Joint Location-Value Privacy Protection for Spatiotemporal Data Collection via Mobile Crowdsensing

Tong Liu[1,2](✉), Dan Li[1], Chenhong Cao[1], Honghao Gao[1], Chengfan Li[1,2], and Zhenni Feng[3]

[1] School of Computer Engineering and Science, Shanghai University, Shanghai, China
{tong_liu,ld19721539,caoch,gaohonghao}@shu.edu.cn

[2] Shanghai Engineering Research Center of Intelligent Computing System, Shanghai, China

[3] School of Computer Science and Technology, Donghua University, Shanghai, China
fzn@dhu.edu.cn

Abstract. Due to the development of the Internet of Things, mobile crowdsensing has emerged as a promising pervasive sensing paradigm for online spatiotemporal data collection, by leveraging ubiquitous mobile devices. However, privacy leakage of device users is a crucial problem, especially when an untrusted central platform in mobile crowdsensing is considered. Moreover, private information of users like trajectories contained in both location tags and sensed values of their sensing data may be unexpectedly revealed to the platform. In order to solve this problem, we proposed a joint location-value privacy protection approach, which consists of two privacy preserving mechanisms to perturb the locations and sensed values of users, respectively. The approach can be performed by each user locally and independently. The privacy of users can be well preserved, as we theoretically prove that the two mechanisms satisfy local differential privacy. In addition, extensive simulations are conducted, and the results show that accurate estimated values can be derived based on perturbed locations and sanitized sensed values, by adopting the truth discovery method.

Keywords: Mobile crowdsensing · Privacy protection · Local differential privacy · Truth discovery

1 Introduction

With the rapid development of the Internet of Things, mobile devices equipped with diverse embedded sensors (e.g., camera, accelerometer, compass) are pervasive. Mobile crowdsensing (MCS) [4,13] has emerged recently as a promising pervasive sensing paradigm to enable the Internet of Things, which facilitates spatiotemporal data collection in large urban areas like transportation monitoring [28], air

H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 407, pp. 87–103, 2021.
https://doi.org/10.1007/978-3-030-92638-0_6

monitoring [12] and noise mapping [14]. A typical MCS system consists of a central platform resided in cloud and plenty of distributed mobile device users. According to the sensing requests of the platform, users collect location-based sensing data continuously and submit them to the platform for extracting useful information.

A major concern in spatiotemporal data collection via MCS is privacy leakage [7,16], as spatiotemporal sensing data collected by users contain their private information, such as their trajectories and preferences. Moreover, an untrusted platform in the MCS system should be considered, and hence privacy protection should be conducted by each user independently. Note that both location tags and sensed values contained in the sensing data of users should be perturbed, before they are submitted to the platform. On one hand, the trajectory privacy of a user will be exposed to the untrusted platform, if sensing data with unperturbed location tags of the user are continuously submitted. On the other hand, sensed values also leak location privacy of users unexpectedly, since the location of a user may be inferred according to the values of collected data by adopting truth discovery methods. The intuition of inferring locations of users is that sensing data collected in the same location always have close values, while the values of sensing data collected in different locations may be discrepant. Thus, the sensed values of sensing data collected by users should be sanitized before they are submitted to the platform.

To solve the privacy concern in MCS, some privacy preserving approaches [5,11,18,21,22,27] have been proposed based on differential privacy (DP)[2,15] which is an effective tool to provide valid privacy protection and ensure the usability of aggregated sensing data at the same time. These DP-based approaches always assume that the platform is a trusted third party for users, which is responsible to sanitize collected sensing data and limit the disclosure of private information of users. However, the assumption is not true in reality, as the platform may leak the privacy of users for commercial benefits or be attacked by adversaries. Some other approaches [9,10,17,19,20,23,25] are proposed based on local differential privacy (LDP) [3,6], in which users perturb their sensing data locally and independently before submitting sensing data to the platform. Hence, private information of users are protected. Moreover, truth discovery methods [8] can be adopted by the platform to extract true values from the perturbed data. However, there are few works considering preserving the privacy contained in both locations and sensed values of users at the same time.

In this work, we consider a MCS system with an untrusted platform, in which location-based sensing data are collected from mobile users continuously. Each datum submitted by a user consists of the identity of the user, the sensed value of a monitored object, the location tag and the time stamp. With sensed values collected by multiple users in an interested location, the platform applies a truth discovery method to obtain the estimated value of the monitored object. Obviously, the trajectory of the user can be released to the untrusted platform over the time, which is a succession of the timestamped locations. Moreover, even if the locations of users are perturbed, we consider that the platform can also infer the trajectories of users from their sensed values in the submitted data. Considering that there are few work considering the problem that sensing data of users may lead to unexpected location privacy leakage, we try to design a privacy protection approach for online spatiotemporal data collection, which protects the location privacy of participating users by perturbing both sensed values and locations in their submitted data.

However, the joint location-value privacy protection problem in MCS is particularly difficult due to the existence of the following challenges. *Firstly*, the locations of users contained in sensing data submitted to the platform should be perturbed to protect their privacy, which will lead to the platform mismatches the collected sensing data to a wrong location. Furthermore, the accuracy of values estimated based on sanitized sensing data with perturbed locations is impacted. *Secondly*, considering there is no trusted third party, the joint location-value privacy protection approach should be performed by each user locally and independently. It makes the truth discovery conducted by the platform becomes particularly difficult. *Finally*, there exists a natural intrinsic tradeoff between the level of privacy protection and the utility of perturbed data. In other words, a high-level privacy protection approach inevitably decreases the utility of sensing data, i.e., the accuracy of estimated values.

In response to these challenges, we propose a privacy protection approach for online spatiotemporal data collection via MCS, in which a location privacy preserving mechanism and a value privacy preserving mechanism are provided respectively. Specially, the location privacy preserving mechanism is designed based on random response that each user can perturb their locations locally. In the Gaussian-mechanism-based value privacy preserving mechanism, each user sanitizes the collected sensed values by adding random Gaussian noise independently. Spatiotemporal sensing data with perturbed locations and sanitized sensed values are submitted to the platform continuously.

