TrPF: A Trajectory Privacy-Preserving Framework for Participatory Sensing

Sheng Gao, Jianfeng Ma, Weisong Shi, *Senior Member, IEEE*, Guoxing Zhan, and Cong Sun

***Abstract*—The ubiquity of the various cheap embedded sensors on mobile devices, for example cameras, microphones, accelerometers, and so on, is enabling the emergence of participatory sensing applications. While participatory sensing can benefit the individuals and communities greatly, the collection and analysis of the participators' location and trajectory data may jeopardize their privacy. However, the existing proposals mostly focus on participators' location privacy, and few are done on participators' trajectory privacy. The effective analysis on trajectories that contain spatial-temporal history information will reveal participators' whereabouts and the relevant personal privacy. In this paper, we propose a trajectory privacy-preserving framework, named TrPF, for participatory sensing. Based on the framework, we improve the theoretical mix-zones model with considering the time factor from the perspective of graph theory. Finally, we analyze the threat models with different background knowledge and evaluate the effectiveness of our proposal on the basis of information entropy, and then compare the performance of our proposal with previous trajectory privacy protections. The analysis and simulation results prove that our proposal can protect participators' trajectories privacy effectively with lower information loss and costs than what is afforded by the other proposals.**

***Index Terms*—Participatory sensing, trajectory privacy-preserving framework, trajectory mix-zones graph model, information loss, entropy.**

I. Introduction

With the development of wireless communication technologies, such as WLAN, 3G/LTE, WiMax, Bluetooth, Zigbee, and so on, mobile devices are equipped with a variety of embedded sensors surveyed in [1] as well as powerful sensing, storage and processing capabilities. Participatory sensing [2] (urban sensing [3]), which is the process that enables individuals to collect, analyze and share local knowledge with their own mobile devices, emerges as required under these well conditions. Compared with WSNs, participatory sensing offers a number of advantages on deployment costs, availability, spatial-temporal coverage, energy consumption and so forth. It has attracted many researchers in different areas such as Intelligent Transportation System, healthcare and so on. There are lots of existing prototype systems that include *CarlTel* [4], *BikeNet* [5], *DietSense* [6], *PEIR* [7] and so on.

Nowadays, participatory sensing applications mainly depend on the collection of data across wide geographic areas. The sensor data uploaded by participators are invariably tagged with the spatial-temporal information when the readings were recorded. According to the analysis in [8], the possible threats to a participator's privacy information that include monitoring data collection locations, tracing his/her trajectory, taking photographs of private scenes and recording the intimate chat logs. Once participators realize the serious consequences with the disclosure of their sensitive information, they are unwilling to participate in the campaign and use the services. Since the success of participatory sensing campaign strongly depends on the altruistic process of data collection, if the participators are reluctant to contribute their collected data, it would weaken the popularity and impact of this campaigns deployed at large scale while also reducing the benefits to the users. Therefore, the privacy problems are the significant barriers to data collection and sharing. How to ensure the participators' privacy is the most urgent task.

In typical participatory sensing applications, the uploaded data reports may reveal participators' spatial-temporal information. Analysts could obtain some valuable results from the published trajectories for decision making, for example, merchants may decide where to build a supermarket that can produce maximum profit by analyzing trajectories of customers in a certain area and the Department of Transportation can make an optimized vehicle scheduling strategy by monitoring trajectories of vehicles. However, it may introduce serious threats to participators' privacy. Adversary may possibly analyze the trajectories which contain rich spatial-temporal history information to link multiple reports from the same participators and determine certain private information such as the places where the data reports are collected. Thus, it is necessary to unlink the participators' identities from sensitive data collection locations. To best of our knowledge, existing work on privacy in participatory sensing mainly concentrate on data contribution and reporting process [9]–[12]. If an adversary has *a priori* knowledge of a participator's trajectory, it is effortless to deanonymize his/her reports.

Manuscript received May 24, 2012; revised January 24, 2013; accepted February 27, 2013. Date of publication March 15, 2013; date of current version May 16, 2013. This work was supported by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1078), by the Key Program of NSFC-Guangdong Union Foundation (U1135002), by the Major National S&T Program (2011ZX03005-002), and by the Fundamental Research Funds for the Central Universities (JY10000903001, K5051203010). The work of W. Shi was supported in part by the Introduction of Innovative R&D team program of Guangdong Province (201001D0104726115). The associate editor coordinating the review of this manuscript and approving it for publication was Sen-Ching Samson Cheung.

S. Gao, J. Ma, and C. Sun are with the School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China (e-mail: sgao@mail.xidian.edu.cn; jfma@mail.xidian.edu.cn; suncong@xidian.edu.cn).

W. Shi and G. Zhan are with the Department of Computer Science, Wayne State University, Detroit, MI 48202 USA (e-mail: weisong@wayne.edu; gxzhan@wayne.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2013.2252618

In this paper, we propose a trajectory privacy-preserving framework, named TrPF, for participatory sensing. We observe that the locations on or nearby participators' trajectories may not all be sensitive, and with this thought, our proposal only deals with the sensitive trajectory segments that will be discussed in the following. Moreover, mix-zones are regions [13], [14] where no applications can track participators' movements. Some works [15], [16] focused on road network mix-zones, which are not applicable in participatory sensing. For one thing, they all build mix-zones at road intersection, which may restrict the random data collection time and the number of ingress/egress locations; for another thing, the trajectory segments at the road intersection may not be sensitive, while the others would be. Therefore, we improve the theoretical mix-zones model [13], [14] to construct trajectory mix-zones model for protecting sensitive trajectory segments from the perspective of graph theory. Compared with existing trajectory privacy-preserving proposals, our proposal has advantages of lower costs and information loss while the privacy level would not decrease.

In this paper, the main contributions of our work are summarized as follows:

- We propose a framework TrPF of participatory sensing for trajectory privacy protection;
- We improve the theoretical mix-zones model with considering time factor from the perspective of graph theory to construct trajectory mix-zones model for protecting participators' sensitive trajectory segments;
- We formalize privacy level metric, privacy loss metric and information loss metric, and then analyze the attack models with different background knowledge;
- Compared with previous trajectory privacy protections, we run a set of simulation experiments to evaluate the effectiveness of our proposals and then make a comparison of the performance.

The remainder of this paper is organized as follows. Section II presents the related work. In Section III, we propose the trajectory privacy-preserving system framework TrPF for participatory sensing. In Section IV, we present two algorithms, called *GraphConstruct* and *WeightConstruct* to construct Trajectory Mix-zones Model from the perspective of graph theory. In Section V, we measure the privacy in terms of the privacy level metric and the privacy loss metric based on the information entropy and then define the information loss metric. In Section VI, we analyze two kinds of attack models with different background knowledge. In Section VII, we run a set of simulations to evaluate the effectiveness of our proposal, and then compare the system performance with previous work. Finally, we conclude this paper and present the future work in Section VIII.

II. Related Work

In this section, we discuss the current state of the art of privacy-preserving techniques in participatory sensing. Kapadia *et al.* [17] proposed the first implementation of a privacy-aware architecture, named AnonySense, for the anonymous task allocation and data reporting. From the perspective of cryptography, De Cristofaro *et al.* [18] analyzed the realistic architectural assumptions and privacy requirements, and then provided an instantiation that achieved privacy protection in participatory sensing with provable security. Christin *et al.* [19] surveyed the privacy and security implications in three types of application scenarios. In [8], [20], they analyzed the privacy challenges in participatory sensing applications in detail. Chow *et al.* [21] surveyed the privacy protection in terms of data privacy protection, location privacy protection and trajectory privacy protection in location-based services. Liu [22] reviewed the definitions, the models and the appropriate location privacy protection techniques from the perspective of mobile data management.

A. Location Privacy Protection

There are several works [23]–[25] that survey the location privacy-preserving schemes. We classify them into the following aspects.

1) Dummy Locations: Mechanisms proposed in [26], [27] mainly employ the idea of dummy locations to protect a user's location privacy. Our previous work [28] focused on the tradeoff between location and trajectory privacy protection and QoS based on the dummy events.

2) Location k-Anonymity: Much of the work regarding location privacy protection derive from *k-anonymity* model which is first proposed by Sweeney in database [29]. For example, spatial and temporal cloaking on the basis of this model to protect location privacy was first proposed by Gruteser and Gruwald [30]. Take the individuals' requirements on location privacy into consideration, Gedik and Liu [31] proposed a scalable architecture for location privacy protection. Bettini *et al.* [32] proposed a formal framework to protect a user's anonymity when requesting location-based services. They provided the safeguards that were specific for different kinds of knowledge available to attacker.

3) Obfuscation: Duckham *et al.* [33] proposed to protect a user's location privacy by deliberately degrading the accuracy of his/her spatial-temporal information. Obfuscation is a class of the important approaches in location privacy. Much of the work that belong to it can be enforced through perturbation [12] or generalization [34].

4) Mix-Zones: Pseudonym [35] is used to break the linkage between a user's identity and his/her events. The process of its change is usually performed in some pre-determined areas called *mix-zones* [13], [14] and the idea of building mix-zones at road intersections has been proposed in [15], [16]. The problems of optimal placement of mix-zones are studied in [36], [37], where rectangular or circular shaped zones that commonly used by these mix-zones techniques. To best of our knowledge, only [15], [16] take into account the effect of timing attack in the construction process. However, they may not be applicable in participatory sensing for the constraints at road intersections. In this paper, we take the time interval into consideration and improve the theoretical mix-zones model from the perspective of graph theory to protect the data collectors' trajectories privacy in participatory sensing.

