectories indistinguishable at each timestamp.

3. Preliminary

In this section, we discuss various notions of location privacy-preserving methods such as k-anonymity, differential privacy, and location transfer model. We consider a scenario where a user wants to post a query about points of interest at the current location by using a personal device (e.g., smartphone) to query a public service provider. The users expect their accurate location to be private regardless of the process of the search. Our goal is to develop a real-time privacy mechanism that provides privacy protection in a formal notion to achieve users' expected privacy protection level. A list of frequently used symbols in this paper is all motioned in Table 1.

3.1. k-Anonymity. k-Anonymity is one of the privacy protection methods widely used in most location-based systems. These systems focus on protecting the user's identity, making the adversary unable to infer which user is the true target among k users. One way is to generate $k-1$ properly pseudo points and use the actual location and pseudo locations to perform k queries to the service provider. Another way to achieve k-anonymity is through a cloaking area. This approach involves creating a cloaking area that includes k users sharing some points of interest, then querying the server using this cloaking area instead of the accurate location. Unfortunately, the adversaries can identify the target user when adequate side knowledge is available. Pseudo locations are only useful if they have enough similarity with the real locations from the adversaries' point of view.

As a result, notions that abstract from adversaries' knowledge, such as differential privacy, have more popularity later than k-anonymity approaches.

3.2. Differential Privacy and Laplace Mechanism. Differential privacy (DP) [11] is a notion of private information inspired by the concept of statistics. DP guarantees to maximize the accuracy of data queries when querying from the statistical database while minimizing the chance of identifying other records. DP removes the individual characteristics while preserving statistical characteristics to protect the user's privacy. DP has gradually become the de facto standard in data privacy due to its strong privacy guarantees in statistical analysis. Moreover, differential privacy is a semantic model that does not need to rely on the adversary's background knowledge and provides a higher level of semantic security from private information. Differential privacy ensures that adversaries cannot infer whether a particular user is present in the original data. Releasing data according to differential privacy ensures that adversaries cannot infer any information about personal information from the "sanitized data." The definition of differential privacy is demonstrated as follows.

Table 1: Summary of notations.

Symbol	Meaning
a_u^t	User's information at timestamps t
X_i, Y_i	Axis in the coordinate system
Δ	The special location set a user is located in
$p_t[i]$	The probability of a user in region i at timestamps t
l	Current location of the user
l'	Puppet of the user's location
R_t	The special region a user is located in
$\widehat{l}$	Nearest location to the target user's location in the special region
S_f	The sensitivity
TM	Markov transfer mechanism
D, D'	Neighboring database
T_{a_u}	Trajectory of a user

Definition 1. (differential privacy). A mechanism M satisfies ε-differential privacy if any outputs $\in S$ and database D and its neighboring database D' can be obtained by either adding or removing a single record, and the following holds:

$$\Pr\left(M(D) \in s\right) \leq e^{\varepsilon} \times \Pr\left(M\left(D'\right) \in s\right). \tag{1}$$

The Laplace mechanism [33] is commonly used to achieve ε-differential privacy. It is built on the sensitivity defined as follows.

Definition 2. (sensitivity). For any query $f(D)$: $D \to R^d$, l_1-norm sensitivity is the maximum l_1-norm of $f(D) - f(D')$, where D and D' are any two instances in neighboring databases as the following equation holds how to capture the sensitivity of two neighboring databases:

$$S_f = \max_{D,D'} \left\| f(D) - f\left(D'\right) \right\|_1. \tag{2}$$

The Laplace mechanism implements differential private protection by adding noise of Laplace distribution to the query result $\mathrm{Ans}(D) = f(D) + \mathrm{Lap}(\beta)$, where $\beta = S_f/\varepsilon$. As shown above, the concept of differential privacy is generally applied in the joint publishing of compound data. The standard concept makes it unsuitable for applications that involve only one person. In this paper, we propose a more rigorous privacy guarantee with k-anonymity and differential privacy methods.

3.3. Coordinate System. We divide a map into grids where each grid is a state in the Markov model. The users' real locations can be denoted by the state grids in a Markov model. Denote R is the area that includes all the state grids. These areas R can be divided into many spaces $R = \{r_1, r_2, r_3, \cdots, r_i\}$, where each r_i means a unit grid in region R. We set up a spatial coordinate system by which the user's accurate longitude and latitude can be represented as X-axis and Y-axis coordinates. Vector coordinates represent each grid unit, which more clearly shows the user's current position and the corresponding state grid in the Markov model. In Figure 3, all grids have the same size, but in the real world, the sizes of each region are not necessarily equal.

The following example illustrates how to use these grids to denote the user's current location. If the user is located in the area r_6, we denote into this state coordinate system, where $l_u = r_6 = [1, 3]$ with $x = 1$ and $y = 3$. As time goes by, the trajectory of a user's movement is represented by a series of state l_{u_i} in the map coordinate system.

3.4. Location Transition Model and HMM. This paper is based on the study of moving users' trajectories, so we propose to use the random process Markov chain [34–36] to simulate the movement of the user from one point to another under temporal correlations. Other constraints, such as the road networks, can also be captured by the Markov model. The kernel of the Markov model is the state transition matrix. The current state only depends on the transition matrix and its previous moment state. As mentioned before, the user's real locations are unobservable and only known by him. Hence, for the adversaries, the user's movement process is a hidden Markov model.