The main contributions of this work can be summarized as follows:

- We consider the privacy preserving problem in a MCS system to collect location-based sensing data over time, in which we observe that not only location tags but also sensed values submitted to an untrusted platform will expose the private information of users.
- We propose a LDP-based privacy protection approach, which includes two privacy preserving mechanisms to perturb location tags and sensed values respectively. The approach can be performed by each user locally and independently. We theoretically prove that the two mechanisms achieve certain local differential privacy.
- Extensive simulations are conducted to validate the performance of our proposed privacy protection approach. The simulation results show that the privacy of users is well preserved and the estimated values obtained by the truth discovery method is relatively accurate.

This paper is organized as follows. We first discuss related works in Sect. 2, and present the motivation of joint location-value privacy protection in Sect. 3. Then, we present our system model and some preliminaries in Sect. 4. Section 5 and Sect. 6 elaborate our proposed privacy protection approach and the theoretical analysis, respectively. Finally, simulation results are presented in Sect. 7, and the paper is concluded in Sect. 8.

2 Related Work

Privacy protection has received a lot of attention in MCS, while differential privacy is seen as a promising technology in recent studies [5,9–11,17–23,25,27]. These privacy protection approaches in MCS can be classified into two categories, i.e., DP-based approaches and LDP-based approaches.

2.1 DP-Based Approaches in MCS

DP-based privacy preserving approaches assume there exists a trusted third party(e.g., a platform or a central server) has been widely adopted and used in many areas [1,26]. In the MCS system, there are some DP-based approaches [5,11,18,21,22,27] which sanitize the sensing data collected from mobile users for privacy protection.

To *et al.* [18] introduce a framework for protecting location privacy of works participating spatial crowdsourcing tasks, which needs users' cellular service providers to take coordination role between users and MCS platforms. Wang *et al.* [22] study the privacy protection problem in a crowd-sourced system for continuous real-time spatiotemporal data publishing, and an online privacy preserving scheme is proposed to monitor population statistics over infinite streams. Then, an enhanced RescueDP framework in [21] is proposed which leverages neural networks to accurately predict the values of statistics and improve the utility of released data. In [5,11,27], privacy preserving auction-based incentive mechanisms are designed to preserve the users' bid privacy. Specifically, the mechanism designed by Jin *et al.* [5] approximately minimizes the platform's total payment with a guaranteed approximation ratio. Besides, Lin *et al.* [11] propose two score functions to realize frameworks for privacy-preserving aution-based incentive mechanisms which achieves approximate social cost minimization. Differently, the joint effect of users privacy concerns and the positive network effect are considered in [27].

However, the assumption of a trusted third party is unpractical sometimes, as the platform may leak the privacy of users for commercial interests or be attacked by adversaries.

2.2 LDP-Based Approaches in MCS

Recently, some LDP-based approaches toward data statistics and analysis in MCS are widely adopted to alleviate the privacy concerns caused by untrusted third party [24]. Mobile users can sanitize their private sensing data locally and submitting the perturbed data to the platform.

There are some works [17,23] focus on studying the privacy preserving data distribution estimation with LDP in MCS. Wang *et al.* [23] provide an optimal LDP-based privacy preserving mechanism for distribution estimation over user-contributed data, in which the private information of users contained in both qualitative data and discrete quantitative data can be protected. Ren *et al.* [17] develop LDP-based privacy-preserving algorithms for multi-dimensional data

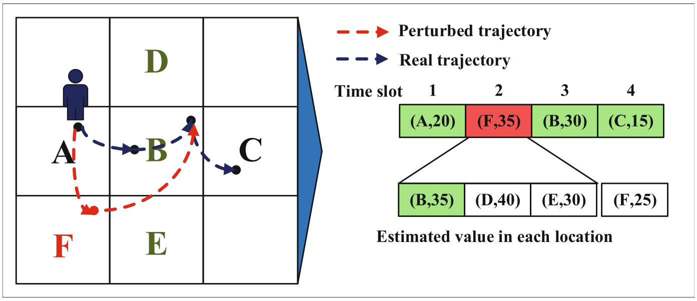

Fig. 1. An example of inferring the real location of a user based on sensed values, even though the location is perturbed.

distribution estimation and data publication, which achieve high computation efficiency and data utility.

[19,20,25] design privacy-preserving frameworks to satisfy the privacy demands of users. In order to protect the location privacy of users, [20] design a LDP-based privacy-preserving framework which consists of a data adjustment function and an optimal location obfuscation, and they propose an inference algorithm to improve the inference accuracy of obfuscated data. While [19] leverage distortion privacy with differential privacy together to provide more comprehensive protection for users' location privacy. Differently, a privacy-preserving task allocation framework in MCS is proposed in [25], in which provides personalized location privacy protection to meet different demands of users. Moreover, Lin *et al.* [10] propose a randomized response-based privacy-preserving crowdsensing data collection and analysis method to ensure users' privacy, and Li *et al.* [9] provide a privacy preserving truth discovery mechanism with theoretical guarantees of both utility and privacy.

Unfortunately, there are few works considering both location tags and sensed values contained in sensing data may unexpectedly disclose the private information of users and further proposing a joint location-value privacy protection approach accordingly.

3 Motivation

In this section, we aim to emphasize that joint location-value privacy protection is necessary for spatiotemporal data collection via MCS. Only perturbing the locations contained in sensing data collected by a user is not enough, as the platform can infer the real locations of users according to the sensed values. Here, we give a simple example to illustrate how the platform infers the real location of a user based on his/her sensed values, even though the location in submitted sensing data is perturbed.

