RoPriv: Road Network-aware Privacy-preserving Framework in Spatial Crowdsourcing

Mengyuan Wang, Hongbo Jiang, *Senior Member, IEEE,* Ping Zhao, *Member, IEEE,* Jie Li, *Member, IEEE,* Jiangchuan Liu, *Fellow, IEEE* Geyong Min, *Senior Member, IEEE,* Schahram Dustdar, *Fellow, IEEE*

Abstract—Spatial Crowdsourcing (SC) has been an indispensable Location-based Service where the SC server assigns tasks to workers based on the locations of task requesters and workers, raising strong privacy concerns. Limited by the computational and time complexity, existing works prefer differential privacy-based methods to protect location privacy. However, most differential privacy-based works ignore the road network, perturbing locations on two-dimensional plane, resulting in more failures in tasks and moreover extensive privacy disclosure in practice. This paper aims to implement a multi-task assignment with both high utility and efficiency while protecting the location privacy of both task requesters and workers on road networks. Specifically, we design a Road Network-aware Exponential Mechanism and propose an Obfuscated Locations Selection algorithm to guarantee location privacy of all participants and extensive privacy. Then, we propose region distance. Based on this, we further formulate multi-task assignment as a Binary Linear Programming problem and a utility-aware optimization problem. We solve the first problem to obtain optimal efficiency and then propose a utility-aware optimization algorithm for the second problem to improve the utility. Our experiments demonstrate sufficient and stable privacy guarantee and the well-performance on both utility and efficiency of our framework.

Index Terms—Spatial Crowdsourcing, Location privacy, Road network, Multi-task assignment.

1 Introduction

With the development of mobile devices and GPS positioning, Spatial Crowdsourcing (SC) [1–16] has become an indispensable Location-based Service in people's lives, such as ride-hailing services (e.g., Uber, Didi), food delivery (e.g., Grubhub, Meituan) [17, 18]. In the SC system, multiple tasks containing locations are sent to the SC server by task requesters. Then, based on workers' locations, the SC server assigns these tasks to multiple suitable workers and these assigned workers need to complete corresponding tasks on time.

However, the information exchange in SC can lead to the leakage of participants' (including task requesters and workers) location information [12–16], which poses severe threats to their privacy and security. Specifically, 1) The SC server may leak participants' exact locations to advertisers or vicious companies. 2) Malicious workers may acquire tasks' exact locations even though without being assigned, which threatens the location privacy of task requesters.

The existing works concerning location privacy protection in SC can be primarily classified into *differential privacy-based frameworks* [1, 2, 12–15, 17] and *cryptography-based frameworks* [3, 19–22]. Limited by the computational and time complexity in multi-task assignment, these existing works prefer the differential privacy-based algorithms to those based on cryptography. However, most of the differential privacy-based works [1, 2, 12–15, 17] perturb locations based on the Euclidean distances on the two-dimensional plane rather than the distances on road networks, and thus the behavior of perturbing data results in *more failures of tasks* and *extensive privacy disclosure* (e.g., personality traits far beyond location privacy) [23]. Several differential privacy-based works [1, 17], considering the distances on road networks, *ignored the location privacy of task requesters*, and moreover *cannot guarantee the location privacy of workers in worker-dense areas* and *the data utility in the remote areas as the perturbed locations are significantly far from the exacted ones in the remote suburb*.

To address these problems above, we design a differential privacy-based framework that perturbs locations based on the distances on road networks, protecting both the location privacy of task requesters and workers. Moreover, it preserves the behavior of perturbing locations and thereby the extensive privacy. Our main idea is that the locations of task requesters and workers are perturbed locally, and on the basis of the perturbed locations, the SC server obtains a well-performed multi-task assignment on both efficiency (i.e., average travel distance) and utility (i.e., assignment success rate). Though the basic idea sounds straightforward, we are facing the following challenges.

This study is supported in part by NSFC Nos. U20A20181, Hunan Provincial Natural Science Foundation (2020JJ4211, 2020JJ5089, 2022JJ40878), the science and technology innovation Program of Hunan Province under grant 2021RC4023, Key Research and Development Program of Hunan Province under grant 2021WK2001 and the Postgraduate Scientific Research Innovation Project of Hunan Province under Grant CX20220411. (Corresponding authors: Hongbo Jiang, Ping Zhao, Jie Li.)