We use P_t to represent the location of the user at timestamp t. $P_t[r_i] = \Pr\left(a_u^t = r_i\right)$ represents the probability that user u appears in the area r_i at timestamp t. Therefore, we construct the Markov process as follows:

$$\Pr\left(a_u^t = r_i\right) = \Pr\left(a_u^{t_i} = r_i \mid a_u^{t_1} = r_1, a_u^{t_2} = r_2, \cdots, a_u^{t_{i-1}} = r_{i-1}\right). \tag{3}$$

In the first-order Markov model, there is the hypothesis that the transition probability of state and the output probability of observation are only dependent on the current state. So the Markov process can be simplified to

$$\Pr\left(a_u^t = r_i\right) = \Pr\left(a_u^{t_i} = r_i \mid a_u^{t_{i-1}} = r_{i-1}\right). \tag{4}$$

The transition probability,

$$\Pr_u^{r_{ij}} = \Pr\left(a_u^{t_{i+1}} = r_j \mid a_u^t = r_i\right), \tag{5}$$

is the one-step transition probability from timestamp t to $t + 1$. The transition probability satisfies the following properties:

$$\begin{gathered} \Pr_u^{r_{ij}} \geq 0, \\ \sum_{i=0}^{\infty} \Pr_u^{r_{ij}} = 1, \\ j = 1, 2, 3, \cdots. \end{gathered} \tag{6}$$

The sum of the transition probabilities of all possible locations of the user from timestamp t to $t + 1$ is 1. We implement the Markov model on the trajectory of moving users and get $P_t = P_{t-1} * \mathrm{TM}$ to denote the probability

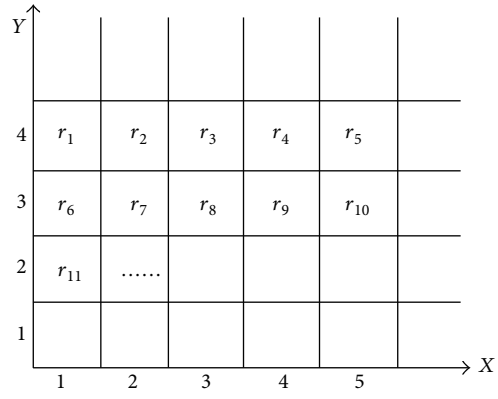


Figure 3: State coordinate system.

transfer at each timestamp. The transition matrix TM is given in our system.

4. UG-Indistinguishable System Model

To apply k-anonymity and differential privacy in the area where moving users share locations on consecutive timestamps, we conduct a rigorous privacy analysis and set a special region R_t (contains an actual user and pseudo users). Even if the adversaries capture this special region, they still cannot identify the target user.

4.1. BB Cloaking Region R. The essence of applying differential privacy in location sharing is to "hide" a real location in a database by adding or removing one record to obtain a "neighboring database." The special region R can be regarded as a set of locations. The adversaries cannot infer whether the target user is in this database R or not with any kinds of queries. However, such a dataset is not completely suitable for our problem. So we proposed a new notion, the black B cloaking region. The black B cloaking region is also noted as a special region to hide users' accurate location at every timestamp. We need to compute the amount of information as follows to obtain the cloaking region at timestamp t:

$$H\left(a_u^t\right) = \sum \Pr\left(a_u^t = r_i\right) \cdot \log\Pr\left(a_u^t = r_i\right), \tag{7}$$

which is also known as the amount of prior information that user u is in region r_i at timestamp t. We intend to use $H(a_u^t = r_i \mid K)$ to represent the posterior information that adversaries infer the user's location information through existing background knowledge K.

$$H\left(a_u^t = r_i \mid K\right) = \sum \Pr\left(a_u^t = r_i \mid K\right) \cdot \log\Pr\left(a_u^t = r_i \mid K\right). \tag{8}$$

In summary, the amount of the user's location information disclosed to the adversary is as follows:

$$H' = H\left(a_u^t\right) - H\left(a_u^t = r_i \mid K\right). \tag{9}$$

Thus, we can obtain a definition of generating a special location set.

Definition 3. (θ-location set). We can set the probability that the adversary can infer the user's current location $\Pr(a_u^t = r_i \mid K)$ as the posterior probability, and the probability of the user at the current location is $\Pr(a_u^t = r_i)$; then, the privacy requirement is as follows:

$$\Pr\left(a_u^t = r_i \mid K\right) - \Pr\left(a_u^t = r_i\right) \le \theta. \tag{10}$$

θ is the privacy threshold of the user's current location at timestamp t and $0 < \theta < 1$. In this article, we set parameter θ as no greater than 0.3. We assume that the parameter θ is given in our framework. When the user's location information exposed to the adversary is greater than the privacy threshold, the cloaking region needs to be generated for protecting the real location at timestamp t.

We define a special region based on k-anonymity and differential privacy, which intuits that the released area will not help an adversary to distinguish any instances in the region. According to Definition 1, we make a transformation that adjusts to a special region R in our article. The new definition is shown as follows.

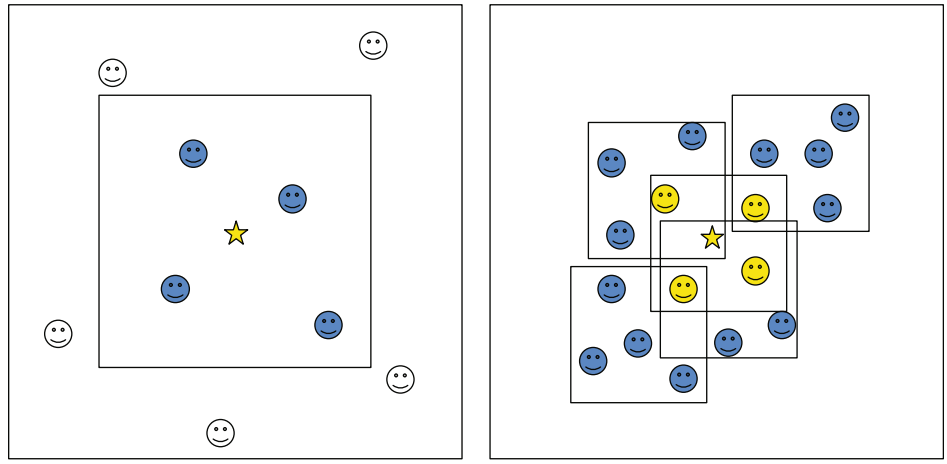
Definition 4. (BB cloaking region differential privacy$*$). At any timestamp$t$, the cloaking region generation by mechanism$M$is represented asR_t, the query function represented as$q()$, and the query result of$q()$on the cloaking region satisfiesε-differential privacy, and the following holds:

$$\Pr\left(q\left(a_{u_1}^t\right) \in R_t\right) \le e^{\varepsilon} \cdot \Pr\left(q\left(a_{u_2}^t\right) \in R_t\right). \tag{11}$$