B. Trajectory Privacy Protection

We realize that once a user's trajectory is identified, the user's locations are exposed. Some works [21], [38] have summarized the trajectory privacy protection techniques, where the most immediate and simple ways are dummy trajectories and suppression technique. For example, You *et al.* [39] proposed to produce a user's dummy trajectories through *random pattern* and *rotation pattern*. To be specific, the former generated a dummy trajectory randomly from the staring point towards the destination and the later did it by rotating the user's trajectory. However, the trajectory similarity may affect the anonymity quality. Thus, how to generate dummy trajectories that look like a normal user's trajectory is one of the main challenges of this kind of work. To prevent adversary from inferring a user's unknown locations by using his/her partial trajectory knowledge, Terrovitis *et al.* [40] proposed a location suppression technique to convert a database of trajectories, which can prevent the disclosure of the user's whole trajectory with high probability. However, those trajectory segments that are suppressed would cause the collected data lost.

Trajectory *k-anonymity* that extends from location *k-anonymity* is widely used in trajectory privacy protection. For convenience, we only address some typical and recent studies. Nergiz *et al.* [41] proposed to group the trajectories based on log cost metric and then enforce *k-anonymity* on each sample location. Finally, a random reconstruction method was presented to enhance anonymized trajectory privacy further. Motivated by the inherent uncertainty of localization, Abul *et al.* [42] proposed the concept of (k, δ)-*anonymity* for mobile object databases, where δ represented the possible location imprecision. Then, they proposed *Never Walk Alone (NWA)* to achieve (k, δ)-*anonymity* through clustering and space translation. Specifically, when $\delta = 0$, it had degenerated into the traditional micro-aggregation that replaced trajectories by the trajectories clustering center over the same time interval. To anonymize mobile objects with dynamic sensitive attributes, Yarovoy *et al.* [43] presented to achieve the new notion of k-anonymity they proposed for mobile objects through *extreme-union* and *symmetric anonymization*. Xu *et al.* [44] exploited historical locations to construct trajectory *k-anonymity* and then presented algorithms for spatial cloaking. In a recent study, Huo *et al.* [38] investigated the selection of trajectory k-anonymity sets based on graph partition. In the follow-up work, they proposed a method called *You Can Walk Alone (YCWA)* [45] to improve *NWA* by anonymizing the stay points that were extracted efficiently on people's trajectories. They generated *k-anonymity* zone based on two algorithms called *grid-based approach* and *clustering-based approach*.

As mentioned above, we can see that dummy trajectories and almost all the trajectory anonymity techniques deal with the whole trajectory, which result in the increase of costs such as computation, storage and query with a certain privacy level. Sensitive location suppression technique may reduce the overhead costs with a same privacy level. However, if the sensitive locations on trajectories are suppressed too much, it might cause lots of information loss. We observe that not all the locations on the trajectory are sensitive. There have been some work [45]–[48] to analyze the sensitive locations on or nearby the published trajectory. For example, Zheng *et al.* [48] proposed a method to find interesting locations and frequent travel sequences in a given geographic region. Palma *et al.* [46] proposed a method of clustering-based stops and moves of trajectories to compute important places based on the change of the speed of the trajectory. However, privacy is rarely considered in these work. Monreale *et al.* [47] distinguished the semantics of the visited place between sensitive and quasi-identifier places and proposed a algorithm for generalizing the visited place based on taxonomy. Huo *et al.* [45] improved the stay point extraction strategies proposed in [48], and proposed duration-based strategy and density-based strategy to extract the sensitive stay point. To overcome the defects above, we propose a preferable trajectory privacy protection method to reduce the costs and information loss; meanwhile the privacy level will not decrease.

III. Overview of TrPF System

In this section, we firstly depict the trajectory privacy-preserving framework TrPF for participatory sensing, and then emphasize the privacy problem with the disclosure of users' trajectories. Finally, we define some basic notions.

A. The Architecture of TrPF for Participatory Sensing

Mix Network functions as an anonymizing intermediary between Mobile Nodes and the Report Server that is widely used in [9], [17], [49]. Take [49] for example, it routes reports via multi-hop transmission, adding delays and mixing with the data from other sources to other destinations. Such process makes adversary can neither link a mobile node's reports together nor identify which mobile node sent the report, or learn when and where the reports were reported. Based on [49], we propose a trajectory privacy-preserving framework TrPF for participatory sensing system depicted as Fig. 1. Compared with the previous architecture, we consider the factor of participators' privacy and substitute the mix network in [49] with a Trusted Third Party Server component. Due to the removal of mix network, it will optimize the data reports transmission. The addition of Trusted Third Party Server can function as a privacy-preserving agent, which can trade off the efficiency of data transmission and privacy protection. It can reduce the network hops of data reports transmission route via wireless network. According to the different roles of function characteristics, the main components of TrPF are made up of the following entities.

1) Data Collectors: Mobile Nodes are devices with the capabilities of sensing, computation, memory and wireless communication, which act as data collectors in participatory sensing system. They can be used for context-aware data capture and carried along with each participator. Note that the involvement of data collectors in this sensing campaign is voluntary. Any participator who wants to provide application server with shared data needs to obtain a certificate from Trusted Third Party Server. To prevent adversary from disguising as a legitimate participator to upload malicious data, only the one who has been validated can access the participatory sensing system and upload his/her collected data reports.

We formalize the data reports collected by participator P_i as: $R_i = \langle ID_{P_i}, Data, Location, Time \rangle$, where ID_{P_i} represents the identity of P_i, *Location* and *Time* are the spatial-temporal

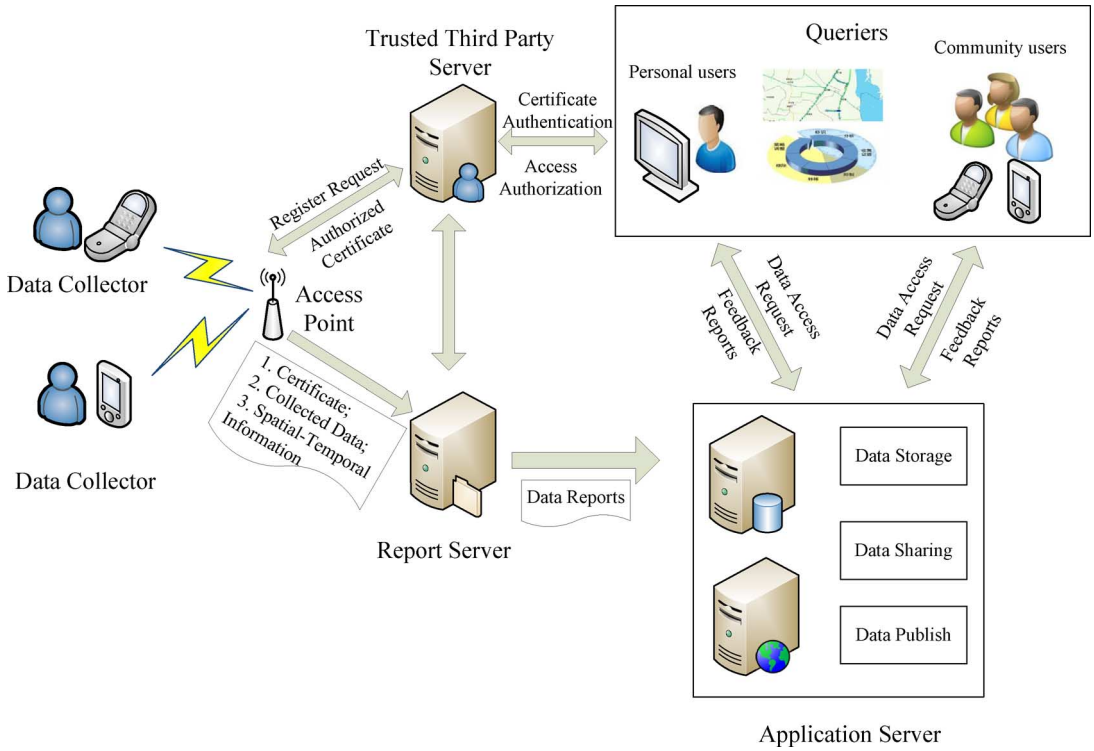

Fig. 1. Architecture of TrPF for participatory sensing.

information tagged with the collected data that compose trajectories of data collectors.

2) Trusted Third Party Server (TTPs): To ensure system security and participators' privacy, TTPs stores participators' relevant information such as certificates and pseudonyms information. Certificates are used for verifying participators' validity so as to exclude malicious attacker. The disclosure of the spatial-temporal information may also threaten the participators' privacy. We remove the linkage between the participators' spatial-temporal information and their identities based on pseudonym technique. It will be discussed in Section IV.

3) Report Server: Report Server is responsible for dealing with two aspects: (a) Interact with TTPs to verify the validity of the participators' identities by the certificates contained in the data reports; (b) Simplify the uploaded data reports such as data aggregation, and then send the data reports to Application Server.

4) Application Server: Application Server acts as a data center. It can provide kinds of data services for end users and play the following roles: (a) Data Storage: store the processed data reports received from data report server; (b) Data Sharing: any legitimate end user can access the available data services; (c) Data Publish: publish the data reports for the end users to query.

However, in our system architecture, Application Server may be untrustworthy. It may leak participators' sensitive information to adversary. For example, the disclosure of participators' trajectories may indicate where the data reports are collected. Maybe some of the locations such as home address are sensitive. Adversary can use the published trajectories to link participators' data reports with sensitive locations. As a result, the participators are aware that their privacy might be invaded seriously so that they may not want to share their collected data reports with end users.