Example: Suppose there is a platform requiring users to collect ambient noise from various interested locations. A location privacy preserving mechanism is provided to perturb their original locations to other possible locations with a certain probability. Assume there is a mobile user who collects 20 dB, 35 dB, 30 dB, and 15 dB of ambient noise in location A, B, and C over four time slots, where ambient noise is collected twice at location B. The location of the user at the second time slot (i.e., location B) is perturbed to location F. The real trajectory and perturbed trajectory of the user are $A \rightarrow B \rightarrow B \rightarrow C$ and $A \rightarrow F \rightarrow B \rightarrow C$, respectively. The sensing data of users submitted to the platform are shown in Fig. 1. In addition, we assume the platform can obtain relatively accurate estimations of the ambient noise in each location over time.

According to the estimated values in the second time slot, the platform can easily find F is not the real location of the user. In addition, according to the locations of the user in the first and third time slot, the platform can infer that the possible location of the user in the second time slot can be D, B, or E. Then, by comparing the estimated values of these three locations with the sensed values collected by the user in the second time slot, the platform can successfully infer that the real location of the user in the second time slot is B.

4 System Model and Preliminaries

In this section, we present the model of online spatiotemporal data collection in MCS, and introduce some preliminaries including truth discovery and local differential privacy.

4.1 System Model

In this work, we consider a typical crowdsensing system consists of a central platform located in cloud and a set of registered users equipped with smart devices. We denote the set of users by $\mathcal{U} = \{u_1, u_2, \cdots, u_n\}$. Users are mobile and distributed in an urban area. The platform requires users to collect location-based and time-sensitive sensing data around several interested locations continuously. The locations of interested points in the urban area are represented as $\mathcal{L} = \{L_1, L_2, \cdots, L_m\}$, where m is the number of interested locations. For convenience, we divide time into equal-interval time slots, i.e., $\mathcal{T} = \{t_1, t_2, \cdots, t_\tau, \cdots\}$. In each time slot t_τ, the subset of users located around location $L_j \in \mathcal{L}$ is denoted as $\mathcal{U}_j^\tau \subseteq \mathcal{U}$.

Let u_i denote a user located around interested point L_j in time slot t_τ (i.e., $u_i \in \mathcal{U}_j^\tau$). We denote the location of user u_i in t_τ as l_i^τ, and we use the location of his/her nearby interested point to replace it, i.e., $l_i^\tau = L_j$. The sensed value of sensing data collected by user u_i in time slot t_τ is represented by v_i^τ. Each user submits the identity, the sensed value, the location tag, and the time stamp to the platform in real time.

Submitting original sensed values and locations of users will expose their private information (e.g., trajectories) to the platform and adversaries, since an untrusted platform may leak privacy of users for commercial interests and financial benefits or be attacked by adversaries. In this work, we consider users preserve their private information by submitting sanitized values of sensing data and perturbed locations to the platform. Specially, the perturbed location and the sanitized value of user u_i in time slot t_τ is denoted by $\tilde{l}_i^\tau$ and $\tilde{v}_i^\tau$, respectively. Note that we assume $\tilde{l}_i^\tau \in \mathcal{L}$.

With receiving all sensing data collected in t_τ, the platform aggregates the sanitized value $\tilde{v}_i^\tau$ of user u_i according to perturbed location $\tilde{l}_i^\tau$. Specially, we define the set of users whose perturbed location is L_j as $\tilde{\mathcal{U}}_j^\tau = \{u_i \in \mathcal{U} | \tilde{l}_i^\tau = L_j\}$. According to the sanitized values $\{\tilde{v}_i^\tau | u_i \in \tilde{\mathcal{U}}_j^\tau\}$ collected in L_j, the platform can obtain the estimated value $\bar{V}_j^\tau$ in location L_j by employing truth estimation as follows,

$$\bar{V}_j^\tau = \frac{\sum_{u_i \in \tilde{\mathcal{U}}_j^\tau} \tilde{w}_i^\tau \cdot \tilde{v}_i^\tau}{\sum_{u_i \in \tilde{\mathcal{U}}_j^\tau} \tilde{w}_i^\tau}, \tag{1}$$

where $\tilde{w}_i^\tau$ is the weight of user u_i, calculated based on sanitized value at time slot t_τ. Correspondingly, we denote w_i^τ as the weight of user u_i calculated based on the original sensed value at time slot t_τ.

4.2 Preliminaries

Truth Discovery [8]**:** Given an initialization of the weights of users, the truth discovery method iteratively conducts the following steps until the estimated value converges.

- **Truth estimation:** Given the weights of users and sanitized values collected in location L_j at time slot t_τ, the estimated value $\bar{V}_j^\tau$ is calculated as (1).
- **Weight update:** According to difference between sanitized values submitted to the platform and estimated value $\bar{V}^\tau$ of the monitored object, the weight of user u_i can be updated as

$$\tilde{w}_i^\tau = \log\left(\frac{\sum_{u_r \in \tilde{\mathcal{U}}_j^\tau} (\tilde{v}_r^\tau - \bar{V}_j^\tau)^2}{(\tilde{v}_i^\tau - \bar{V}_j^\tau)^2}\right). \tag{2}$$

Local Differential Privacy [3]**:** LDP is a promising technology used to provide privacy protection with a quantified guarantee, which is applied to the systems without a trusted third party.

Let $M(x)$ denote the perturbed output of a randomization mechanism M given an input x. M achieves (ϵ, δ)-LDP if it satisfies the following definition.

Definition 1 ((ϵ, δ)-LDP). *A randomization mechanism M with its output domain $range(M)$ achieves (ϵ, δ)-LDP, if for an arbitrary pair of inputs x, y and any possible subset $S \subseteq range(M)$, there exists*

$$Pr\{M(x) \in S\} \leq \exp(\epsilon) \cdot Pr\{M(y) \in S\} + \delta, \tag{3}$$

where $\epsilon > 0$ is privacy budget and $\delta \geq 0$ is relaxation variable.