The definitions guarantee that the accurate location is always protected in a location set Δ at each timestamp. The released region R_t is differentially private at timestamp t for continuous location sharing under temporal correlation. We use the following context to explain how the special region work. In the beginning, a user moves to a new location where he may send a query (e.g., find the nearby restaurant) [37, 38]. At each timestamp, we denote the user's individual information as $A_{u_i}^t = \{\text{time, location, sex, age, query}\}$. The user can be treated as a target user. Then, we assume a mechanism in our system that can obtain a set of k nearest neighbor users with the same query. This set allows the existence of the dummy users, and our system can release the dummy users. Our anonymous generation process is more complicated, and the best effect is verified by experiments when $k = 5$. We regard these four nearest neighbor users as the new target users, respectively. Hence, we have a set of users, as shown in Figure 4.

4.2. Razor Mechanism. After obtaining a set of nearest neighbor users, we propose a new method, Razor Mechanism, to filter the similar terms. We use this mechanism to eliminate users whom we do not want to appear in the nearest neighbor users set.

Even though the adversaries have side knowledge as much as they can have, they still cannot know which nearest neighbor users are generated by the first anonymity. The Razor Mechanism uses the principle of similarity measurement



(a) The first anonymity (b) The second anonymity

Figure 4: Obtain the set of users.

[36] to filter out the pseudo users generated in the first anonymity. In services based on location information, we usually set the distance between locations as a measure of similarity. The similarity measurement of users in the anonymous area is shown in the following formula:

$$\mathrm{Sim}\left(a_{u_i}^t, a_{u_j}^t\right) = \frac{\mathrm{dist}\left(u_i, u_j\right)}{\mathrm{maxdist}\left(R_t\right)}. \tag{12}$$

As shown in Figure 4(a), the data generated in the anonymous data preprocessing stage are removed as noisy data by the Razor Mechanism. Figure 4(b) shows all the remaining data in the special cloaking region, excluding the location coordinate data of yellow dots.

4.3. Drift and Puppet. We use the Razor Mechanism to filter out noisy points that we do not want to appear because the special region contains almost all similar users and possible location information. The target user's information is also eliminated with a small probability (technically, $\mathrm{dist}(u_i, u_j) = \mathrm{maxdist}(R_t)$). This phenomenon is referred to as "drift" and can be solved with the puppet approach in the special region. We use $l_t(\mathrm{lon}, \mathrm{lat})$ and $l_t{}'(\mathrm{lon}', \mathrm{lat}')$ to denote the user's accurate location and puppet, respectively. The definition of the puppet mechanism is shown as follows.

Definition 5. (puppet). A puppet $l_t{}'$ is a cell in the special set which has the closest distance to the target user's location l_t:

$$l_t'\left(\mathrm{lon}', \mathrm{lat}'\right) = \arg\max_{\widehat{l} \in \Delta} \mathrm{dist}\left(\widehat{l}, l\right). \tag{13}$$

In this equation, Δ represents the special location set, and the function $d(\widehat{l}, l)$ denotes the distance between two users in the special region. Note that the puppet approach does not leak any information about the target user. If the target user is in the special set, we protect the target user in the region; otherwise, the puppet is then protected in the special set. Using a puppet does not disclose whether the user's location l_t is in R_t or not. We have mentioned before that our location release mechanism is treated as a black box region. It is still a black box after replacing the accurate location with a puppet.

5. Location and Trajectory Release Algorithm

5.1. Framework. The framework of the special location region release algorithm is shown in Algorithm 1. We generate a special location set at every timestamp to protect a single user's accurate location continuously. The procedure of the generation of the special location set at timestamp t is explained in the context above. First, from lines 1 to 6 in Algorithm 1, the model makes a prediction based on the hidden Markov model. At each timestamp t, we compute the probability p_{t-1}. If the current location at timestamp t satisfied the privacy threshold ($p_{t-1} \le \theta$), the procedure moves to the next timestamp. Otherwise, the special cloaking region is generated at the current location. The process of the special location set is shown in lines 8 to 19. If the target user is filtered out of the special location set, we use a puppet in Δ_t at timestamp t as if it is the "target" user in the release mechanism. Our proposed algorithm uses l_1-norm to capture the sensitivity of the special location set. After all these steps, we can obtain a special location set. According to this special location set, we can generate a special region for a single user at timestamp t. In this region, no matter how much side knowledge the adversaries may have, the adversaries can no longer distinguish the target user from the users.