5) Queriers: Queriers are end users that request sensor reports in a given participatory sensing application, which can be personal users or community users. They access and consult the data gathered by the data collectors according to their requirements. The queriers include, for example, data collectors are intending to consult their own collected data, doctors checking their patients' records, environmentalists querying the climate data of a certain area or the general public for other purposes. Note that only the registered end users can access the shared data reports. End users send certificate authentication requests to TTPs. Anyone who has registered before can get the access authorization and only the valid end users can access the shared data reports that are provided by data collectors.

B. Problem Statement

In participatory sensing system, data reports collected by participators are tagged with spatial-temporal information. Since the location information that attached to the collected data reports are commonly shared, a prominent attack is thus the *Trajectory Inference*. For example, suppose an adversary learns through background knowledge that a data collector P_i has visited a specific location at a certain time t_i, while the location happens to be the only sample on P_i's trajectory at time t_i in the data reports. The adversary would synthesize this information to infer the whole trajectory of P_i, which may relate to a certain sensitive attribute. Additionally, the analysis of trajectories over several data reports may help adversary to exploit the frequently visited locations and reveal participators' identities, e.g., a data collector P_i usually spends the same time on arriving at a specific location from a fixed location everyday in the

morning. Adversary can use the frequent information to deduce the starting location in the morning may be his/her home and the location reached after the time may be the work place. Consequently, the participators' privacy would suffer a huge threat with the disclosure of sensitive locations.

To prevent from linking participators' identities with their uploaded data reports, we propose a method to protect participators' identities and trajectories privacy from the perspective of graph theory based on mix-zones model and pseudonym technique. In fact, only part of the locations on or nearby their trajectories are sensitive mentioned in Section II. On this basis, we only need to protect the sensitive parts of participators' trajectories in their collected data reports.

C. Basic Notion

In the following, we first formalize trajectory presentation and then define sensitive area and sensitive trajectory segment.

Definition 3.1 (Trajectory Presentation): A participator P_i, whose trajectory Tr can be considered as a set of discrete locations at sampling time in three-dimensional space, represented as: $Tr = \{ID_{P_i}, (x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_n, y_n, t_n), t_1 < t_2 < \cdots < t_n\}$, where ID_{P_i} represents the trajectory identity of participator P_i, (x_i, y_i, t_i) represents participator P_i's sample location at time t_i on or nearby the trajectory.

Definition 3.2 (Sensitive Area): Suppose the sensitive location is $o = (x_i, y_i)$ on or nearby a participator P_i's trajectory, we define the sensitive area S_i is a horizontal disk with center at o, radius r, which is formalized as $S_i = \{ID_{P_i}, s_i \,|\, (x_i, y_i) \in S(o, r)\}$.

Definition 3.3 (Sensitive Trajectory Segment): A piece of trajectory segment of P_i's trajectory is called sensitive trajectory segment when it gets through the sensitive area. Suppose the sensitive area S_i includes the sensitive location $o = (x_i, y_i)$, we define the sensitive trajectory segment as $TF_i = \{ID_{P_i}, (x_i, y_i) \,|\, \{(x_{i-1}, y_{i-1}), (x_{i+1}, y_{i+1})\} \subseteq S_i, i = 2, 3, \ldots, n-1\}$. TF is the set of all sensitive trajectory segments depicted as $TF = \{TF_i, 0 \leq i \leq n\}$. Specially, when $i = 0$, that is $TF = \emptyset$, it means there is no sensitive trajectory segment; while $i = n$ means the whole trajectory is sensitive.

To illustrate the concepts above intuitively, we describe them as Fig. 2. The whole space is composed by three-dimensional coordinate, where (X, Y) represents the trajectory geographic information and T represents the time information. As we can see, the red circle represents the sensitive location, the cylinder represents sensitive area. The trajectory segments are sensitive when they get through the sensitive area. For example, in road traffic condition, dangerous areas are set up essential warning sign to warn the vehicles that get through these areas carefully. The dangerous areas can be considered as sensitive area and the traffic routes of the vehicles in the area can be seemed as sensitive trajectory segments. In participatory sensing, we consider all the trajectories of data collectors that get through the sensitive area are sensitive. To ensure the sensitive trajectory privacy, we construct the sensitive area defined by Definition 3.2

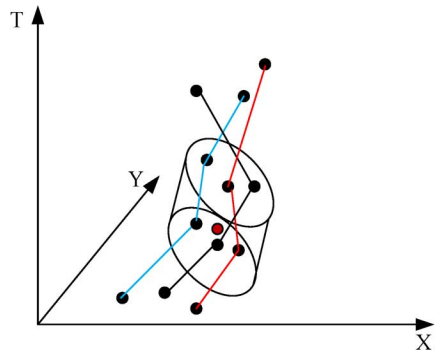

Fig. 2. Sensitive area and trajectory segment.

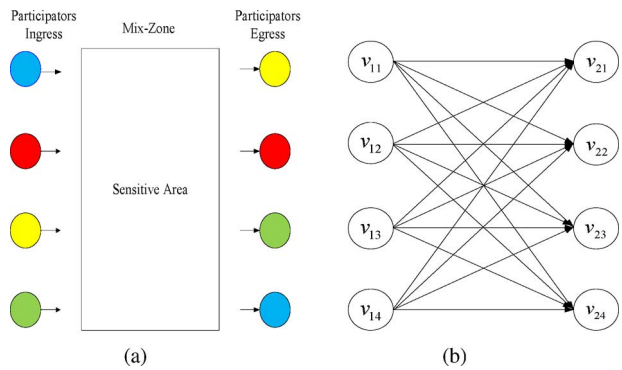

Fig. 3. Trajectory mix-zones graph model. (a) Trajectory mix-zones; (b) directed graph model.

and then divide the whole trajectory into the sensitive trajectory segments and nonsensitive trajectory segments.

IV. Trajectory Mix-Zones Graph Model

In this section, we propose to anonymize the sensitive trajectory segment from the perspective of graph theory. To reduce information loss and costs at a certain privacy-preserving level, we divide the whole area into several parts. According to the sensitive locations on or nearby the trajectories, we divide the whole trajectories into sensitive trajectory segments and nonsensitive trajectory segments. We only protect sensitive trajectory segments based on mix-zones model and pseudonym technique. Fig. 3 presents the Trajectory Mix-zones Graph Model. The trajectory mix-zones are described as Fig. 3(a). We abstract it into Directed Graph Model depicted by Fig. 3(b). The process will be discussed in Section IV-A in detail.

Any data collector who enters the Sensitive Area should select a pseudonym provided by TTPs to anonymize the linkability between his identity and his collected data reports. Meanwhile, they record their ingress and egress time. A participator's information we describe as a tuple: $I_i = (ID_{P_i}, R_i, S_i, t_{ingress}, \Delta t_{egress})$, where ID_{P_i} represents the participator P_i's pseudonym provided by TTPs, R_i is the mapping from participator's identity to his pseudonym, S_i is the sensitive area the participator passes by, $t_{ingress}$ presents the set of participators' enter time and Δt_{egress} is the participator's egress time interval.

A. Trajectory Graph Construction

We propose to model the Trajectory Mix-zones as *Directed Weighted Graph (DWG)*, which is formalized as $G = \{V, E\}$. V is the set of vertexes which are constructed by the pseudonyms provided by TTPs. A participator enters the sensitive area with a pseudonym and leaves it with another pseudonym. It can be depicted as $V = \{(v_{11}, v_{12}, \ldots, v_{1n}), (v_{21}, v_{22}, \ldots, v_{2n})\}$. E is the set of edges that represent the participators' trajectory mapping from the ingress to the egress in the sensitive area. In Fig. 3, as a result of pseudonym technique, there may be some difficulties for adversary to link the ingress and egress participator with the same identity.

In fact, *DWG* is a complete bipartite graph with different weights on each edge. The time of participators stays in mix-zones can either be constant or vary. Palanisamy *et al.* [16] analyzed the two different cases in road network. They pointed out that if the residence time was constant, it would encounter *First In First Out (FIFO)* attack. That is to say, the first exit participator corresponds to the first one that enters the mix-zones and the pseudonym technique takes no effect. In this paper, we assume that the arrivals of participators at the trajectory mix-zones follow a Poisson process. Given a time interval T, k participators enter the trajectory mix-zones with mean arrival rate λ to achieve *k-anonymity*. Note that the time interval and the arrival rate decide the number of participators that enters the trajectory mix-zones. Additionally, the participators' arrival time should not differ by a large value, or adversary could infer the first exit might correspond to the first enter.

The time of data collectors that spend in mix-zones is random. However, in [16], due to the constraints of road network mix-zones at road intersection, they cannot construct an effective mix-zone for the limited randomness on the time of participators that spent in it. In this paper, we argue that the data collection time of each participator follows normal distribution depicted by $\Delta'(t) \sim N(\mu, \sigma)$. Take the process of road traffic information collection for example, in the rush hours, the traffic roads are very congested, the data collection time will be more; otherwise, in other time, the data collection time may be less for its good road conditions. Thus, the data collection time is dependent on the conditions of road congestion. Since the time interval of data collection in sensitive area is random, even though adversary obtains the related information such as ingress and egress order and time, it cannot link the ingress pseudonym to egress pseudonym.

According to our discussions above, in this paper, participators enter and stay in sensitive area with random time interval. Even if adversary observes the time information, they cannot identify the pseudonyms mapping. Each edge weight represents the mapping probability between an ingress pseudonym and egress pseudonym. The construction process will be discussed in Section IV-B in detail.