Specially, randomization mechanism M is called ϵ-differential privacy when $\delta = 0$. Note that a lower value of privacy budget ϵ and δ indicates a stronger privacy protection level can be achieved, vice versa.

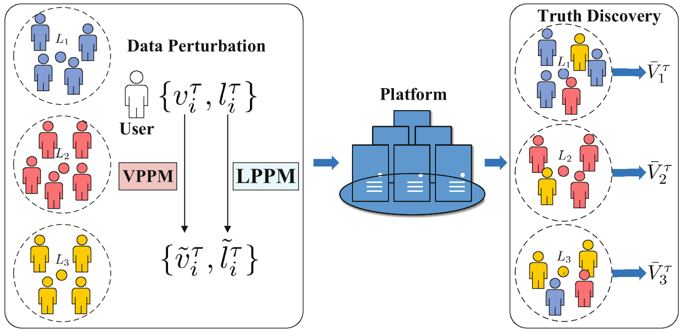

Fig. 2. An illustration of the MCS system and our proposed privacy protection approach.

5 Methodology

In this section, we first introduce the overview of our proposed privacy protection approach, which includes two mechanisms for preserving location privacy and sensed value privacy of users, respectively. Then, we describe the detailed designs of these two mechanisms in the next two subsections.

5.1 Overview

As shown in Fig. 2, our MCS system consists of a central platform resided in cloud and a set of mobile smart device users distributed in an urban area. Sensing data around interested locations are continuously collected by the users nearby and submitted to the platform. In each time slot, the operations conducted by each user and the platform are illustrated in the following.

Each user first collects the sensed value $v_i^\tau$of the monitored object in his/her current location l_i^τ. Then, each user performs the LPPM and VPPM locally and independently, to perturb the location and sanitize the sensed value as $\tilde{l}_i^\tau$ and $\tilde{v}_i^\tau$, respectively. Finally, the perturbed location and sanitized sensed value, as well as the identity of the user and the time stamp, are submitted to the platform.

The platform first aggregates sanitized values according to the perturbed locations of users after receiving all sanitized sensing data. Then, based on the sanitized values $\{\tilde{v}_i^\tau | u_i \in \tilde{\mathcal{U}}_j^\tau\}$ in each location, the platform conducts the truth discovery method to estimate the true value of the monitored object $\bar{V}_j^\tau$ in location L_j at time slot t_τ.

By conducting our privacy protection approach online in a MCS system, we can guarantee that the private information of users can be preserved and the value of the monitored object can be estimated accurately.

5.2 Location Privacy Preserving Mechanism (LPPM)

Original location tags contained in sensing data collected by users over time will disclose their trajectories to the platform or adversaries, which may pose severe threats to their real life and public security. In order to protect the location information of users, we provide a LPPM based on random response [3]. The main idea of this mechanism is that the original location of a user is perturbed to another interested location with a certain probability. The details are illustrated in the following.

We represent our LPPM by a function A, whose both input domain and output range are $\mathcal{L}$. Given a predefined probability $p \in (0,1)$, we perturb the original location $l_i^\tau = L_j \in \mathcal{L}$ of user u_i in time slot t_τ as follows,

$$\tilde{l}_i^\tau = A(l_i^\tau, p) = \begin{cases} L_j, & \text{with probability } 1-p, \\ L_r \in \mathcal{L} \setminus \{L_j\}, & \text{with probability } \frac{p}{m-1}. \end{cases}$$

In Sect. 6, we prove that our location privacy preserving mechanism satisfies LDP. Note that although the location tags of users are perturbed, the platform can still extract accurate estimated values in different locations by applying the truth discovery method. Because the sensed value with a mismatched location will be assigned a low weight in truth discovery, and there could be less impact on the accuracy of the estimated result.

5.3 Value Privacy Preserving Mechanism (VPPM)

The trajectory of a user can be still inferred by the platform or adversaries, through comparing the sensed values collected by the user over time and the estimated true values obtained by truth discovery. Thus, besides perturbing locations of users, their sensed values should be sanitized as well. In this subsection, we propose a LDP-Gaussian-based VPPM to sanitize sensed values of users. The main idea of this mechanism is adding noise on sensed values to obtain a sanitized version of them, where the noise is sampled by users from their private Gaussian distributions. The details of this mechanism are illustrated as follows.

In each time slot t_τ, the platform first publishes a predefined parameter λ to all users, where λ is determined by specific privacy demands (i.e., privacy budget ϵ_2 and δ). Then, each user u_i generates a private Gaussian distribution $\mathcal{N}(0, \sigma_i^2)$ locally, where σ_i^2 is sampled from the exponential distribution $\mathcal{E}(\lambda)$, according to the parameter published by the platform. Finally, user u_i independently samples noise ζ_i^τ from his/her private Gaussian distribution and adds the noise on the sensed value. Summarily, letting function B denote the VPPM, the process can be formulated as

$$\begin{aligned} \tilde{v}_i^\tau = B(v_i^\tau, \sigma_i^2) = v_i^\tau + \zeta_i^\tau, \\ \text{where } \zeta_i^\tau \sim \mathcal{N}(0, \sigma_i^2) \text{ and } \sigma_i^2 \sim \mathcal{E}(\lambda). \end{aligned} \tag{4}$$

Intuitively, a larger value of parameter λ indicates a smaller expectation of σ_i^2, which leads to a smaller expectation of noise added to sensed values and a lower privacy protection level accordingly.