M. Wang and H. Jiang are with College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China. Emails: wmy1997@hnu.edu.cn, hongbojiang2004@gmail.com.
P. Zhao is with College of Information Science and Technology, Donghua University. Shanghai 201620, China. Email: pingzhao2014ph@gmail.com.
J. Li is with the Department of School of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China. E-mail: jieli2014@hnu.edu.cn.
J. Liu is with School of Computing Science, Simon Fraser University, Burnaby, BC V5A1S6, Canada; and also with Jiangxing Intelligence Inc, Nanjing 210000, China. Email: jcliu@cs.sfu.ca.
G. Min is with College of Engineering, Mathematics and Physical Sciences, University of Exeter EX4 4QF, U.K. Email: g.min@exeter.ac.uk.
S. Dustdar is with the Distributed Systems Group, TU Wien, 1040 Vienna Austria. Email: dustdar@dsg.tuwien.ac.at.
Digital Object Identifier XXXX/TMC.XXXX.XXXXXX

1) It is nontrivial to design a differential privacy-based mechanism for location privacy of both task requesters and workers on road networks, since the topology of road networks is irregular rather than a continuous two-dimensional plane. To this end, we propose a location obfuscation scheme on road networks. It first samples the road network into discrete locations. Then, based on that, it specially defines ϵ-RN-differential privacy. After that, it utilizes a Road Network-aware Exponential Mechanism (RNEM), we proposed, to perturb real locations of task requesters and workers on road networks, which is theoretically proved to be subordinate to ϵ-differential privacy. Furthermore, perturbed locations of task requesters and workers are well restricted to road networks by our mechanism, based on which their behaviors of perturbing data are also preserved from disclosure.

2) It is hard to implement the multi-task assignment on perturbed locations of both task requesters and workers, as there are significant errors between distances among perturbed locations and distances among real locations for tasks requesters and workers. For that, we propose the region distance based on Bayesian inference to replace the distance among perturbed locations. Furthermore, we formulate the multi-task assignment on region distances as a Binary Linear Programming (0-1 LP) problem, based on which we can obtain the assignment with the minimal average travel distance (ATD).

3) It is complicated to implement a multi-task assignment with both high utility and efficiency, since minimal ATD does not promise high assignment success rate (ASR). Therefore, we attempt to exchange workers' tasks to improve ASR. Yet, task exchange will break out the assignment for the minimal ATD, where there will be a increase in ATD. For that, we first set an upper threshold η^τ for the increase rate η of ATD compared to minimal ATD, and then formulate a task exchange problem to increase ASR while keeping the increase of ATD within η^τ. To solve this problem, we propose the ASR-aware Optimization algorithm to improve assignment success rate, where the increase rate of ATD $\eta \leq \eta^\tau$.

In addition, we conduct extensive experiments on the real taxi dataset from the Roma [24] and set several representative frameworks [17, 18, 25] as baselines to investigate the location privacy, efficiency and utility of multi-task assignment, compared with our framework. The experimental results indicate that our framework can provide sufficient and stable location privacy protection for both task requesters and workers on road networks, whether downtown or in a remote suburb. Compared to the state-of-the-art, our framework performs much better on average travel distance, where ATD < 1.0km downtown and ATD < 3.75km in a remote suburb. Furthermore, our framework achieves both high efficiency and utility in multi-task assignment. Specifically, in extreme cases, our framework can increase the ASR (i.e., utility) by 17.2% while keeping the ATD (i.e., efficiency) growth less than 5%.

The rest of this paper is organized as follows. We demonstrate the related work in Section. 2. Then, we introduce preliminaries in Section. 3 and raise the problem statement of our framework in Section. 4. An overview of our framework is demonstrated in Section. 5 and then, we present details of privacy protection in our framework in Section 6 and multi-task assignment on obfuscated locations in Section 7. Thereafter, we evaluate the performance of our framework in Section. 8. Finally, we draw the conclusion in Section. 9.

2 Related Work

In this section, we briefly review the related work.

2.1 Cryptography-based Frameworks

To protect location privacy in SC, several methods based on cryptography have been developed to encrypt location information while implementing multi-task assignment. Shu et al. [22] proposed a non-interactive privacy-preserving task recommendation framework (PPTR), where the locations of both task requesters and workers were encrypted to protect their location privacy, and multiple task requesters would be matched with multiple workers. In 2019, Yuan et al. [21] proposed an efficient task assignment algorithm (PriRadar), aiming at improving the assigned time by instantly assigning tasks to nearby workers while protecting the data of tasks and workers. Based on encrypted locations, iTAM [3], proposed by Zhao et al., focused on minimizing travel distance of workers. Li et al. [19] designed a grid-based privacy-preserving framework for online SC (GPSC) to obtain a trade-off between efficiency and security considering the preferences of task requesters. According to investigated interests of workers, Song et al. [20] proposed a privacy-preserving task matching framework (PPTM) to achieve efficient task matching, meanwhile, protect the privacy of locations and interests of task requesters and workers.