To achieve *k-anonymity*, the graph vertexes are constructed by the participators' pseudonyms set of size k at least for each ingress and egress participator. The egress pseudonym should be different from the ingress pseudonym of the same participator to prevent adversary from identifying the trajectory linkage in the sensitive area for data collection. We call the *WeightConstruct* function to compute the weight of each edge. The process of the trajectory graph construction can be described by Algorithm 1 in detail.

Algorithm 1 GraphConstruct

Input: Trajectory Tr and Pseudonym set P

Output: Directed Weighted Graph (DWG)

```
Procedure
Sensitive trajectory segments set TF ← Tr;
for each TF_t ∈ TF do
  Construct a sensitive area S_t;
  Select k pseudonyms as vertexes to achieve
   k-anonymity;
  for each i < k do
   Random select ingress pseudonym P_i ∈ P;
   V_i ← P_i;
   Random select egress pseudonym P_j ∈ P;
   if P_j ≠ P_i then
     E_ij ← ⟨P_i, P_j⟩;
   else
    Select another different P_o ∈ P;
   end if
   W_ij ← WeightConstruct;
   G_i ← ⟨V_i, E_i, W_ij⟩;
  end for
end for
Return G = ⟨V, E, W⟩;
```

B. Weight Construction Algorithm

A participator v_i enters the mix-zones at time $t_{ingress}(v_i)$ and exits the mix-zones in a time interval from t_j to t_{j+1}. Let $P(v_i, t)$ present the probability of participator v_i exits the mix-zones in above-mentioned time interval $[t_j, t_{j+1}]$. $P(v_i, t)$ numerically equals to the probability that participator v_i takes data collection time in mix-zones from $t_j - t_{ingress}(v_i)$ to $t_{j+1} - t_{ingress}(v_i)$. As our assumption above, the data collection time $\Delta'(t)$ in mix-zones follows normal distributions $\Delta'(t) \sim N(\mu, \sigma)$. Therefore, we have

$$P(v_i, t) = \int_{t_j - t_{ingress}(v_i)}^{t_{j+1} - t_{ingress}(v_i)} f(\Delta'(t))dt \tag{1}$$

where $f(\Delta'(t))$ is the probability density function (PDF) of data collection time in mix-zones.

Similarly, the other participators exit in the time interval $[t_j, t_{j+1}]$ can be computed as above. Thus, the probability of all participators exit in the interval time $[t_j, t_{j+1}]$ can be computed by (2).

$$P(v', t) = \sum_{i=1}^{k} P(v_i, t) \tag{2}$$

However, only one of them is the real participator. Therefore, the probability that participator v_i exits in $[t_j, t_{j+1}]$ is v_i, denoted as $p(v_i[t_j, t_{j+1}])$ is given by the following conditional probability (3).

$$p(v_i[t_j, t_{j+1}]) = \frac{P(v_i, t)}{P(v', t)}, i = 1, 2, \ldots, k \tag{3}$$

The computation of edge weight is described by Algorithm 2. As we mentioned above, we use adjacency matrix to represent the constructive graph model. It can be described as follows:

$$\mathbf{Graph\ G} = \begin{pmatrix} w_{11} & w_{12} & \cdots & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & \cdots & w_{2k} \\ w_{31} & w_{22} & \cdots & \cdots & w_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{k1} & w_{k2} & \cdots & \cdots & w_{kk} \end{pmatrix}$$

Graph G can be represented as $k \times k$ matrix. The elements in Graph G represent the possibilities that are computed by (3) between the ingress pseudonyms and the egress pseudonyms. It meets the conditions (4).

$$\forall i \in [1,k], s.t. \sum_{j=1}^{k} w_{ij} = 1, 0 \le w_{ij} \le 1 \tag{4}$$

Each participator enters with one of the k pseudonyms and exits the sensitive area with a different one after he/she finishes the data collection. The uncertainty and privacy level are dependent on the distribution of edge's weight. It will be discussed in Section V.

Algorithm 2 WeightConstruct

Input: $t_{ingress}$ and $\Delta t_{egress} = [t_j, t_{j+1}]$ and $\Delta'(t)$

Output: Edge Weight W

```
Procedure
for j = 1 to k do
  for i = 1 to k do
    δ_i^1 ← t_j − t_ingress(v_i), δ_i^2 ← t_{j+1} − t_ingress(v_i);
    P(v_i, t) ← integral(f(Δ'), δ_i^1, δ_i^2);
    P(v', t) += P(v_i, t);
  end for
end for
for j = 1 to k do
  for i = 1 to k do
    p(v_i[t_j, t_{j+1}]) ← (P(v_i, t))/(P(v', t)),
    w_ij ← p(v_i[t_j, t_{j+1}]);
  end for
end for
Return W = {w_ij | w_ij = e(v_i, v_j), i, j = 1, 2, ..., k};
```

V. Metric

In this section, we introduce privacy level metric and privacy loss metric to evaluate the effectiveness of our proposal, and then define the information loss to compare the performance of our proposal with previous proposals.

A. Privacy Level Metric

We evaluate the privacy level that our proposal can achieve based on Information Entropy. The concept of information entropy defined by Shannon [50] is a quantitative measure of information content and uncertainty over a probability distribution. In this paper, the probability distribution represents the chance that adversary can identify each participator. The more uniform the probability distribution is, the higher the information entropy is and the more difficult the real participator can be identified. Otherwise, if there is a significant difference in the probability distribution, it will be easy to confirm the real participator for the low information entropy. Thus, it is feasible to measure the trajectory privacy level that our proposal can achieve using information entropy.

We represent the ingress pseudonym set as $V_{ingress} = \{v_{1i} \,|\, i = 1, 2, \ldots, k\}$ and egress pseudonym set as $V_{egress} = \{v_{2i} \,|\, i = 1, 2, \ldots, k\}$. The information entropy of sensitive trajectory segment represents the degree of uncertainty in the set of possible mapping, which can be measured by (5).

$$H(TS_i) = -\sum_{j=1}^{k}\sum_{i=1}^{k} P_{v_{2j} \leftarrow v_{1i}} \log_2 P_{v_{2j} \leftarrow v_{1i}} \tag{5}$$

where TS_i represents the ith sensitive trajectory segment and $P_{v_{2j} \leftarrow v_{1i}}$ represents the mapping possibility from v_{1i} to v_{2j}. In the graph matrix model, the mapping possibilities are presented by the weights that $w_{ij} = P_{v_{2j} \leftarrow v_{1i}}, i, j = 1, 2, \ldots, k$.

To evaluate trajectory privacy-preserving level, it is also useful to find the maximum entropy of the mapping in the trajectory mix-zones graph model. If adversary has no background knowledge about participators, the possibilities of the mapping are the same with $P_{v_{2j} \leftarrow v_{1i}} = (1/k)$. In this case, the information entropy of sensitive trajectory segment reaches the maximum value depicted by (6).

$$\max\{H(TS_i)\} = \log_2(k) \tag{6}$$

The trajectory privacy-preserving level is defined as the ratio of the entropy of each sensitive trajectory segment to the maximum entropy. The ith sensitive trajectory segment privacy-preserving level, denoted as PL_i, can be computed by (7) quantitatively.

$$PL_i\% = \frac{H(TS_i)}{\max\{H(TS_i)\}}\% \tag{7}$$

We can see that a higher value of PL_i indicates a higher trajectory privacy-preserving level. Conversely, the privacy leak is lower. It gives a hint of how far the trajectory privacy is from the theoretical privacy upper bound.

B. Privacy Loss Metric

To quantify the privacy further more, we define the privacy loss based on the model proposed in [51]. In this paper, privacy loss is defined as the probability that an adversary will be able to gain sensitive trajectory segment about a participator. It could be calculated by combining the identity leakage and the pseudonym mapping index.

The number of participators generated in the ingress time interval T constitutes the anonymity set, which meets *k-anonymity*. Thus, adversary can identify the target with the same probability. Finally, for each target, the identity leakage, denoted as *IDL*, is the inverse of the size of anonymity set and can be computed as: $IDL = (1/k)$, where k represents the size of the anonymity set.

The locations of participators cannot be tracked inside the mix-zones, adversary can observe the time and locations of all these ingress and egress participators. The trajectory of target participator can be identified by linking the ingress pseudonym with egress pseudonym. Since the time intervals in the mix-zones of different participators are different, the mapping probability of the target can be computed by (3). That is, the mapping index MI_i for each ith target can be calculated as follows:

$$MI_i = p(v_{1i} \rightarrow v_{2i'}) = \frac{P(v_i, t)}{P(v', t)}, i' = 1, 2, \ldots, k \quad (8)$$

where $P(v_i, t)$ and $P(v', t)$ could be computed by (1) and (2), respectively.

Hence, the total privacy loss of the ith target participator, denoted as Lo_i, can be calculated as a product of the identity leakage (IDL_i) and the mapping index (MI_i). That is,

$$Lo_i = IDL_i \cdot MI_i \quad (9)$$

In face, the privacy level measures how much the privacy loss of our model is. Thus, the more privacy loss is, the lower the privacy level is. In the worse case, the privacy loss of target participatory is equal to 1, that's mean the privacy of the target is exposed completely, the privacy level of the target may tend to be zero.

C. Information Loss Metric

The information loss is defined as the reduction in the probability with which people can accurately determine the position of an object in [43], [45]. In this paper, similar to [38], [42], we use the sum of area size of anonymity regions to measure the information loss. It can be computed by (10).

$$IL = \sum_{i=1}^{k} \sum_{j=1}^{n} Area\{S(o_i, t_j)\} \quad (10)$$

where IL represents the information loss with different number of trajectories, $Area\{S(o_i, t_j)\}$ represents the area size of the generalized regions of o_i at time t_j, k is the number of trajectories and n is the number of timestamps in anonymity regions. Based on the definition of information loss, we theoretically analyze some previous proposals such as dummy trajectories [39], suppression technique [40], and trajectory *k-anonymity* [42] and our proposal as follows.