Algorithm 1. Joint location-value privacy protection approach for spatiotemporal data collection

```
Input: Set of interested locations $\mathcal{L}$, the locations $\{l_i^\tau\}_{i=1}^n$ and sensed values $\{v_i^\tau\}_{i=1}^n$
    of all participating users at time slot $t_\tau$, predefined parameters $p$ and $\lambda$
Output: Estimated values $\{\bar{V}_j^\tau\}_{j=1}^m$ of all interested locations at time slot $t_\tau$
for each user $u_i$, $(i = 1, \cdots, n)$ independently do
    perturbs his/her location $l_i^\tau$ as
        $\tilde{l}_i^\tau = \begin{cases} L_j, & \text{with probability } 1-p, \\ L_r \in \mathcal{L} \setminus \{L_j\}, & \text{with probability } \frac{p}{m-1}. \end{cases}$
    generates a private Gaussian distribution $\mathcal{N}(0, \sigma_i^2)$, where $\sigma_i^2$ is sampled from
        exponential distribution $\mathcal{E}(\lambda)$.
    samples a noise $\zeta_i^\tau$ from $\mathcal{N}(0, \sigma_i^2)$, and obtains the sanitized sensed value as
        $\tilde{v}_i^\tau = v_i^\tau + \zeta_i^\tau.$
    submits the perturbed location $\tilde{l}_i^\tau$ and sanitized value $\tilde{v}_i^\tau$ to the platform.
end for
The platform aggregates sanitized values based on perturbed locations submitted
    by users.
for each interested location $L_j \in \mathcal{L}$ do
    conducts truth discovery to obtain the estimated value of location $L_j$.
end for
return Estimated values $\{\bar{V}_j^\tau\}_{j=1}^m$ of all interested locations at time slot $t_\tau$
```

So far, the locations and sensed values of users submitted to the platform are perturbed. Then, the platform can use the aforementioned truth discovery method to estimate the true values in different locations. Our privacy protection approach is summarized in Algorithm 1.

6 Theoretical Analysis

In the following, we theoretically analyze that both the location and value privacy preserving mechanisms satisfy LDP.

Theorem 1. *Given a set of locations whose size is m, the LPPM with perturbation probability p satisfies ϵ_1-local differential privacy, where $\epsilon_1 = \ln((1-p)(m-1)/p)$.*

Proof. According to Eq. (3), for any two possible locations L_j and L_r, the LPPM satisfies LDP if we could calculate the probability ratio $Pr\{A(L_j) = \tilde{l}_i^\tau\}/\,Pr\{A(L_r) = \tilde{l}_i^\tau\}$ and find its maximum. Accordingly, the ratio is maximized when function A outputs perturbed location $\tilde{l}_i^\tau$ which is identical to one of the input locations. Mathematically, when $L_j \neq L_r$ and $\tilde{l}_i^\tau = L_j$, the ratio reaches its maximum. Then we have,

$$\frac{Pr\{A(L_j) = \tilde{l}_i^\tau\}}{Pr\{A(L_r) = \tilde{l}_i^\tau\}} \leq \frac{Pr\{A(L_j) = L_j\}}{Pr\{A(L_r) = L_j\}} = \frac{1-p}{\frac{p}{m-1}}$$

Thus, the LPPM satisfies ϵ_1-LDP with $\epsilon_1 = \ln((1-p)(m-1)/p)$.

From Theorem 1, we can observe that when perturbation probability becomes larger or the size of location set becomes smaller, the value of ϵ_1 will become smaller, which indicates low level of privacy protection, vice versa.

In what follows, we present the theoretical analysis on the VPPM in each location at each time slot. We first introduce some parameters just for theoretical analysis. Generally, value v_i^τ of sensing data collected by user u_i in location L_j follows Gaussian distribution $\mathcal{N}(V_{j_{truth}}^\tau, {\rho_j^\tau}^2)$ [29], where $V_{j_{truth}}^\tau$ and ${\rho_j^\tau}^2$ represent the ground truth and the error variance at L_j, respectively. Then, we give the definition of L1-Sensitivity as follows.

Definition 2 (L1-Sensitivity). *L1-Sensitivity Δ_j^τ of a user in L_j at time slot t_τ is defined as*

$$\Delta_j^\tau = \max_{v_i^\tau, \grave{v}_i^\tau \in \mathcal{D}_j^\tau} |v_i^\tau - \grave{v}_i^\tau|,$$

where $\mathcal{D}_j^\tau$ is the range of values that may be sensed by users in L_j at t_τ, and v_i^τ and $\grave{v}_i^\tau$ are two possible values of sensing data collected by u_i.

Obviously, Δ_j^τ depends on ρ_j^τ. We present the relation between the ρ_j^τ and Δ_j^τ in the following lemma.

Lemma 1. *The value of sensitive information Δ_j^τ is smaller than $a\sqrt{2}\rho_j^\tau$ with probability at least $1 - \frac{1}{a}e^{\frac{-a^2}{2}}$, where $a \geq 0$ and decided by the sensed values collected by the users.*

Proof. According to the description mentioned, the error between sensed value v_i^τ and $V_{j_{truth}}^\tau$ follows Gaussian distribution $\mathcal{N}(0, {\rho_j^\tau}^2)$, and $v_i^\tau \sim \mathcal{N}(V_{j_{truth}}^\tau, {\rho_j^\tau}^2)$. Hence, for any two possible values v_i^τ and $\grave{v}_i^\tau$ may sensed by u_i, the difference between v_i^τ and $\grave{v}_i^\tau$ follows Gaussian distribution $\mathcal{N}(0, 2{\rho_j^\tau}^2)$. Based on the Gaussian tail bounds, we have,

$$Pr\{|v_i^\tau - \grave{v}_i^\tau| > a\sqrt{2}\rho_j^\tau\} \leq \frac{1}{a}e^{\frac{-a^2}{2}}, \tag{5}$$

where $a \geq 0$ and a is decided by values of sensing data collected by users. Thus, the lemma is proved.

Next, we take $\Delta_j^\tau = a\sqrt{2}\rho_j^\tau$ to analyze the LDP property achieved by the VPPM.