These cryptography-based works achieve reliable location privacy protection by encrypting the locations of task requesters and workers. However, constrained by the computational and communication complexity, cryptography-based frameworks are difficult to be implemented in practice. Therefore, differential privacy-based frameworks are more preferred in Spatial Crowdsourcing.

2.2 Differential Privacy-based Frameworks

In SC, locations of task requesters and workers are involved in multi-task assignment, in which location privacy of task requesters and workers may be disclosed to adversaries. Based on differential privacy and geocasting, To et al. [26] assumed a trusted Cell Service Provider (CSP) to provide location privacy for workers, which is the first work focused on location privacy issues in SC. After that, to improve the overhead of the assignment, To et al. [2] improved the framework [26] by factoring the geocast system overhead. Nevertheless, the trusted third parties were not practical. For that, Wang et al. [12, 18] focused on protecting workers' location privacy without involving any trusted third parties while achieving the optimal task assignment with minimal travel distance. Unfortunately, the location privacy of task requesters has not been protected in [2, 12, 18, 26]. Wei et al. [15] proposed a differential privacy-based location protection framework (DPLP), aiming to achieve the trade-off between utility and system overhead while protecting the location privacy of both task requesters and workers.

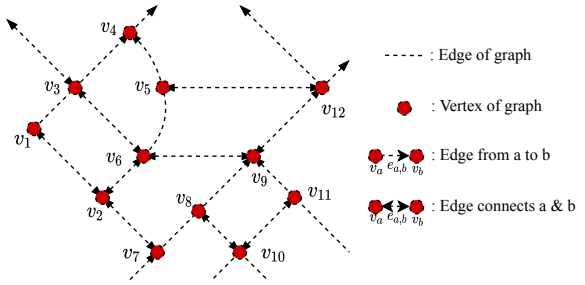


Fig. 1. A part of road networks abstracted from real city traffic map.

Tao et al. [13] designed a novel privacy mechanism based on Hierarchically Well-Separated Trees to minimize the total distance for the SC system. However, all these works above perturbed locations on the two-dimensional plane and ignored the road network, which might lead to the disclosure of the behavior of perturbing locations. Furthermore, due to ignoring the road network, there would be more workers unable to complete tasks because detour results in more failures tasks. Qiu et al. [1, 17] considered road networks and designed a differential privacy-based framework to protect workers' location privacy. But it was a pity that they ignored the location privacy of task requesters, and moreover could not guarantee the data utility in remote areas.

To tackle these problems above, our framework takes into account the road networks and designs a differential privacy-based mechanism to protect the location privacy of both task requesters and workers, while achieving a high assignment success rate and a low average travel distance. Moreover, with perturbed locations constrained to the road network, participants' behaviors of perturbing locations are completely preserved from disclosure.

3 PRELIMINARIES

In this section, we first demonstrate the road network model adopted in our privacy-preserving framework. Then, we introduce the de facto standard ϵ-differential privacy. Finally, several metrics represent the location privacy, the efficiency and the utility of multi-task assignment.

3.1 Road Network Model

With reference to the related work [6, 27], we use directed weighted graph to process the information of the road network. As shown in Fig. 1, we extract the city's road networks structure from the real city traffic map, which consists of critical locations and roads. We represent the road networks by directed weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where locations and roads are represented by the set of vertices $\mathcal{V}$ and the set of edges $\mathcal{E}$, respectively. Each edge and vertex are presented by $e_{i,j} \in \mathcal{E}$ and $v_i \in \mathcal{V}$ respectively, where v_i denotes the vertex numbered i and $e_{i,j}$ denotes the edge connected by v_i and v_j. For example, in Fig. 1, $\mathcal{V} = \{v_1, v_2, ..., v_{12}\}$ and $\mathcal{E} = \{e_{1,2}, e_{1,3}, e_{2,1}, ..., e_{12,9}\}$, which be worth noting that $e_{1,2}$ and $e_{2,1}$ are considered as different edges in directed weighted graph. We assume that both task requesters and workers are located in the road networks $\mathcal{G}$, where l_i^t and l_i^w denote the task requester and the worker's locations respectively.

It is hard to abstract road networks as a directed weighted graph from the real city traffic map. Thanks to OpenStreetMap [1] and igraph [2], we overcome this complicated problem and easily build the road network model for our framework.