Suppression technique proposed in [40] is used to reduce the probability of disclosing the whole trajectories. The information loss can be considered as the ratio of suppressed trajectory segments to the whole trajectory. If a certain data collection location is suppressed, the data reports collected at this location are lost. In [39], [42], they anonymize the whole trajectory information. The information loss can be calculated by (10). In this paper, we propose to anonymize only those sensitive trajectory segments. The information loss among these proposals will be further analyzed in Section VII.

VI. Threat Models

Follow by our analysis in Section III-B, the main goal of adversary is to identify the participator's complete trajectory associated with the true identity. Adversary's knowledge is an important factor in evaluating the privacy of our model. It is a variable and has different values for different abilities and scenarios. In this section, we consider both the weak and strong adversary attack model the same as in [15]. The uncertainty of trajectory mapping index under the two types of attack models is different, which is analyzed as follows.

A. Weak Adversary Attack Model

The weak adversary has little knowledge about the participators. It is only aware of the set of participators moving in and out of the mix-zones but not of their time intervals and trajectories. In this case, the type of probability distribution function suggests the same uniform probability for all the trajectory mapping indexes. Hence, the upper bound on the achievable uncertainty of trajectory mapping index can be computed by (6). That is, the actually achievable uncertainty of our proposal is less than the upper bound. We can see that the maximum uncertainty of trajectory mapping index depends exclusively on the number k of pseudonyms that used in the mix-zones. Thus, since the first exit participator does not corresponding to the first enter one for its random time interval, the weak adversary with little knowledge about the participators has no ability in distinguishing related participators to specific trajectories. It provides an upper bound to the achievable uncertainty of the trajectory mapping index.

B. Strong Adversary Attack Model

The strong adversary can launch the time attack such as *FIFO* by gathering entering time and exiting time intervals. Hence, besides the number of participators, the effectiveness of the mix-zones also relies on the time intervals. As we discussed, in a time interval T, k participators arrive at the mix-zones, where k is determined by the mean arrival rate λ. Additionally, we argue that the data collection time of each participator spends in the mix-zones follows normal distribution. In this case, the strong adversary can record the arrival time and leave time intervals. When adversary observes a participator exiting, he tries to map the exit participator to the related pseudonym identity. According to the distribution of time intervals, the mapping probabilities showed by the weight of graph matrix in Section IV are computed by (1). But only one of them is the real one with the probability that can be computed through inference based on the likelihoods of the others to exit at this time interval, denoted by (3). Once the mapping probabilities are computed, the uncertainty of each trajectory can be determined by (5). Thus, it presents a lower bound to the achievable uncertainty of the trajectory mapping index.

VII. Evaluation

In this section, we study the experimental evaluation of our trajectory mix-zones graph model through two components: 1) the effectiveness of our proposal against threat models in terms of privacy level and privacy loss; 2) the performance analysis in terms of information loss and costs compared with previous trajectory privacy-preserving proposals such as suppression technique [40] and trajectory *k-anonymity* [42]. Before reporting our experimental results, we first describe the experimental setup, including device configuration and parameters setting.

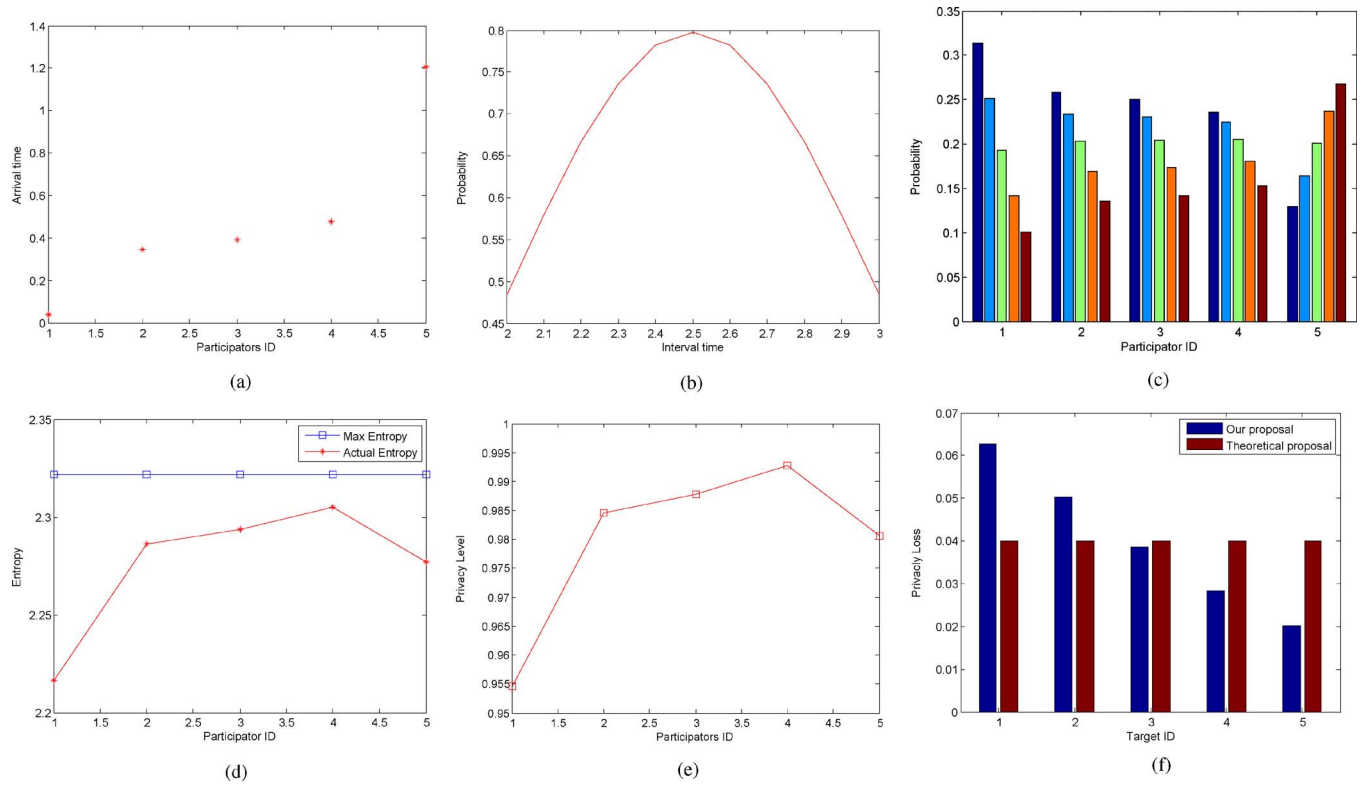

Fig. 4. Simulation results of the first group of statistical parameters. (a) Arrival time. (b) PDF. (c) Weight of edge. (d) Entropy. (e) Privacy level. (f) Privacy loss.

TABLE I
TWO GROUPS OF STATISTICAL PARAMETERS

Ingress time interval(T)	Arrival rate (λ)	Time interval parameter (μ, σ)	The number of Participators (k)
0.5	5	(2.5,0.5)	5
1	10	(3,1)	20

A. Experiment Settings

The simulation experiments are run on MATLAB platform. The machine equips with an Intel Core 4 core, Quad 2.83 GHz, Windows XP system equipped with 4 GB main memory. We model trajectory mix-zones by the following parameters: 1) Pseudonyms information; 2) Participators ingress time interval T and egress time intervals $\Delta t^i_{egress}, i = 1, 2, \ldots, k$; 3) Participator arrival rate λ and random time interval parameters (μ, σ).

The vertexes of graph model are constructed by participators' pseudonyms provided by TTPs. Participators' arrival process follows Poisson distribution with the arrival rate λ. The number of participators should satisfy the requirement of *k-anonymity* within the ingress time interval T. The time interval of participators that stays in mix-zones is modeled by normal distribution with parameters (μ, σ). Participators' egress time intervals are represented as $\Delta t^i_{egress}, i = 1, 2, \ldots, k$. In our experiments, the first exit time is set to 3 and size of participators' exit time intervals is set to 0.1.

We generate a series of random values with two groups of statistical parameters showed in Table I to analyze the effectiveness of our proposal against the different attack models in Section VI. The ingress time interval T and the arrival rate λ decide the number of participators k that enters the mix-zones. It must satisfy the requirement of *k-anonymity*.

B. Effectiveness Analysis

We run two groups of experiments with different statistical parameters in Table I. As a result of participators' different arrival rates $\lambda(\lambda = 5, 10)$ depicted by Figs. 4(a) and 5(a), the number of participators that enters the mix-zones is different. Specifically, to ensure *k-anonymity*, we consider the number of participators with $[2\lambda T]$, which is showed in Table I during ingress time interval $T(T = 0.5, 1)$. The arrival time should not differ at a large value so as to prevent from time attack in Section VI. Figs. 4(b) and 5(b) present the probability density function of time interval in mix-zones with $(\mu, \sigma) = \{(2.5, 0.5), (3, 1)\}$. The egress time intervals of k participators are $[t_j, t_{j+1}], j = 1, 2, \ldots, k$. The graph is constructed by Algorithm 1 and the weight of edge can be calculated by Algorithm 2. Figs. 4(c) and 5(c) present the weight of edge, which means each probability mapping from the ingress pseudonym to the egress pseudonym.