Theorem 2. *Given L1-Sensitivity Δ_j^τ and an exponential distribution with parameter with λ, the VPPM is (ϵ_2, δ)-local differential private, where $\epsilon_2 \geq \frac{{\Delta_j^\tau}^2}{2\sigma_i^2}$ and $\delta > 1 - e^{\frac{-\lambda {\Delta_j^\tau}^2}{2\epsilon_2}}$.*

Proof. According to Eq. (4), user u_i adopts VPPM to add noise sampled from Gaussian distribution $\mathcal{N}(0, \sigma_i^2)$ on v_i^τ to obtain sanitized value $\tilde{v}_i^\tau$. Besides, noise variance σ_i^2 is sampled from an exponential distribution with parameter λ. For any two possible sensed values v_i^τ and $\grave{v}_i^\tau$, we have,

$$\frac{Pr\{B(v_i^\tau, \sigma_i^2) = \tilde{v}_i^\tau\}}{Pr\{B(\grave{v}_i^\tau, \sigma_i^2) = \tilde{v}_i^\tau\}} = \frac{\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(\tilde{v}_i^\tau - v_i^\tau)^2}{2\sigma_i^2}}}{\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(\tilde{v}_i^\tau - \grave{v}_i^\tau)^2}{2\sigma_i^2}}} \tag{6}$$
$$= e^{\frac{(\tilde{v}_i^\tau - \grave{v}_i^\tau)^2 - (\tilde{v}_i^\tau - v_i^\tau)^2}{2\sigma_i^2}} \leq e^{\frac{(v_i^\tau - \grave{v}_i^\tau)^2}{2\sigma_i^2}} \leq e^{\frac{\Delta_j^{\tau\,2}}{2\sigma_i^2}} \leq e^{\epsilon_2}$$

According to Eq. (6), when $\sigma_i^2 \geq \frac{\Delta_j^{\tau\,2}}{2\epsilon_2}$, mechanism B meets ϵ_2-local differential privacy. As σ_i^2 follows the exponential distribution with parameter λ, and we constrain the probability of event $\{\sigma_i^2 : \sigma_i^2 \geq \frac{\Delta_j^{\tau\,2}}{2\epsilon_2}\}$ happens with at least $1 - \delta$. Thus, $Pr\{\sigma_i^2 \geq \frac{\Delta_j^{\tau\,2}}{2\epsilon_2}\} = e^{-\frac{\lambda \Delta_j^{\tau\,2}}{2\epsilon_2}} \geq 1 - \delta$. Therefore, $\lambda \leq \frac{2\epsilon_2 ln(\frac{1}{1-\delta})}{\Delta_j^{\tau\,2}}$.

Next we partition $\mathbb{R}^+$, the domain of noise variance, as $\mathbb{R}^+ = R_1 \cup R_2$, where $R_1 = \left\{\sigma_i^2 \in \mathbb{R}^+ : \sigma_i^2 \geq \frac{\Delta_j^{\tau\,2}}{2\epsilon_2}\right\}$ and $R_2 = \left\{\sigma_i^2 \in \mathbb{R}^+ : \sigma_i^2 \leq \frac{\Delta_j^{\tau\,2}}{2\epsilon_2}\right\}$. For subset $S_1 \in \mathbb{S}$ and $S_2 \in \mathbb{S}$, where $\mathbb{S}$ is the range of $B(v_i^\tau, \sigma_i^2)$, we define $S_1 = \left\{B(v_i^\tau, \sigma_i^2) | \sigma_i^2 \in R_1\right\}$ and $S_2 = \left\{B(v_i^\tau, \sigma_i^2) | \sigma_i^2 \in R_2\right\}$. Then we have,

$$\begin{aligned}
&\underset{\sigma_i^2 \in \mathbb{R}^+}{Pr}\{B(v_i^\tau, \sigma_i^2) \in S\} \\
&= \underset{\sigma_i^2 \in R_1}{Pr}\{B(v_i^\tau, \sigma_i^2) \in S_1\} + \underset{\sigma_i^2 \in R_2}{Pr}\{B(v_i^\tau, \sigma_i^2) \in S_2\} \\
&\leq \underset{\sigma_i^2 \in R_1}{Pr}\{B(v_i^\tau, \sigma_i^2) \in S_1\} + \delta \\
&\leq e^{\epsilon_2}(\underset{\sigma_i^2 \in R_1}{Pr}\{B(\grave{v}_i^\tau, \sigma_i^2) \in S_1\}) + \delta \\
&\leq e^{\epsilon_2}(\underset{\sigma_i^2 \in \mathbb{R}^+}{Pr}\{B(\grave{v}_i^\tau, \sigma_i^2) \in S\}) + \delta,
\end{aligned}$$

Thus, mechanism B yields (ϵ_2, δ)-local differential privacy, where $\epsilon_2 \geq \frac{\Delta_j^{\tau\,2}}{2\sigma_i^2}$ and $\delta > 1 - e^{\frac{-\lambda \Delta_j^{\tau\,2}}{2\epsilon_2}}$.

From Theorem 2, we can find that when σ_i^2 becomes larger, the lower bound of ϵ_2 becomes smaller. In addition, the lower bound of δ will be smaller when the value of λ is smaller. When ϵ_2 and δ have smaller values, higher privacy protection can be achieved.

7 Performance Evaluation

7.1 Simulation Setup

The default settings in our simulations are set as follows. We consider there is an urban area consists of 10 interested locations that need to monitor ambient noise, and the total number of users is 400. The sensed values of users are simulated by a Gaussian distribution $\mathcal{N}(V^{\mathcal{T}}_{j_{truth}}, 3)$, where $V^{\mathcal{T}}_{j_{truth}}$ represents the ground truth and is uniformly distributed in [20, 100]dB. We set perturbation probability p as 0.3 (i.e., $\epsilon_1 = 3.004$), privacy budget ϵ_2 as 0.7, and relaxation variable $\delta = 0.3$. Besides, the weight of each user are equally initialized to 1 at the beginning of each time slot.