5.2. Linking Differential Privacy to Trajectory. A user's location trajectory is a moving path or trace reported by a moving object in the geographical space. The user's trajectory T_{a_u} is represented by a set of n time-order points, $T_{a_u} : p_1 \longrightarrow p_2 \longrightarrow \cdots \longrightarrow p_n$, where each point p_i consists of a geospatial coordinate set (X_i, Y_i) and timestamp t (i.e., $p_i = (X_i, Y_i, t)$, where $1 \le i \le n$). Such temporal and spatial attributes of a location trajectory can be considered powerful quasi-

Framework.
Input: a_u^t, TM, p_{t-1}, l, l'
Output: Cloaking region R_t
$p_t \longleftarrow p_t$TM;
Pr $(a_u^t = r_i)$ − Pr $(a_u^t = r_i \mid K) = p_{t-1}$;
if $p_{t-1} \geq \theta$ **then**
    Construct a special set of this location;
else
    Go to next timestamps;
end if
Construct a special location set:
Run k-anonymity$\longrightarrow$set Δ^*;
$r_i = (x_i ; y_i)$;
for (int $k = 0, k \leq 5, k$++) **do**
    r_i = Random $(x_i^* ; y_i^*)$
    if dist$(r_i ; r_i^*) \in$ (dist$_{\min}$; dist$_{\max}$) **then**
        add r_i^* to Δ^*
        algorithm goes on
    else
    go to line 11
    end if
end for
    $\Delta^* \longrightarrow$ Razor Mechanism;
    Sim$(a_{u_i}^t ; a_{u_j}^t)$ = dist$(u_i ; u_j)$/max (R_t)
while Check l' **do**
    if $l \in R_t$ **then**
        algorithm goes on;
    else $\{l \notin R_t\}$
        $l' \longleftarrow$ surrogate;
    end if
end while
    Obtain sensitivity of the special set Δ_1;
    Δ_1 + Lap$(\Delta f/\varepsilon) = R_t$;
    Releases this region R_t;
end;
return Algorithm;
$\longrightarrow$ Go to the next timestamps

Algorithm 1

identifiers that can be linked to various other kinds of physical data objects [39, 40]. From the adversaries' point of view, these trajectories may disclose users' individual information such as users' work, home, and points of interest (POI). Although such trajectories can be made anonymous by replacing the identifier of users with random identifiers, the users may still suffer from privacy threats.

In this paper, our approach uses the Markov model to denote users' movement from one special region to another. We use the equation $p_t = p_{t-1} * \mathrm{TM}$ to denote a single user moving from one region (at timestamp $t-1$) to another region (at timestamp t), and TM denotes the transfer mechanism of users' movement. The transfer mechanism uses Laplace noise to make users' trajectories indistinguishable. As shown in Figure 5, we add Laplace noise to the Markov transfer mechanism to make users' transition probability basically the same. For example, when a user moves from region r_1 to another, according to his habits, the transfer

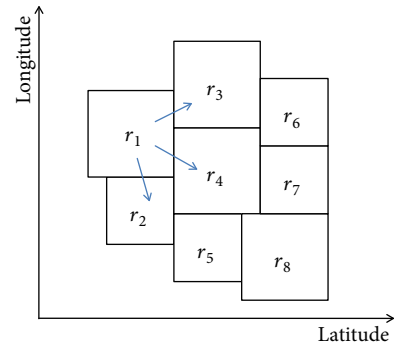


Figure 5: Possible transfer state.

probability at each region is different (e.g., the probability from r_1 to r_2 is 5/10, r_1 to r_3 is 2/10, and r_1 to r_4 is 3/10). According to our method, after adding Laplace noise to the transfer mechanism, the transition probability from r_1 to r_2 is 4/10, r_1 to r_3 is 3/10, and r_1 to r_4 is 3/10. In the following section, we will show the performance by the experiment results.

6. Experiment and Evaluation

In this section, we present the evaluation of our method. All algorithms are implemented in Python on macOS with the real-world datasets GeoLife and Gowalla [41–43]. The GeoLife dataset is collected in (Microsoft Research Asia) GeoLife project by 182 users from April 2007 to August 2012. A time-stamped sequence of points represents the GPS trajectories in this dataset. Each point contains information on latitude, longitude, and altitude. This dataset has 17,621 trajectories with a total distance of 1,292,951 kilometers and a total duration of 50,176 hours. The trajectories are updated at a frequency of every 1 ~ 60 seconds. The Gowalla dataset is collected by Stanford University and is a location-based social networking site where users can share their location information by signing in. The dataset collects a total of 6,442,890 check-in locations and 19,651 check-in information. The check-in data is used to train the Markov model. We implement the proposed model by the following steps.

Step 1. Input the training dataset (Gowalla) to train the Markov model and output the prediction results.

Step 2. If the prediction results we obtain from Step 1 did not satisfy the privacy threshold θ, then we need to generate the special location set by our mechanism at the current timestamp. Otherwise, move to the next timestamp and continue Step 1.

Step 3. Check whether the real location is in the special region or not. In the process of generating a special region, we have a small probability of filtering out the true location. So we use the nearest and the most similar location as a puppet in Δr instead of it.