3.2 ϵ-Differential Privacy

ϵ-differential privacy [28, 29], as we know, has been a de facto standard privacy-preserving conception in recent years, which can provide a provable privacy guarantee. Specifically, if a mechanism satisfies ϵ-differential privacy, with two inputs of adjacent datasets, the outputs can not be distinguished by an adversary with side information.

Definition 1 (ϵ-Differential Privacy). Given two adjacent datasets $\mathcal{D}$, $\mathcal{D}^{'}$ and $\epsilon > 0$, a privacy mechanism K satisfies ϵ-differential privacy iff for different records x and $x^{'}$:

$$\frac{\Pr[K(x \in \mathcal{D}) = y]}{\Pr[K(x^{'} \in \mathcal{D}^{'}) = y]} \leq e^{\epsilon}, y \in \mathcal{W}, \quad (1)$$

where $K(x \in \mathcal{D})$ and $K(x^{'} \in \mathcal{D}^{'})$ differ in an individual record of inputs, ϵ is the privacy budget, the smaller ϵ, the higher privacy.

3.3 Performance Metrics

In SC, the obfuscated locations of task requesters and workers guarantee location privacy. However, it causes the dilemma that noisy locations do harmful to multi-task assignment. With this, we need to consider two aspects of our work, location privacy (i.e., Expected Estimation Error (E3) [25, 30, 31]) and multi-task assignment, which is evaluated with two metrics, the efficiency (i.e., Average Travel Distance (ATD) [2, 18]) and the utility (i.e., Assignment Success Rate (ASR) [2, 18]).

3.3.1 Location Privacy

The standard metric to measure location privacy is Expected Estimation Error, which represents the distance between the real location and the location inferred by adversaries. More specifically, we assume that adversaries have the side information about an obfuscated location l_o, and the side information can be expressed by a prior probability distribution on possible locations $\mathcal{W}^p$. $\Pr(w_i)$ is the probability assign to the possible location $w_i \in \mathcal{W}^p$. $\Pr(l_o|w_i)$ is the probability that the reported location l_o is converted from w_i. Based on Bayesian inference [25, 28, 32], the posterior probability model of the victim's real location can be calculated as follows:

$$\Pr(w_i|l_o) = \frac{\Pr(l_o|w_i)\Pr(w_i)}{\sum_{w_j \in \mathcal{W}^p} \Pr(l_o|w_j)\Pr(w_j)}, \ w_i \in \mathcal{W}^p. \quad (2)$$

Then, based on the posterior probability model, the adversary strives to estimate the real location with the largest posterior probability $w_{max} = arg\max_{w_i \in \mathcal{W}} \Pr(w_i|l_o)$. Expected Estimation Error is defined as follows:

1. https://www.openstreetmap.org/
2. https://igraph.org/

Definition 2 (Expected Estimation Error (E3)). We define the Expected Estimation Error as the distance between w_{max} and the real location l_r on road networks.

$$E3 = d(w_{max}, l_r) \ , \tag{3}$$

where $d(\cdot)$ denotes the Euclidean distance between two locations, and the higher degree of E3 achieved, the higher location privacy guaranteed.

3.3.2 Efficiency of Multi-task Assignment

In Spatial Crowdsourcing, Average Travel Distance (ATD) is a common metric used to measure the efficiency of task assignment [2, 18]). A worker's travel distance represents the distance between the assigned worker's true location and the task requester's true location, while the average travel distance is defined as follows:

Definition 3 (Average Travel Distance (ATD)). Assume there are N tasks in an assignment, ATD is the average value of multiple travel distances.

$$ATD = \frac{1}{N}\sum_{i=1}^{N} d_r(l_i^t, l_i^w) \ , \tag{4}$$

where $d_r(l_i^t, l_i^w)$ denotes the shortest distance between task requester's location l_i^t and worker's location l_i^w. The lower degree of ATD denotes better efficiency of multi-task assignment.

3.3.3 Utility of Multi-task Assignment

In the multi-task assignment, workers who fail in their tasks also need to be cared about. In more details, assume there are five tasks with several travel distances 3.1km, 2.4km, 1.3km, 8.2km, 0.8km, where ATD $=$ 3.16km. $d_r = 8.2$km means that the assigned worker is far away from the task requester and could not complete the task on time. Hence, we introduce Assignment Success Rate (ASR) to represent the utility of multi-task assignment. We define a success assignment as $d_r(l_i^t, l_i^w) \leq d_\tau^w$, where d_τ^w denotes the threshold of workers' acceptable distance.