We compare our proposed approach with two baselines:

- *No Privacy Protection(NPP)*: Each user submits the original sensing data. Then the estimated values obtained by truth discovery.
- *Original Location with Sanitized Value (OLSV)*: Each user submits the original locations and sanitized values obtained by value privacy preserving mechanism to the platform. Then the estimated values obtained by truth discovery.
- *Perturbed Location with Original Value (PLOV)*: Each user submits the perturbed locations obtained by location privacy preserving mechanism and original sensed values to the platform. Then the estimated values obtained by truth discovery.
- *Privacy Protection with Mean (PPM)*: Each user submits the perturbed locations and sanitized values obtained by our privacy preserving mechanisms to the platform. Then the estimated values obtained by taking the average of the sanitized values submitted by the user in each interested location.

In order to measure the performance achieved by different approaches, we first adopt the commonly used Mean Absolute Error (MAE) as our metric, which calculates the differences between ground truth and estimated values as

$$\mathrm{MAE} = \frac{1}{m}\sum_{j=1}^{m}\left|V^{\mathcal{T}}_{j_{truth}} - \bar{V}^{\mathcal{T}}_{j}\right|.$$

The smaller values of MAE indicate that the perturbation and sanitization have little impact on the accuracy of estimated results for the monitored object in all interested locations.

Besides, we compare the average accuracy of estimated values under different settings, which is calculated as

$$Accuracy = \frac{1}{m}\sum_{j=1}^{m}\left(1 - \frac{|V^{\mathcal{T}}_{j_{truth}} - \bar{V}^{\mathcal{T}}_{j}|}{V^{\mathcal{T}}_{j_{truth}}}\right).$$

7.2 Performance Evaluation

In the following, we take different numbers of interested locations into consideration to compare the MAE achieved by our privacy preserving approach and other baselines first. We plot the MAE and the accuracy achieved by five approaches when the number of locations varies from 5 to 25 in Fig. 3 and Fig. 6. It can be found that the MAE and the accuracy of our approach remain stable, which indicates that our approach is scalable to the amount of interested points.

Specially, when our approach can achieve about 94.61% accuracy of estimated values, which is only 8.13%, 7.26% and 1.94% lower than NPP, OLSV and PLOV but 4.67% higher than PMM, respectively.

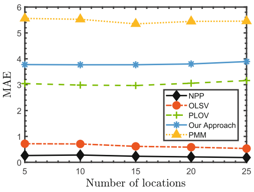

Fig. 3. MAE vs. number of locations.

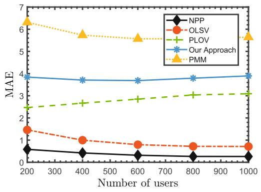

Fig. 4. MAE vs. number of users.

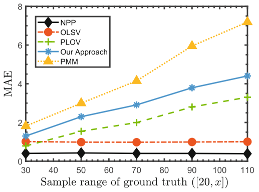

Fig. 5. MAE vs. sample range of ground truth.

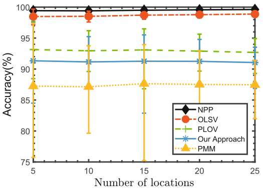

Fig. 6. MAE vs. number of locations.

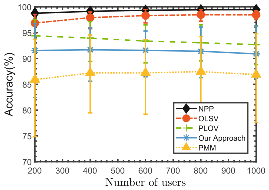

Fig. 7. MAE vs. number of users.

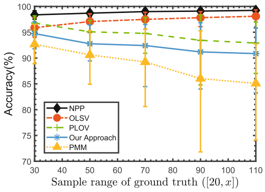

Fig. 8. MAE vs. sample range of ground truth.

As shown in Fig. 4 and Fig. 7, we evaluate the performance of five approaches, by varying the total number of users from 200 to 1000. It can be observed that the MAE and the accuracy achieved by our approach keeps stable regardless of the number of users, which indicates that our approach applies to a large-scale MCS system with plenty of users. Specially, when there are 600 users, our approach achieves 91.68% accuracy, which is only 7.51%, 6.36% and 3.41% lower than NPP, OLSV and PLOV, respectively.

For further studying the performance of our privacy preserving mechanisms on estimation quality, we change the range of user sensed values, i.e., adjusting the range of $V_{\jmath truth}^{\tau}$. The sampling interval of $V_{\jmath truth}^{\tau}$ is $[20, x]$ and we vary x from 30dB to 110dB. In Fig. 5 and Fig. 8, the MAE increases but the accuracy decreases, when x becomes larger. This is because as the sample range of $V_{\jmath truth}^{\tau}$ increasing, sensed values of users becomes more diverse. Specifically, our approach can achieve about 92.39% average accuracy of estimated values when varies the sample range of $V_{\jmath truth}^{\tau}$, which is only 6.38%, 4.82% and 1.02% lower than NPP, OLSV and PLOV but 2.97% higher than PMM, respectively.

To summarize, although the performance of our approach is inevitably worse than NPP, OLSV and PLOV, our approach still achieves relatively high accuracy of estimated values and provides joint location-value privacy protection for users. Moreover, our approach always outperforms PMM, since we adopt a more reliable truth discovery method to eliminate the influence of unreliable or protected sensing data on the truth estimation.

8 Conclusion

In this work, we consider the joint location-value privacy protection problem in a MCS system with an untrusted platform, since not only location tags but also sensed values of users contained in their spatiotemporal sensing data will expose the privacy. Therefore, we propose a privacy protection approach, comprising of two privacy preserving mechanisms to perturb the locations and sensed values of users respectively. Specially, the LPPM is designed based on random response, and the VPPM is designed based on Gaussian mechanism. Both of the two mechanisms are proved to satisfy local differential privacy. Moreover, we conduct extensive simulations to show that the true values in interested locations can be accurately estimated based on perturbed locations and sanitized sensed values, by adopting the truth discovery method.