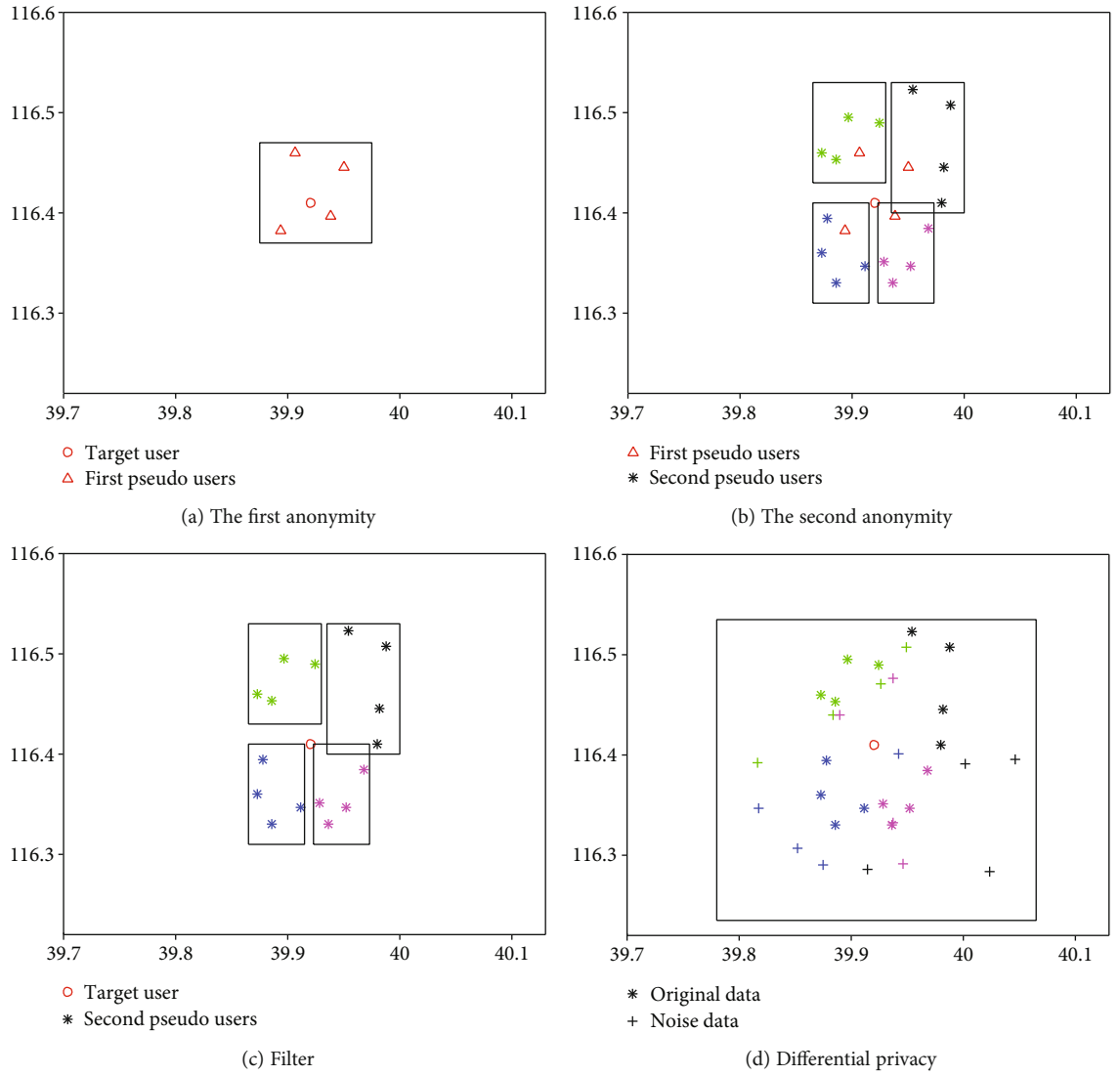


Figure 6: Differential privacy and k-anonymity.

Step 4. Use the special region R_t in each timestamp to divide the real-world map into several neighboring grids. In each grid, the adversaries cannot distinguish between the target user and the pseudo user.

Step 5. Finally, we add noises to the Markov model. In each timestamp, we add noise to the transfer mechanism to make the trajectories indistinguishable.

The performance of the release mechanism as a user moves over time is explained as follows. We treat our release mechanism at each timestamp with $\varepsilon = 1$. Each method is run over 20 times and shows outstanding performance. Figure 6(a) shows how to hide a user's accurate location by the first anonymity. The X-axes and Y-axes represent the longitude and latitude, respectively. The symbol "$\circ$" denotes the user's true location. "Δ" are the pseudo locations generated in the first anonymity. We have the SSE (sum of the squared errors) as the core indicator for the selection of k

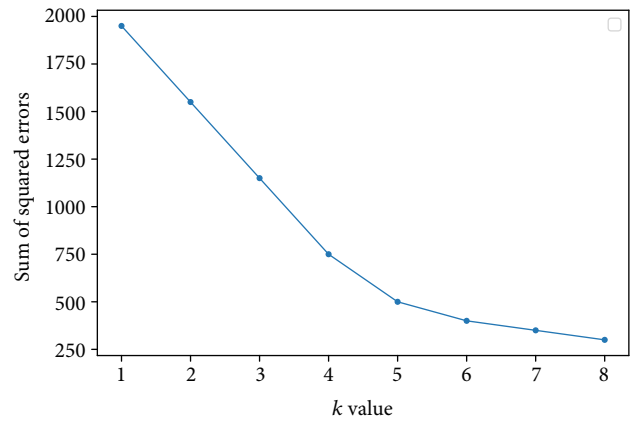
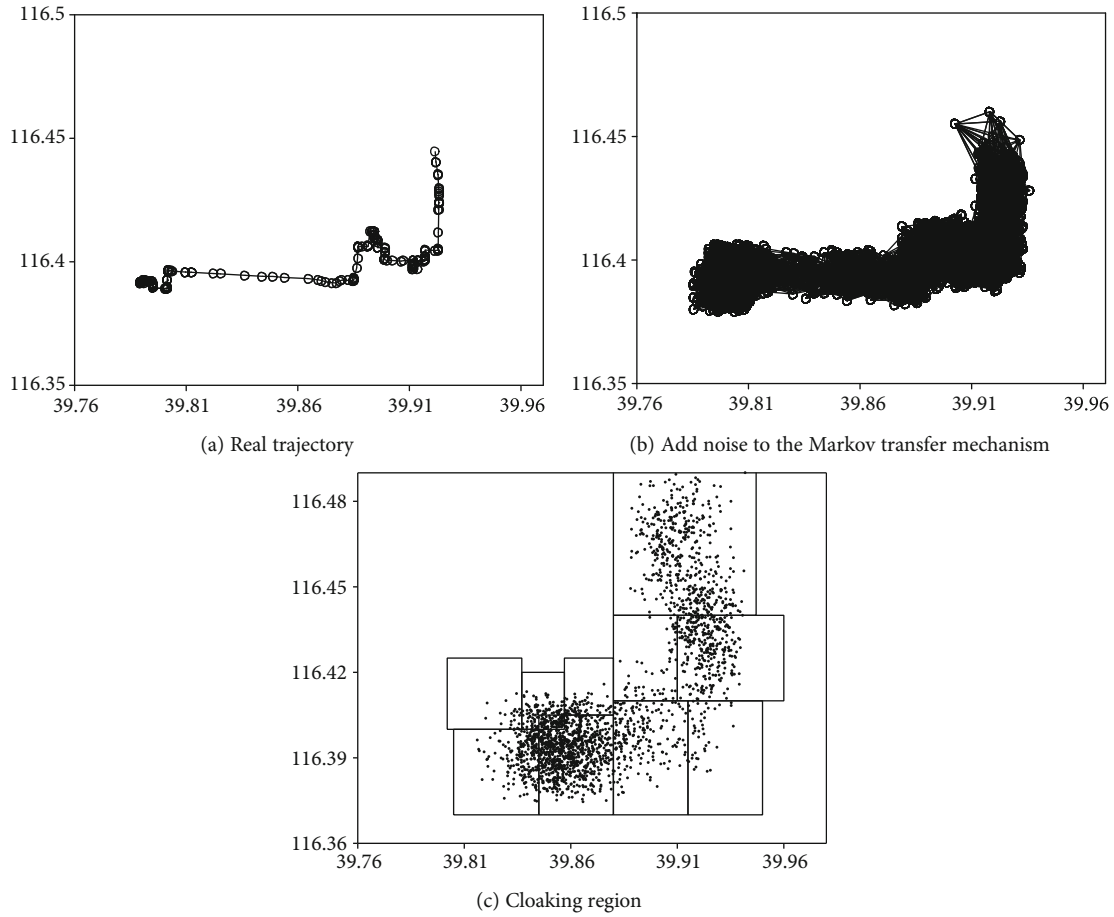