Definition 4 (Assignment Success Rate (ASR)). ASR indicates the percentage of success assignments in all assignments:

$$ASR = \frac{\text{The number of success assignments}}{\text{The total number of assignments}} \ , \tag{5}$$

where we always want to obtain ASR as high as possible, while keeping the increase of ATD tolerable.

4 Problem Statement

In this section, we first describe the attack model. Then, we present the goal of our framework.

4.1 Attack Model

In Spatial Crowdsourcing, we consider locations of both task requesters and workers as private information, which needs to be protected. The potential privacy disclosure is composed of two kinds of entities, *Semi-honest SC server* and *Curious-but-honest workers*.

Semi-honest SC Server. As the existing works [11, 14, 33, 34], we consider the SC server is semi-honest. That is, the SC server will strictly execute its functions, i.e., honestly transmitting information and implementing multi-task assignment. But it may attempt to disclose task requesters and workers' location privacy. More specifically, the SC server can access the location information of all participants throughout the entire Spatial Crowdsourcing and may leak their locations to advertisers or even evil companies, who may abuse the location information of task requesters and workers. That seriously threatens task requesters and workers' location privacy.

Curious-but-honest Workers. As the existing works [14, 15], we consider workers in SC are curious-but-honest. Locations of tasks may be disclosed to the adversary by some specific workers. If a mass of workers has access to the exact locations of tasks, the location privacy of task requesters will be seriously threatened.

4.2 Goals of Our Framework

Goals of our framework consists of *Privacy Goal* and *Goal of Multi-task assignment on Obfuscated Locations*.

Privacy Goal: There are three privacy goals of our framework. **The first privacy goal** is that the simi-honest server cannot determine the real location of each participant. In particular, each location of participant will be perturbed locally before sent to the server, preventing the server from the real location. **The second privacy goal** is that anyone of the workers is not allowed to access the exact location of a task requester until one of them is assigned, where we consider the assigned worker is trusted. In order to preserve the behavior of perturbing locations, we give **the third privacy goal** that no one can identify whether a participant's location is a perturbed location.

Goal of Multi-task Assignment on Obfuscated Locations: Obfuscated locations result in many difficulties in multi-task assignment. Our goal is to implement a multi-task assignment with both high utility and efficiency. That is a high ASR and a low ATD.

5 Overview of Our Framework

As shown in Fig. 2, our framework is composed of *Location Obfuscation on Road Networks* and *Multi-task Assignment on Obfuscated Locations*.

Location Obfuscation on Road Networks: In order to achieve the privacy goal (introduced in Section. 4.2), we first need to sample the road network to generate the set $\mathcal{W}$ of possible obfuscated locations. Thereby, we propose the Obfuscated Locations Selection Algorithm 1 to generate $\mathcal{W}$ on participants' devices utilizing their real locations (cf. Section. 6.2). Then, we adopt $\mathcal{W}$ as the basis of our privacy-preserving mechanism and design the Road Network-aware Exponential Mechanism (RNEM) to locally perturb locations of task

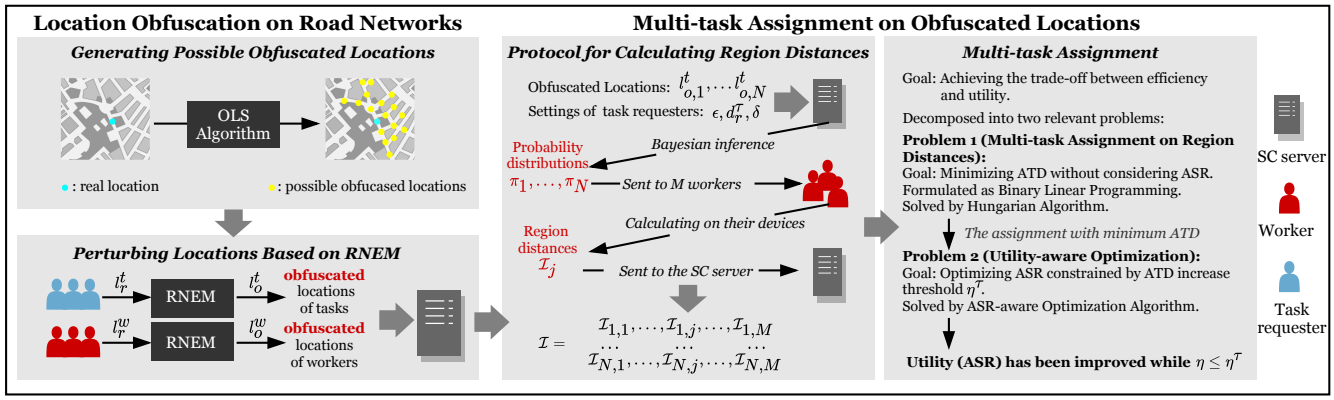


Fig. 2. Overview of RoPriv.

requesters and workers, satisfying ϵ-differential privacy (cf. Section. 6.1).