Acknowledgements. This research is supported by Grant No. 61802245 from NSFC and Grant No. 20CG47 from Shanghai Chen Guang Program. We also appreciate the High Performance Computing Center of Shanghai University and Shanghai Engineering Research Center of Intelligent Computing System (No. 19DZ2252600) for providing the computing resources.

References

1. Abowd, J.M.: The US census Bureau adopts differential privacy. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 2867 (2018)
2. Dwork, C.: Differential privacy. In: Henk, C., van Tilborg, A., Jajodia, S. (eds.) Encyclopedia of Cryptography and Security pp. 338–340 (2011). https://doi.org/10.1007/978-1-4419-5906-5_752
3. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Found. Trends® Theor. Comput. Sci. **9**(3–4), 211–407 (2014)
4. Ganti, R.K., Ye, F., Lei, H.: Mobile crowdsensing: current state and future challenges. IEEE Commun. Mag. **49**(11), 32–39 (2011)
5. Jin, H., Su, L., Ding, B., Nahrstedt, K., Borisov, N.: Enabling privacy-preserving incentives for mobile crowd sensing systems. In: 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS), pp. 344–353. IEEE (2016)
6. Kairouz, P., Oh, S., Viswanath, P.: Extremal mechanisms for local differential privacy. In: Advances in neural information processing systems, pp. 2879–2887 (2014)
7. Kouicem, D.E., Bouabdallah, A., Lakhlef, H.: Internet of things security: A top-down survey. Comput. Netw. **141**, 199–221 (2018)
8. Li, Q., Li, Y., Gao, J., Zhao, B., Fan, W., Han, J.: Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, pp. 1187–1198. ACM (2014)

9. Li, Y., et al.: Towards differentially private truth discovery for crowd sensing systems. arXiv preprint arXiv:1810.04760 (2018)
10. Lin, B.C., Wu, S.H., Tsou, Y.T., Huang, Y.: PPDCA: privacy-preserving crowdsensing data collection and analysis with randomized response. In: 2018 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE (2018)
11. Lin, J., Yang, D., Li, M., Xu, J., Xue, G.: Frameworks for privacy-preserving mobile crowdsensing incentive mechanisms. IEEE Trans. Mob. Comput. **17**(8), 1851–1864 (2017)
12. Liu, L., Liu, W., Zheng, Y., Ma, H., Zhang, C.: Third-eye: a mobilephone-enabled crowdsensing system for air quality monitoring. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **2**(1), 20 (2018)
13. Ma, H., Zhao, D., Yuan, P.: Opportunities in mobile crowd sensing. IEEE Commun. Mag. **52**(8), 29–35 (2014)
14. Maisonneuve, N., Stevens, M., Niessen, M.E., Steels, L.: NoiseTube: measuring and mapping noise pollution with mobile phones. In: Athanasiadis, I.N., Rizzoli, A.E., Mitkas, P.A., Gómez, J.M. (eds.) Information Technologies in Environmental Engineering. Environmental Science and Engineering, pp. 215–228. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-88351-7_16
15. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: FOCS, vol. 7, pp. 94–103 (2007)
16. Pournajaf, L., Garcia-Ulloa, D.A., Xiong, L., Sunderam, V.: Participant privacy in mobile crowd sensing task management: a survey of methods and challenges. ACM Sigmod Record **44**(4), 23–34 (2016)
17. Ren, X., et al.: LoPub: high-dimensional crowdsourced data publication with local differential privacy. IEEE Trans. Inf. Forensics Secur. **13**(9), 2151–2166 (2018)
18. To, H., Ghinita, G., Shahabi, C.: A framework for protecting worker location privacy in spatial crowdsourcing. Proc. VLDB Endow. **7**(10), 919–930 (2014)
19. Wang, L., Zhang, D., Yang, D., Lim, B.Y., Han, X., Ma, X.: Sparse mobile crowdsensing with differential and distortion location privacy. IEEE Trans. Inf. Forensics Secur. **15**, 2735–2749 (2020)
20. Wang, L., Zhang, D., Yang, D., Lim, B.Y., Ma, X.: Differential location privacy for sparse mobile crowdsensing. In: 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 1257–1262. IEEE (2016)
21. Wang, Q., Zhang, Y., Lu, X., Wang, Z., Qin, Z., Ren, K.: Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy. IEEE Trans. Dependable Secure Comput. **15**(4), 591–606 (2016)
22. Wang, Q., Zhang, Y., Lu, X., Wang, Z., Qin, Z., Ren, K.: RescueDP: real-time spatio-temporal crowd-sourced data publishing with differential privacy. In: IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, pp. 1–9. IEEE (2016)
23. Wang, S., et al.: Local differential private data aggregation for discrete distribution estimation. IEEE Trans. Parallel Distrib. Syst. **30**, 2046–2059 (2019)
24. Wang, T., Zhao, J., Zhang, X., Yang, X.: A comprehensive survey on local differential privacy toward data statistics and analysis in crowdsensing. arXiv preprint arXiv:2010.05253 (2020)
25. Wang, Z., et al.: Personalized privacy-preserving task allocation for mobile crowdsensing. IEEE Trans. Mob. Comput. **18**(6), 1330–1341 (2018)
26. Yang, X., Wang, T., Ren, X., Yu, W.: Survey on improving data utility in differentially private sequential data publishing. IEEE Trans. Big Data **7**, 729–749 (2017)

27. Zhang, M., Yang, L., Gong, X., Zhang, J.: Privacy-preserving crowdsensing: privacy valuation, network effect, and profit maximization. In: 2016 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE (2016)
28. Zhang, X., Yang, Z., Liu, Y.: Vehicle-based bi-objective crowdsourcing. IEEE Trans. Intell. Transp. Syst. **99**, 1–9 (2018)
29. Zhao, B., Rubinstein, B.I., Gemmell, J., Han, J.: A Bayesian approach to discovering truth from conflicting sources for data integration. Proc. VLDB Endow. **5**(6), 550–561 (2012)