Figure 7: k size vs. sum of the squared errors.



(a) Real trajectory

(b) Add noise to the Markov transfer mechanism

(c) Cloaking region

Figure 8: Compared with the real trajectory.

value. As the k value increases, the sample division gets more refined, and the degree of aggregation of each cloaking region gradually increases. Then, the SSE naturally gradually becomes smaller. In our method, we set the parameter $k = 5$. We have experimented for many times that the parameter k size is better than others, as shown in Figure 7.

We choose four users who are the most similar to the real user and sent the same query at timestamp t. In Figure 6(b), we can see the second k-anonymity after Figure 6(a). In the second anonymity, we consider the first four users generated by the first anonymity as the "real" user, respectively. Then, the model generates more anonymous users by these four "real" users. In Figure 6(b), the symbol "$*$" denotes the pseudo users generated in the second anonymity. Through these two operations, we obtained a much bigger anonymous area with many similar anonymous users. Next, through the "Razor Mechanism" in Section 4, the model filters out several pseudo users generated from the first k-anonymity with "Razor," as shown in Figure 6(c). In this part, we make full use of the Jaccard Razor. While using the "Razor Mechanism," the actual user may be filtered out with a very small probability, which is known as the "drift" phenomenon.

When a "drift" happens by a minuscule probability, we use a surrogate user in ΔR_t to impersonate the target user.

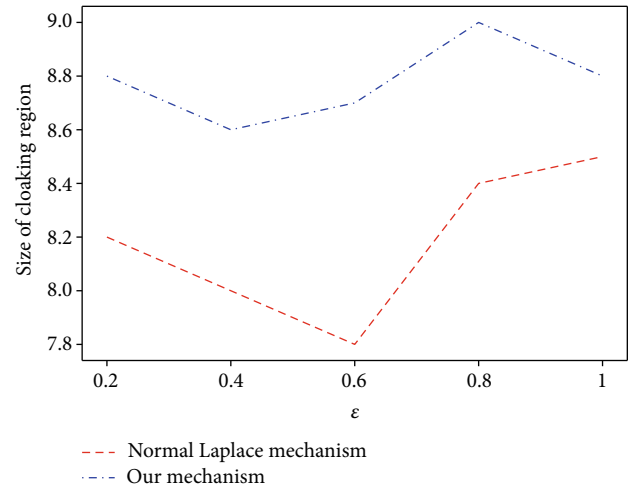
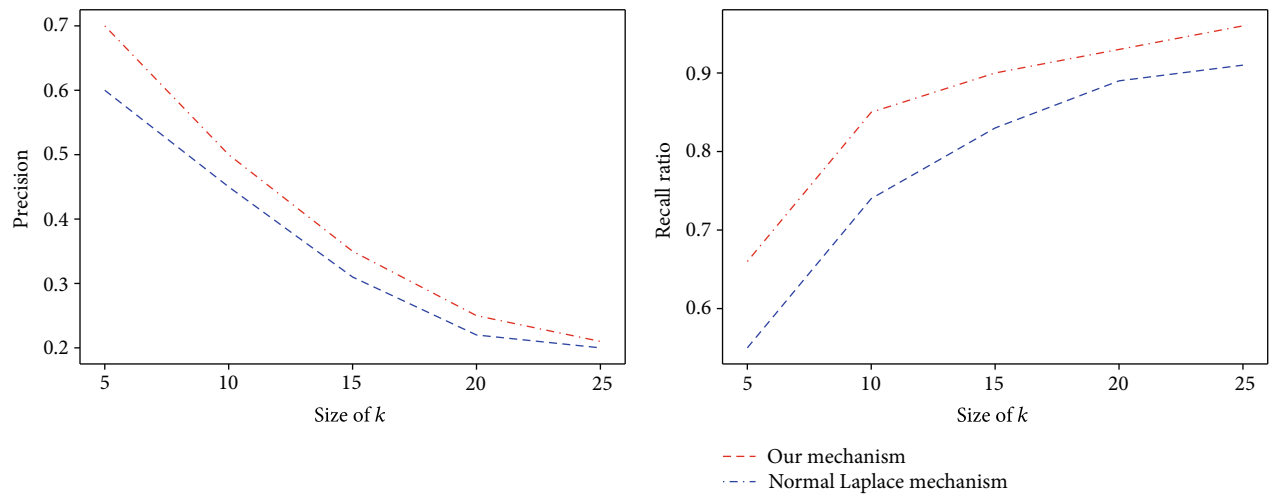


Figure 9: Size of the cloaking region vs. ε.