Multi-task Assignment on Obfuscated Locations: There are significant errors on distances among obfuscated locations, making it hard to obtain a well-performed multi-task assignment on the SC server. For that, in Section. 7.1, we design region distances to replace these distances among obfuscated locations and propose a protocol to calculate region distances utilizing workers' real locations without disclosing their location privacy. Then, to obtain high utility and efficiency, we decompose Multi-task Assignment on Obfuscated Locations into two relevant problems, Multi-task Assignment on Region Distances (P1) in Section. 7.2 and Utility-aware Optimization (P2) in Section. 7.3. Firstly, to solve P1, we formulate it into a Binary Linear Programming (0-1 LP) problem based on region distances and adopt the Hungarian algorithm to obtain the assignment with minimum ATD. Since improving ASR may increase ATD, we formulate this problem to maximize ASR subject to the increase threshold η^{τ} of ATD. To solve P2, we propose the ASR-aware Optimization Algorithm 3 to exchange tasks between failed workers and successful assigned workers to improve ASR while keeping the increase rate η of ATD within the threshold η^{τ}.

6 Location Obfuscation on Road Networks

To protect location privacy on road networks in SC, we need to consider two issues. First, the irregular topology of road networks makes it difficult to design a privacy-preserving mechanism satisfying ϵ-differential privacy on road networks. Second, this mechanism needs to guarantee the location privacy of both task requesters and workers. Motivated by several related works [25, 35, 36], we employ exponential mechanism and replace the road network with discrete locations to solve the first issue in Section. 6.1. Then, in Section. 6.2, we propose the Obfuscated Locations Selection Algorithm 1 to select discrete locations locally based on participants' locations without considering any external parameter, which solves the second issue.

6.1 Road Network-aware Exponential Mechanism

With a set $\mathcal{W}$ of discrete locations abstracted from the road network, we consider $\mathcal{W}$ as the set of possible obfuscated locations of task requesters and workers. Then, we give the definition of differential privacy on road networks concisely.

Definition 5 (ϵ-RN-differential privacy). A mechanism K satisfies ϵ-RN-differential privacy iff for all $l_x, l_x^{'}$:

$$\Pr[K(l_x) = l_y] \leq \exp(\epsilon d_r(l_x, l_x^{'})/\Delta d_r)\Pr[K(l_x^{'}) = l_y] \,, \tag{6}$$

where $l_y \in \mathcal{W}$, $d_r(\cdot)$ represent the shortest distance between two locations l_x, $l_x^{'}$ and Δd_r denotes the sensitivity of $d_r(\cdot)$, defined as $max_{w_i}(d_r(l_x, w_i)), w_i \in \mathcal{W}$.

Corollary 1. If a mechanism K satisfies ϵ-RN-differential privacy, K must also satisfy ϵ-differential privacy [29].

Proof 1. Assume a mechanism K satisfying ϵ-RN-differential privacy (cf. Eq. (6)). We know that $\Delta d_r \geq d_r(\cdot)$, which contribute to $\epsilon d_r(l_x, l_x^{'})/\Delta d_r \leq 1$. Then, $exp(\epsilon d_r(l_x, l_x^{'})/\Delta d_r) \leq exp(\epsilon)$ is proved. After that, the mechanism K satisfies as follows:

$$\Pr[K(l_x) = l_y] \leq \exp(\epsilon)\Pr[K(l_x^{'}) = l_y] \,. \tag{7}$$

Therefore, the mechanism K satisfies ϵ-differential privacy.

With the set of discrete possible obfuscated locations $\mathcal{W}$, it is straightforward to utilize exponential mechanism [37, 38] to perturb task requesters and workers' real locations. Considering the distance between locations on road networks, we design the Road Network-aware Exponential Mechanism as follows.

Definition 6 (Road Network-aware Exponential Mechanism (RNEM)). For an input location l_x and output location $l_y \in \mathcal{W}$, the exponential mechanism K randomly selects l_y as follows:

$$\Pr[K(l_x) = l_y] = \frac{\exp(\epsilon d_r(l_x, l_y)/2\Delta d_r)}{\sum_{l_{y,i} \in \mathcal{W}} \exp(\epsilon d_r(l_x, l_{y,i})/2\Delta d_r)} \,, \tag{8}$$

where $d_r(l_x, l_y)$ denotes the shortest distance between l_x and l_y on road networks $\mathcal{G}$.