As we mentioned, the uncertainty of mapping among pseudonyms can be evaluated by (5). According to the discussion above, the maximum entropy achieves if and only if the mapping probabilities are equal. In this paper, we improve the theoretical mix-zones model with considering the time factor. As illustrated by Figs. 4(d) and 5(d), the maximum entropy and actual entropy can be computed according to the probability distributions depicted by Figs. 4(c) and 5(c) respectively. The probability distributions represent the probabilities of k participators that exit trajectory mix-zones at each egress time interval. As

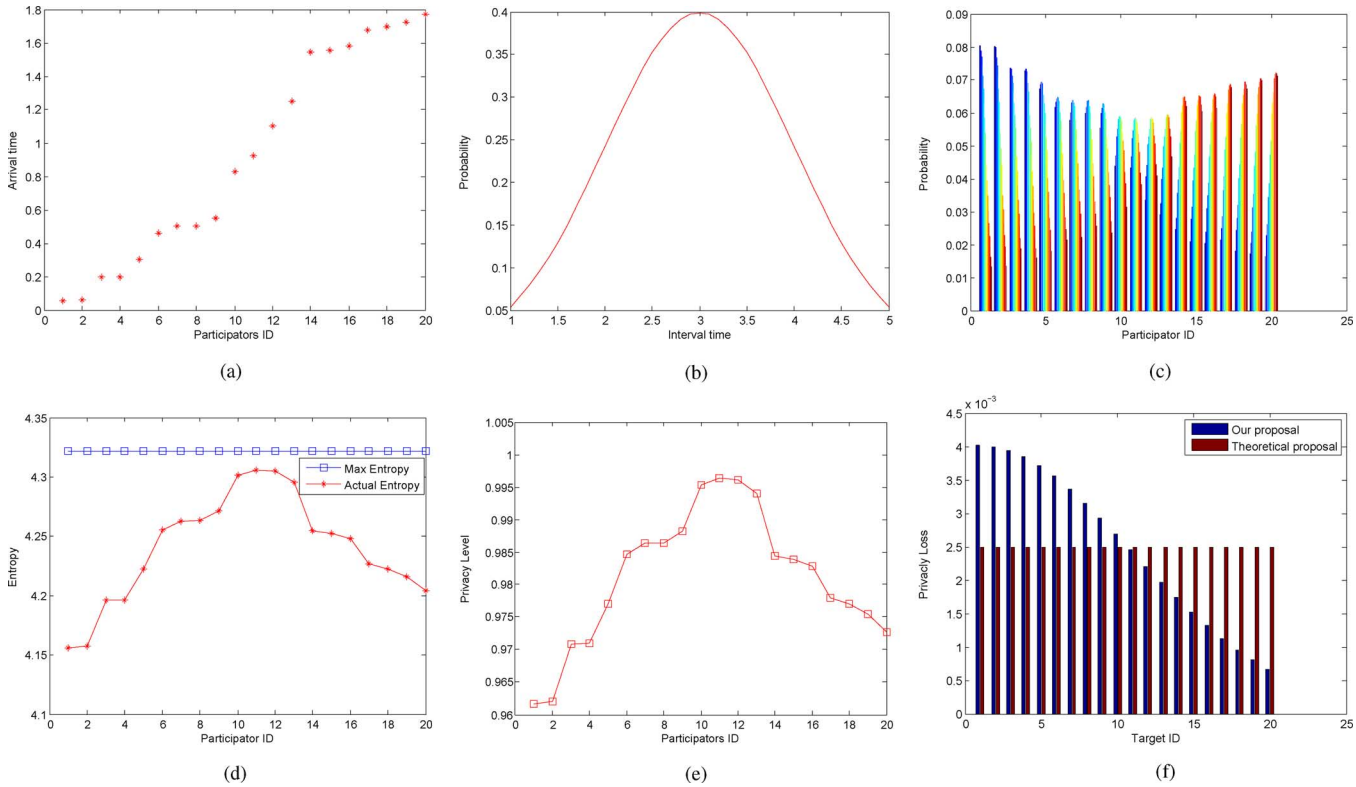

Fig. 5. Simulation results of the second group of statistical parameters. (a) Arrival time. (b) PDF. (c) Weight of edge. (d) Entropy. (e) Privacy level. (f) Privacy loss.

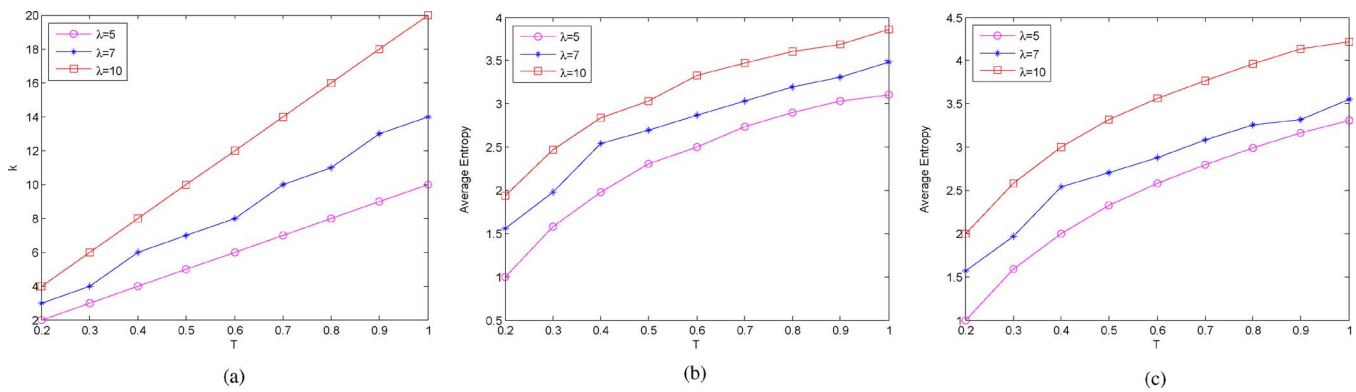

Fig. 6. Comparison of average entropy. (a) Number of participators. (b) Collection time $(\mu, \sigma) = (2.5, 0.5)$. (c) Collection time $(\mu, \sigma) = (3, 1)$.

we can see from these figures, the more uniform of the mapping probability distributions are, the higher the actual entropy is. When the mapping probabilities are equal, the maximum entropy achieves.

According to (7), privacy level can be calculated depicted by Figs. 4(e) and 5(e). It evaluates the privacy-preserving level of our proposal. The higher privacy level is, the stronger the trajectory privacy-preserving proposal is. Consequently, the privacy leak is lower. Moreover, when we define the privacy level, it is important to measure the privacy loss. Considering a given participator such as P_{11}, Figs. 4(f) and 5(f) demonstrate the privacy loss of our model compared with the theoretical mix-zones model according to the computational model proposed in Section V-B. As illustrated by the results, in the theoretical mix-zones, the mapping probabilities of P_{11} from the ingress pseudonym v_{11} to the egress pseudonym $v_{2i}, i = 1, 2, \ldots, k$ are the same. Thus, the privacy loss are the same whatever the target pseudonym the ingress pseudonym is mapping to. However, when taking the other factors such as time interval in the mix-zones into consideration, the privacy loss is different for the different probabilities of mapping index. Additionally, compared Fig. 4(f) with Fig. 5(f), the privacy loss decreases with the number of participators in the mix-zones increases.

Furthermore, based on our discussion above, the number of participators that enters the mix-zones k changes with the arrival rate λ and ingress time interval T varying, as showed in Fig. 6(a). Clearly, the number of participators increases along with the increase of arrival rate and time interval. We compare the average entropy with various values of arrival rate λ and ingress time interval T in Fig. 6(b) and (c) under the same experimental setting in Table I. The figures show that the average entropy of the mix-zones increases with the increasing of

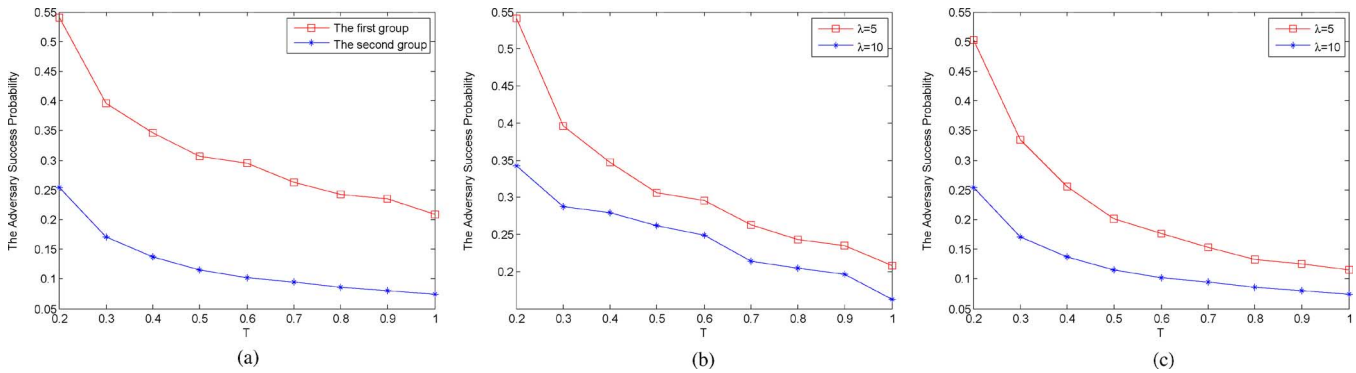

Fig. 7. Adversary success probability. (a) Comparison of the two group parameters. (b) Collection time $(\mu, \sigma) = (2.5, 0.5)$. (c) Collection time $(\mu, \sigma) = (3, 1)$.

the number of participators. That is because a large number of participators raises the uncertainty of the mix-zones. We consider the maximum mapping probability as the adversary success probability. As for a certain data collector such as P_{11}, Fig. 7(a) presents the success probability of an adversary in guessing and tracking P_{11} under the two groups of parameters. Obviously, the first group makes easier for the adversary to guess and track than the second one based on their time intervals. That is because there are less data collectors for a lower entropy of the first group than that of the second one in the same time interval. Additionally, we analyze the effects of arrival rate λ to adversary success probability under the same collection time in Fig. 7(b) and (c). Since a larger arrival rate may increase the number of data collectors in the mix zones, the adversary success probability in guessing and tracking P_{11} decreases.