In the next step, the model adds Laplace noise to this special region ΔR_t at timestamp t, which can provide a rigorous privacy guarantee. We then obtain a new cloaking region that contains the pseudo location and the true location. As shown in Figure 6(d), the area within the square is one of the grids in the real-world maps. The symbol "$*$"



(a) Precision with the size of k

(b) Recall with the size of k

Figure 10: KNN results.

denotes the pseudo users generated in the second anonymity. Finally, the model generates a noisy cloaking region. The added Laplace noise makes the new special region very stable. Noisy users are always around the real user and fake users. In this special region, the adversaries cannot distinguish the target user and the pseudo users. In Figure 8, the true trajectory is compared with the trajectories with noise added to the Markov transfer mechanism. Figure 8(a) is a randomly selected accurate trajectory of a single user in a period of time. The user's movement is shown in Figure 8(a). The most important step in the algorithm is the addition of Laplace noise to the Markov transfer mechanism. This process makes the transfer probability stable. The noisy trajectories after adding noise with the Laplace mechanism are presented in Figure 8(b). As shown in Figures 8(a) and 8(b), the released trajectory is still close to the accurate trajectory. The special regions at every timestamp t are used to divide the real-world map into grids, as shown in Figure 8(c). In this region, the adversaries' side knowledge no longer affects privacy protections. The adversaries cannot either distinguish the accurate trajectory in the released trajectories or recognize the target user in these special regions ΔR_t at each timestamp. Our mechanism is the better one comparing to the normal Laplace mechanism, as shown in Figure 9.

To show the practicality of the release location area, we measure the query accuracy and recall rate of k nearest neighbors for every 500 timestamps in 150 trajectories, as shown in Figure 10. In Figure 10(a), it shows that the precision declines when k rises because when k grows, the nearest neighbors have to be found in larger areas. And a larger location set returned. On the other hand, Figure 10(b) indicates that the recall ratio increases with greater k. Figure 8 shows the comparisons of experiment results by the proposed method's position release mechanism in this paper and those by the Laplace mechanism. The results indicate that the usability of our method is better than that of the Laplace mechanism.

7. Conclusion and Future Work

In this paper, we proposed a L&A-indistinguishable system under temporal correlation. The system uses the Markov model to denote users' movement on the road network and then generates a special user set by k-anonymity and differential privacy approaches. The proposed system can provide perfect privacy protection for a single moving user. The method is based on the hidden Markov model and learns from historical trajectories to obtain prediction results for the future timestamp.

As a direction for future work, we are interested in instantiating the system with different and more advanced mobility models and researching the impact on the system's performance change. We look forward to making the mobile user's personal information protected with a more rigorous privacy guarantee with a smaller loss in data utility. We aim to conduct more profound research to enhance the availability of the region release mechanism based on the existing research studies. We plan to develop a model to recommend points of interest, based on the user's movement position information, and to recommend the community to which the user may move.

Data Availability

The coordinate data used to support the findings of this study have been deposited in the GeoLife dataset repository [44]. The check-in data used to support the findings of this study have been deposited in the Gowalla dataset repository [45].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Our research fund is supported by the National Natural Science Foundation of China (Grant Nos. 61472096, 61272186,

61472095, and 61502410), 2019 Industrial Internet Innovation and Development Engineering, Industrial Internet Security Audit Technology and Product, On-site Emergency Detection Tools in the Field of Industrial Internet Security (KY10600200008, KY10600200021), International Governance Research Center of industrial Internet (3072020CFP0601), and Fundamental Research Funds for the Central Universities (3072020CF0604).

References

[1] Z. Cai and Q. Chen, "Latency-and-coverage aware data aggregation scheduling for multihop battery-free wireless networks," *IEEE Transactions on Wireless Communications*, p. 1, 2020.

[2] X. Zheng, Z. Cai, J. Li, and H. Gao, "Location-privacy-aware review publication mechanism for local business service systems," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, 2017.

[3] J. Freudiger, R. Shokri, and J.-P. Hubaux, "Evaluating the privacy risk of location-based services," in *Proceedings of the 15th International Conference on Financial Cryptography and Data Security, FC'11*, pp. 31–46, Berlin, Heidelberg, 2012.

[4] J. Freudiger, S. Rane, A. E. Brito, and E. Uzun, "Privacy preserving data quality assessment for high-fidelity data sharing," in *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security, WISCS '14*, pp. 21–29, 2014.

[5] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, MobiSys '03*, pp. 31–42, New York, NY, USA, 2003.

[6] Z. Xiong, Z. Cai, Q. Han, A. Alrawais, and W. Li, "Adgan: protect your location privacy in camera data of auto-driving vehicles," *IEEE Transactions on Industrial Informatics*, p. 1, 2020.

[7] S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest neighbor search with strong location privacy," *Proceedings of the VLDB Endowment*, vol. 3, 2010no. 1-2, pp. 619–629, 2010.

[8] C.-Y. Chow, M. F. Mokbel, and W. G. Aref, "Casper: query processing for location services without compromising privacy," *ACM Transactions on Database Systems*, vol. 34, pp. 1–48, 2009.

[9] P. Samarati, "Protecting respondents' identities in microdata release," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, pp. 1010–1027, 2001.

[10] X. Zheng and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial IoTs," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, 2020.