Corollary 2. RNEM satisfies ϵ-RN-differential privacy and ϵ-differential privacy.

Proof 2. With the real location l_x and the obfuscated location $l_x^{'}$ as inputs, we can obtain two probabilities

$\Pr[K(l_x) = l_y]$ and $\Pr[K(l_x^{'}) = l_y]$ on a output location l_y, respectively. Let's divide $\Pr[K(l_x) = l_y]$ by $\Pr[K(l_x^{'}) = l_y]$:

$$\frac{\Pr[K(l_x) = l_y]}{\Pr[K(l_x^{'}) = l_y]} = \underbrace{\frac{\exp(\epsilon\frac{d_r(l_x,l_y)}{2\Delta d_r})}{\exp(\epsilon\frac{d_r(l_x^{'},l_y)}{2\Delta d_r})}}_{\text{Part A}} \underbrace{\frac{\sum_{l_{y,i}\in\mathcal{W}}\exp(\epsilon\frac{d_r(l_x^{'},l_{y,i})}{2\Delta d_r})}{\sum_{l_{y,i}\in\mathcal{W}}\exp(\epsilon\frac{d_r(l_x,l_{y,i})}{2\Delta d_r})}}_{\text{Part B}} .$$

The part A can be calculated as follows:

$$\text{Part A} = \exp(\epsilon\frac{d_r(l_x,l_y) - d_r(l_x^{'},l_y)}{2\Delta d_r}) .$$

By the triangular inequality, $d_r(l_x,l_y) - d_r(l_x^{'},l_y) \leq d_r(l_x,l_x^{'})$. Then, we obtain

$$\text{Part A} \leq \exp(\epsilon d_r(l_x,l_x^{'})/2\Delta d_r) .$$

Then, we assume Part B $\leq \exp(\epsilon d_r(l_x,l_x^{'})/2\Delta d_r)$.

$$\frac{\sum_{l_{y,i}\in\mathcal{W}}\exp(\epsilon\frac{d_r(l_x^{'},l_{y,i})}{2\Delta d_r})}{\sum_{l_{y,i}\in\mathcal{W}}\exp(\epsilon\frac{d_r(l_x,l_{y,i})}{2\Delta d_r})} \leq \exp(\epsilon d_r(l_x,l_x^{'})/2\Delta d_r) ,$$

$$\underbrace{\sum_{l_{y,i}\in\mathcal{W}}\frac{\exp(\epsilon\frac{d_r(l_x^{'},l_{y,i})}{2\Delta d_r})}{\exp(\epsilon\frac{d_r(l_x,l_x^{'})}{2\Delta d_r})}}_{\text{Left}} \leq \sum_{l_{y,i}\in\mathcal{W}}\exp(\epsilon\frac{d_r(l_x,l_{y,i})}{2\Delta d_r}) . \quad (9)$$

If Eq. (9) is true, we can confirm our assumption of Part B. Hereby, we deduce the Left of Eq. (9) as follows:

$$\text{Left} = \sum_{l_{y,i}\in\mathcal{W}}\exp(\epsilon(d_r(l_x^{'},l_{y,i}) - d_r(l_x,l_x^{'}))/2\Delta d_r) . \quad (10)$$

By the triangular inequality, $d_r(l_x^{'},l_{y,i}) - d_r(l_x,l_x^{'}) \leq d_r(l_x,l_{y,i})$, the Eq. (10) is deduced as follows:

$$\text{Left} \leq \sum_{l_{y,i}\in\mathcal{W}}\exp(\epsilon d_r(l_x,l_{y,i})/2\Delta d_r) .$$

Therefore, our assumption of Part B is confirmed and we multiple Part A and Part B to obtain the inequation as follows:

$$\text{Part A} \cdot \text{Part B} \leq \exp(\epsilon\frac{d_r(l_x,l_x^{'})}{2\Delta d_r}) \cdot \exp(\epsilon\frac{d_r(l_x,l_x^{'})}{2\Delta d_r}) ,$$

$$\Pr[K(l_x) = l_y] \leq \exp(\epsilon d_r(l_x,l_x^{'})/\Delta d_r)\Pr[K(l_x^{'}) = l_y] .$$

Therefore, RNEM satisfies ϵ-RN-differential privacy. Furthermore, RNEM satisfies ϵ-differential privacy due to Corollary 1.