C. Comparison

The comparisons between our proposal and previous work can be evaluated from the following two aspects: efficacy of anonymous quality measured by information loss and efficiency of anonymity measured by the costs of storage.

1) Efficacy: As we mentioned in Section V-C, suppression technique [40] is used to suppress the sensitive locations of data collection. The data information may almost all lost in case the sensitive location is suppressed. The more the locations are suppressed, the higher information loss is generated. Trajectory *k-anonymity* such as (k, δ)-*anonymity* [42] uses space trajectory clustering to transform those trajectories in each cluster into a (k, δ)-*anonymity* set, where δ represents the radius of anonymity set. Dummies protect trajectory privacy by adding dummy trajectories to confuse adversary [39]. They all deal with the whole trajectory and the information loss can be computed by (10). In this paper, we only anonymize those sensitive locations on or nearby participator's trajectory by trajectory mix-zones graph model. According to the computational method of information loss, we compare our proposal with trajectory *k-anonymity* under the two groups of statistical parameters in Table I.

Suppose the trajectories are divided into N segments, and there are L sensitive segments. Obviously, $L \leq N$. For simplicity, we set $N = 5$ and $L = 2$ respectively. The basic size of anonymized area of trajectory *k-anonymity* and our proposal is the same, such as (k, δ)-*anonymity*. Given the value of δ, the anonymized area is fixed to $\pi(\delta/2)^2$, where δ is related to the number of trajectories. In [42], the value of δ ranges from 1000 to 4000, step by 1000. In this paper, we consider $\delta = 1500$ as the maximum distance between any two trajectories. Because our proposal only deals with the sensitive trajectory segments instead of the whole trajectory, it can reduce the information loss which is illustrated by Fig. 8. In the worse case, the whole trajectory is sensitive, hence, the information loss of our proposal tends to be that of trajectory *k-anonymity*.

Fig. 8 describes the information loss with different numbers of trajectories. The *x-axis* represents the number of trajectories and the *y-axis* shows the information loss. As we expected, small values of k yield a low information loss, that increases monotonically as k grows. That is because a large k increases the area size of anonymity regions. Compare the results of different proposals in Fig. 8, we can see that our proposal is superior to trajectory *k-anonymity* for only anonymizing the sensitive trajectory segments. The information loss of our proposal is mainly caused by trajectory mix-zones. Moreover, since suppression technique deletes the sensitive trajectory segments from the whole trajectory, the data information on or nearby the sensitive trajectory segments may almost all be lost.

2) Efficiency: The efficiency of different trajectory-preserving proposals can be measured by qualitative analysis. Dummy trajectories [39] and trajectory *k-anonymity* [42] privacy protections must store all the trajectories. The efficiency of our proposal mainly depends on the number of pseudonyms. Given k trajectories and each trajectory contains N segments. The proposals of dummy trajectories [39] and trajectory *k-anonymity* [42] need storage memory with $O(N * k)$ to store the total k trajectories. Nevertheless, not all of the N trajectory segments are sensitive, our proposal only anonymizes the sensitive segments with mix-zones. The use of different pseudonyms can help to prevent the adversary from identifying the participators' actual trajectories. Since these pseudonyms can be used in different trajectory segments, the storage memory of pseudonyms is about $O(k)$. With the rapid increase in the number of trajectories, the storage memory of dummy trajectories and trajectory *k-anonymity* will also increase fast for storing more trajectories. In our proposal, we only need to store the pseudonyms mapping by TTPs. The increase of

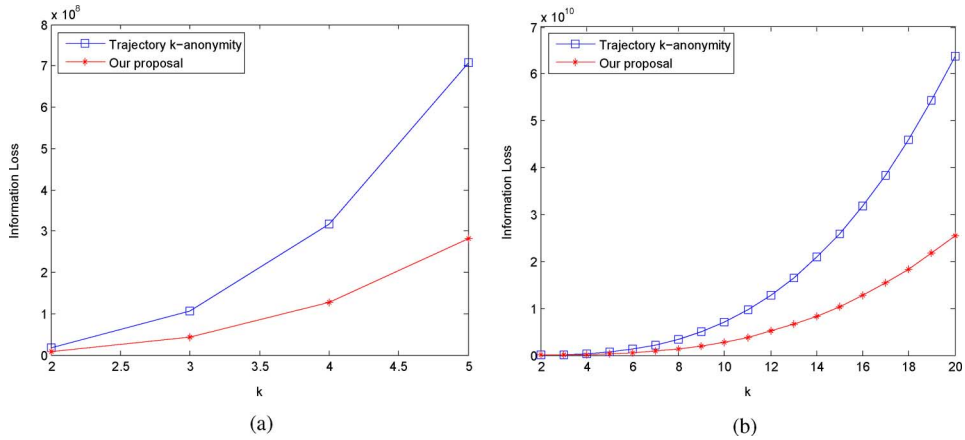

Fig. 8. Information loss. (a) First group. (b) Second group.

trajectories may not affect the number of pseudonyms too much. By comparison, our proposal has lesser storage memory than that of the other proposals.

VIII. Conclusions and Future Work

The disclosure of data collectors' trajectories poses serious threats to participators' personal privacy. It may prevent participators from data sharing. In this paper, we first propose a trajectory privacy-preserving framework TrPF for participatory sensing. Then, we propose a trajectory mix-zones graph model to protect participators' trajectories from the perspective of graph theory. We take the time factor into consideration to improve the mix-zones model. It may be more realistic in practice. Thirdly, we define the privacy metric in terms of the privacy level and privacy loss and information loss metric, and then analyze the threat models with different background knowledge. Finally, we evaluate the effectiveness and performance of our trajectory mix-zones graph model using the metric above with different parameter sets. The simulation results prove that the trajectory mix-zones graph model can protect participators' trajectories privacy effectively and reduce the information loss and costs in contrast to the other proposals. In the future, we will work on the semantic trajectory privacy problems of multiple mix-zones in detail.

Acknowledgment

The authors appreciate the helpful comments and suggestions from the anonymous reviews.

References

[1] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Commun. Mag.*, vol. 48, no. 9, pp. 140–150, Sep. 2010.

[2] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *Proc. Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications*, 2006, pp. 117–134, ACM.

[3] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson, "People-centric urban sensing," in *Proc. 2nd Ann. Int. Workshop on Wireless Internet*, 2006, p. 18, ACM.

[4] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden, "Cartel: A distributed mobile sensor computing system," in *Proc. 4th Int. Conf. Embedded Networked Sensor Systems*, 2006, pp. 125–138, ACM.

[5] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G. S. Ahn, and A. T. Campbell, "Bikenet: A mobile sensing system for cyclist experience mapping," *ACM Trans. Sensor Netw. (TOSN)*, vol. 6, no. 1, p. 6, 2009.

[6] S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin, and M. Hansen, "Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype," in *Proc. 4th Workshop on Embedded Networked Sensors*, 2007, pp. 13–17, ACM.

[7] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda, "Peir, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proc. 7th Int. Conf. Mobile Systems, Applications, and Services*, 2009, pp. 55–68, ACM.

[8] D. Christin, A. Reinhardt, S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *J. Syst. Softw.*, vol. 84, no. 11, pp. 1928–1946, 2011.

[9] K. L. Huang, S. S. Kanhere, and W. Hu, "Preserving privacy in participatory sensing systems," *Comput. Commun.*, vol. 33, no. 11, pp. 1266–1280, 2010.

[10] L. Hu and C. Shahabi, "Privacy assurance in mobile sensing networks: Go beyond trusted servers," in *Proc. IEEE 8th Int. Conf. Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2010, pp. 613–619.

[11] L. Kazemi and C. Shahabi, "Towards preserving privacy in participatory sensing," in *Proc. 9th Int. Conf. Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2011, pp. 328–331.

[12] R. K. Ganti, N. Pham, Y. E. Tsai, and T. F. Abdelzaher, "Poolview: Stream privacy for grassroots participatory sensing," in *Proc. 6th ACM Conf. Embedded Network Sensor Systems*, 2008, pp. 281–294, ACM.

[13] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Comput.*, vol. 2, no. 1, pp. 46–55, 2003.

[14] A. R. Beresford and F. Stajano, "Mix zones: User privacy in location-aware services," in *Proc. 2nd IEEE Ann. Conf. Pervasive Computing and Communications Workshops*, 2004, pp. 127–131, IEEE.

[15] J. Freudiger, M. Raya, M. Flegyhzi, P. Papadimitratos, and J. P. Hubaux, "Mix-zones for location privacy in vehicular networks," in *Proc. 1st Int. Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS 07)*, Vancouver, BC, Canada, 2007.

[16] B. Palanisamy and L. Liu, "Mobimix: Protecting location privacy with mix-zones over road networks," in *Proc. IEEE 27th Int. Conf. Data Engineering (ICDE)*, 2011, pp. 494–505.