[11] C. Dwork, "Differential privacy in new settings," *Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '10*, 2010, pp. 174–183, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2010.

[12] Y. Zak and A. Even, "Development and evaluation of a continuous-time Markov chain model for detecting and handling data currency declines," *Decision Support Systems*, vol. 103, pp. 82–93, 2017.

[13] G. Theodorakopoulos, R. Shokri, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Prolonging the hide-and-seek game: optimal trajectory privacy for location-based services," *Proceedings of the 13th Workshop on Privacy in the Electronic Society, WPES '14*, 2014, pp. 73–82, ACM, New York, NY, USA, 2014.

[14] J. Krumm, "A survey of computational location privacy," *Personal and Ubiquitous Computing*, vol. 13, no. 6, pp. 391–399, 2009.

[15] J. Wang, Z. Cai, and J. Yu, "Achieving personalized k-anonymity-based content privacy for autonomous vehicles in CPS," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4242–4251, 2020.

[16] P. Shankar, V. Ganapathy, and L. Iftode, "Privately querying locationbased services with SybilQuery," in *Proceedings of the 11th International Conference on Ubiquitous Computing, UbiComp '09*, pp. 31–40, New York, NY, USA, 2009.

[17] D. Nussbaum, M. T. Omran, and J.-R. Sack, "Techniques to protect privacy against inference attacks in location based services," *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on GeoStreaming, IWGS '12*, 2012, pp. 58–67, ACM, New York, NY, USA, 2012.

[18] B. Bamba, L. Liu, P. Pesti, and T. Wang, "Supporting anonymous location queries in mobile environments with PrivacyGrid," in *Proceedings of the 17th International Conference on World Wide Web, WWW '08*, pp. 237–246, New York, NY, USA, 2008.

[19] J. Cuellar, M. Ochoa, and R. Rios, "Indistinguishable regions in geographic privacy," *Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC '12*, 2012, pp. 1463–1469, ACM, New York, NY, USA, 2012.

[20] M. Xue, P. Kalnis, and H. K. Pung, "Location diversity: enhanced privacy protection in location based services," *Proceedings of the 4th International Symposium on Location and Context Awareness, LoCA'09*, 2009, pp. 70–87, Springer-Verlag, Berlin, Heidelberg, 2009.

[21] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS '12*, pp. 617–627, New York, NY, USA, 2012.

[22] M. Herrmann, C. Troncoso, C. Diaz, and B. Preneel, "Optimal sporadic location privacy preserving systems in presence of bandwidth constraints," *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society, WPES '13*, 2013, pp. 167–178, ACM, New York, NY, USA, 2013.

[23] Z. Cai, X. Zheng, and J. Yu, "A differential-private framework for urban traffic flows estimation via taxi companies," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6492–6499, 2019.

[24] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2020.

[25] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: theory meets practice on the map," in *2008 IEEE 24th International Conference on Data Engineering*, pp. 277–286, Cancun, Mexico, 2008.

[26] R. Chen, B. C. Fung, B. C. Desai, and N. M. Sossou, "Differentially private transit data publication: a case study on the Montreal transportation system," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12*, pp. 213–221, New York, NY, USA, 2012.

[27] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, 2008.

[28] H. Li, L. Xiong, L. Zhang, and X. Jiang, "DPSynthesizer," *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, vol. 7, 2014no. 13, pp. 1677–1680, 2014.

[29] N. Li, W. Yang, and W. Qardaji, "Differentially private grids for geospatial data," in *Proceedings of the 2013 IEEE International Conference on Data Engineering (ICDE 2013), ICDE '13*, pp. 757–768, Washington, DC, USA, 2013.

[30] S.-S. Ho and S. Ruan, "Differential privacy for location pattern mining," *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS, SPRINGL '11*, 2011pp. 17–24, New York, NY, USA, 2011.

[31] R. Dewri, "Local differential perturbations: location privacy under approximate knowledge attackers," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2360–2372, 2013.

[32] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 577–590, 2018.

[33] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the Third Conference on Theory of Cryptography, TCC'06*, pp. 265–284, Berlin, Heidelberg, 2006.

[34] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," *Proceedings of the 2011 IEEE Symposium on Security and Privacy, SP '11*, 2011, pp. 247–262, IEEE Computer Society, Washington, DC, USA, 2011.

[35] M. Götz, S. Nath, and J. Gehrke, "MaskIt: privately releasing user context streams for personalized mobile applications," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD '12*, pp. 289–300, New York, NY, USA, 2012.

[36] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, "Learning and inferring transportation routines," *Artificial Intelligence*, vol. 171, no. 5-6, pp. 311–331, 2007.

[37] Z. Cai and Z. He, "Trading private range counting over big IoT data," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pp. 144–153, Dallas, Texas, USA, 2019.

[38] Z. Cai and T. Shi, "Distributed query processing in the edge assisted IoT data monitoring system," *IEEE Internet of Things Journal*, p. 1, 2020.

[39] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: a survey of recent developments," *ACM Computing Surveys (Csur)*, vol. 42, pp. 1–53, 2010.

[40] M. E. Nergiz, M. Atzori, and Y. Saygin, "Towards trajectory anonymization: a generalization-based approach," *Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS, SPRINGL '08*, 2008, pp. 52–61, ACM, New York, NY, USA, 2008.

[41] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," in *Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08*, pp. 312–321, New York, NY, USA, 2008.

[42] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, pp. 791–800, New York, NY, USA, 2009.

[43] Y. Zheng, X. Xie, and W.-Y. Ma, "GeoLife: a collaborative social networking service among user, location and trajectory," *Data Engineering*, vol. 33, p. 32, 2010.

[44] https://www.microsoft.com/en-us/download/details.aspx?id=52367.

[45] http://snap.stanford.edu/data/loc-Gowalla.html.