[17] A. Kapadia, N. Triandopoulos, C. Cornelius, D. Peebles, and D. Kotz, "Anonysense: Opportunistic and privacy-preserving context collection," *Pervasive Comput.*, vol. 5013, pp. 280–297, 2008.

[18] E. De Cristofaro and C. Soriente, "Pepsi: Privacy-enhanced participatory sensing infrastructure," in *Proc. ACM 4th Conf. Wireless Network Security (WiSec'11)*, 2011, pp. 23–28.

[19] D. Christin, M. Hollick, and M. Manulis, "Security and privacy objectives for sensing applications in wireless community networks," in *Proc. IEEE 19th Int. Conf. Computer Communications and Networks (ICCCN)*, 2010, pp. 1–6.

[20] A. Kapadia, D. Kotz, and N. Triandopoulos, "Opportunistic sensing: Security challenges for the new paradigm," in *Proc. IEEE 1st Int. Communication Systems and Networks and Workshops*, 2009, pp. 1–10.

[21] C. Y. Chow and M. F. Mokbel, "Trajectory privacy in location-based services and data publication," *ACM SIGKDD Explorations Newsletter*, vol. 13, no. 1, pp. 19–29, 2011.
[22] L. Liu, "From data privacy to location privacy: Models and algorithms," in *Proc. 33rd Int. Conf. Very Large Data Bases (VLDB2007)*, 2007, pp. 1429–1430, VLDB Endowment.
[23] J. Krumm, "A survey of computational location privacy," *Personal and Ubiquitous Comput.*, vol. 13, no. 6, pp. 391–399, 2009.
[24] M. Decker, "Location privacy-an overview," in *Proc. IEEE 7th Int. Conf. Mobile Business (ICMB'08)*, 2008, pp. 221–230.
[25] R. Shokri, J. Freudiger, and J. P. Hubaux, "A unified framework for location privacy," in *Proc. 9th Int. Symp. Privacy Enhancing Technologies (PETS'10)*, 2010, pp. 203–214.
[26] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proc. Int. Conf. Pervasive Services*, 2005, pp. 88–97.
[27] H. Lu, C. S. Jensen, and M. L. Yiu, "Pad: Privacy-area aware, dummy-based location privacy in mobile services," in *Proc. 7th ACM Int. Workshop on Data Engineering for Wireless and Mobile Access*, 2008, pp. 16–23, ACM.
[28] S. Gao, J. Ma, W. Shi, and G. Zhan, "Towards location and trajectory privacy protection in participatory sensing," in *Proc. Mobile Computing, Applications and Services*, Los Angles, CA, USA, 2011, pp. 381–386.
[29] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty Fuzziness and Knowl. Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.
[30] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. ACM 1st Int. Conf. Mobile Systems, Applications and Services*, 2003, pp. 31–42.
[31] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 1–18, Jan. 2008.
[32] C. Bettini, S. Mascetti, X. S. Wang, and S. Jajodia, "Anonymity in location-based services: Towards a general framework," in *Proc. IEEE Int. Conf. Mobile Data Management*, 2007, pp. 69–76.
[33] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," in *Proc. 3rd Int. Conf. Pervasive Computing (PERVASIVE'05)*, 2005, pp. 152–170.
[34] C. Ardagna, M. Cremonini, E. Damiani, S. De Capitani Di Vimercati, and P. Samarati, "Location privacy protection through obfuscation-based techniques," *Data and Applications Security XXI*, pp. 47–60, 2007.
[35] J. Freudiger, M. H. Manshaei, J. Y. Le Boudec, and J. P. Hubaux, "On the age of pseudonyms in mobile ad hoc networks," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.
[36] J. Freudiger, R. Shokri, and J. P. Hubaux, "On the optimal placement of mix zones," in *Privacy Enhancing Technologies*. New York, NY, USA: Springer, 2009, pp. 216–234.
[37] X. Liu, H. Zhao, M. Pan, H. Yue, X. Li, and Y. Fang, "Traffic-aware multiple mix zone placement for protecting location privacy," in *Proc. IEEE INFOCOM*, 2012, pp. 972–980.
[38] Z. Huo, Y. Huang, and X. Meng, "History trajectory privacy-preserving through graph partition," in *Proc. ACM 1st Int. Workshop on Mobile Location-Based Service*, 2011, pp. 71–78.
[39] T. H. You, W. C. Peng, and W. C. Lee, "Protecting moving trajectories with dummies," in *Proc. IEEE Int. Conf. Mobile Data Management*, 2007, pp. 278–282.
[40] M. Terrovitis and N. Mamoulis, "Privacy preservation in the publication of trajectories," in *Proc. IEEE 9th Int. Conf. Mobile Data Management (MDM'08)*, 2008, pp. 65–72.
[41] M. E. Nergiz, M. Atzori, and Y. Saygin, "Towards trajectory anonymization: A generalization-based approach," in *Proc. ACM SIGSPATIAL ACM GIS 2008 Int. Workshop on Security and Privacy in GIS and LBS*, 2008, pp. 52–61.
[42] O. Abul, F. Bonchi, and M. Nanni, "Never walk alone: Uncertainty for anonymity in moving objects databases," in *Proc. IEEE 24th Int. Conf. Data Engineering (ICDE2008)*, 2008, pp. 376–385.
[43] R. Yarovoy, F. Bonchi, L. V. S. Lakshmanan, and W. H. Wang, "Anonymizing moving objects: How to hide a mob in a crowd?," in *Proc. ACM 12th Int. Conf. Extending Database Technology: Advances in Database Technology*, 2009, pp. 72–83.
[44] T. Xu and Y. Cai, "Exploring historical location data for anonymity preservation in location-based services," in *Proc. IEEE 27th Conf. Computer Communications (INFOCOM2008)*, 2008, pp. 547–555.
[45] Z. Huo, X. Meng, H. Hu, and Y. Huang, "You can walk alone: Trajectory privacy-preserving through significant stays protection," in *Proc. 17th Int. Conf. Database Systems for Advanced Applications (DASFAA2012)*, 2012, pp. 351–366.
[46] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, "A clustering-based approach for discovering interesting places in trajectories," in *Proc. 2008 ACM Symp. Applied Computing*, 2008, pp. 863–868, ACM.
[47] A. Monreale, R. Trasarti, D. Pedreschi, C. Renso, and V. Bogorny, "C-safety: A framework for the anonymization of semantic trajectories," *Trans. Data Privacy*, vol. 4, no. 2, pp. 73–101, 2011.
[48] Y. Zheng, L. Zhang, X. Xie, and W. Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proc. ACM 18th Int. Conf. World Wide Web*, 2009, pp. 791–800.
[49] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, "Anonysense: Privacy-aware people-centric sensing," in *Proc. ACM 6th Int. Conf. Mobile Systems, Applications, and Services*, 2008, pp. 211–224.
[50] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
[51] M. Saini, P. K. Atrey, S. Mehrotra, and M. Kankanhalli, "w^3-privacy: Understanding what, when, and where inference channels in multi-camera surveillance video," *Multimedia Tools and Applicat.*, pp. 1–24, 2012.

Sheng Gao received the B.Sc. degree in information and computation science from Xi'an University of Posts and Telecommunications, Xi'an, China, in 2009. He is currently working toward the Ph.D. degree at the School of Computer Science and Technology, Xidian University, Xi'an.

His current research interests include mobile computing, participatory sensing, and mobile data management with focus on security and privacy issues.

Jianfeng Ma received the B.Sc. degree in mathematics from Shaanxi Normal University, Xi'an, China, in 1985, and the M.Sc. and Ph.D. degrees in computer software and communications engineering from Xidian University, Xi'an, China, in 1988 and 1995, respectively.

From 1999 to 2001, he was a Research Fellow with Nanyang Technological University of Singapore. He is a senior member of the Chinese Institute of Electronics (CIE). Currently, he is a Professor and Ph.D. supervisor in the School of Computer Science and Technology at Xidian University, Xi'an, China. He is also Director of Shaanxi Key Laboratory of Network and System Security. His current research interests include information and network security, wireless and mobile computing systems, and computer networks. He has published over 200 refereed articles in these areas and coauthored over ten books.

Weisong Shi (S'99–A'00–M'01–SM'09) is an Associate Professor of Computer Science at Wayne State University. He received the B.Sc. degree from Xidian University in 1995, and the Ph.D. degree from the Chinese Academy of Sciences in 2000, both in computer engineering.

His current research focuses on computer systems, mobile computing, and cloud computing. He has published 120 peer-reviewed journal and conference papers in these areas, with an H-index of 23. He has served the program chairs and technical program committee members of numerous international conferences, including WWW, ICDCS, and so on.

Dr. Shi is a recipient of the NSF CAREER award, one of 100 outstanding Ph.D. dissertations (China) in 2002, Career Development Chair Award of Wayne State University in 2009, and the "Best Paper Award" of ICWE'04, IPDPS'05, HPCChina'12, and IISWC'12.

Guoxing Zhan is currently a software engineer at Cisco Systems Inc. He received the Ph.D. degree in computer science from Wayne State University, in 2012, and the M.S. degree in mathematics from Chinese Academy of Sciences, in 2007.

He has a broad interest in research on data center networking, participatory sensing, wireless sensor network, mobile computing, networking and systems security, trust management, and information processing.

Cong Sun received the B.S. degree in computer science from Zhejiang University, in 2005, and the Ph.D. degree in computer science from Peking University, in 2011.

He is currently an assistant professor in the School of Computer Science at Xidian University. His research interests include system software and information security.