6.2 Generating $\mathcal{W}$ for RNEM

In related works [1, 17, 35], there are two methods to generate the set $\mathcal{W}$ of possible obfuscated locations. The first method [35] simply employs vertices of road networks nearby the real locations to represent $\mathcal{W}$. However, vertices may be reused, which will cause extensive privacy disclosure [23] (i.e., the behavior of perturbing locations). The second method [1, 17] employs locations uploaded by a certain range of workers to represent $\mathcal{W}$. Based on this method, the effect of location privacy protection is heavily affected by the density of workers. Furthermore, this method cannot be applied to the location privacy protection of task requesters.

For that, to generate possible obfuscated locations $\mathcal{W}$, Obfuscated Locations Selection algorithm, we proposed, equably selects discrete locations based on participants' locations and their other settings (i.e., the maximum distance threshold d_r^τ (m), the sample interval $\delta = d_r^\tau/10$ (m)), as shown in Algorithm 1.

Algorithm 1: Obfuscated Locations Selection (OLS).

Input : $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, $v \in \mathcal{V}$, d_{temp}, $\mathcal{V}_a$, $\mathcal{W}$

1 **if** v *not in* $\mathcal{V}_a$ *and* $d_{temp} \leq d_r^\tau$ **then**
2 Add v into $\mathcal{V}_a$;
3 $\mathcal{V}_n$ = adjacent($\mathcal{G}$, v) ;
4 Remove vertices in $\mathcal{V}_a$ from $\mathcal{V}_n$;
5 $d_{tail} = d_{temp}$ mod δ ;
6 **for** v_i *in* $\mathcal{V}_n$ **do**
7 $N_{v_i,v} = (d_r(v_i, v) - d_{tail})$ / δ ;
8 **for** $k = 1$ *to* $N_{v_i,v}$ **do**
9 Select l_k^o based on Eq. (11) ;
10 **if** $d_r(v, l_k^o) + d_{temp} \leq d_r^\tau$ **then**
11 Add l_k^o into $\mathcal{W}$;
12 **end**
13 **end**
14 OLS($\mathcal{G}$, v_i, $d_{temp} + d_r(v_i, v)$, $\mathcal{V}_a$, $\mathcal{W}$) ;
15 **end**
16 **end**

Algorithm 2: Initial of OLS.

Input : $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, $l_r \in \mathcal{V}$
Output: $\mathcal{W}$

1 Initialize d_{temp}, $\mathcal{V}_a$, $\mathcal{W}$ = 0, $\emptyset$, $\emptyset$;
2 OLS($\mathcal{G}$, l_r, d_{temp}, $\mathcal{V}_a$, $\mathcal{W}$) ;
3 Return $\mathcal{W}$;

Initial of OLS. Before performing OLS, some variables need to be predefined. As shown in Algorithm 2, we first initialize the variable d_{temp} and two lists $\mathcal{V}_a$, where d_{temp} is used to record the current distance between real location l_r and the current vertex, $\mathcal{V}_a$ is used to store the vertices in the road network model $\mathcal{G}$ visited by Algorithm 1. Thereafter, we execute Algorithm 1 with the input composed of $\mathcal{G}$, l_r, d_{temp}, $\mathcal{V}_a$ and $\mathcal{W}$. l_r denotes the real location, which has been embedded in $\mathcal{G}$ as a vertex.

Algorithm 1 is a recursive algorithm. In line 1, the algorithm determines whether the current vertex v has been visited and whether d_{temp} exceeds the distance threshold d_r^τ. If v has not been visited and $d_{temp} \leq d_r^\tau$, the algorithm continues to execute lines 2-16. Then, the algorithm records the current vertex as a visited vertex by adding this vertex v into $\mathcal{V}_a$. Based on the current vertex v and the road network $\mathcal{G}$, the algorithm obtains adjacent vertices $v_i \in \mathcal{V}_n$ in line 3 and removes the visited vertices from $\mathcal{V}_n$ in line 4. In addition, the algorithm obtains d_{tail} by taking the remainder of d_{temp} divided by δ in line 5, where d_{tail} is used to keep the distance between the first selected location on the current edge and the previous location equal to δ. Thereafter,

the algorithm is ready to select locations in this recursion. Based on the adjacent vertex v_i, the algorithm tries to extract locations on the edge of v and v_i in lines 7-13. Firstly, the algorithm obtains the number $N_{v_i,v}$ of locations by dividing this edge as $N_{v_i,v} = (d_r(v_i,v) - d_{tail}) \ / \ \delta$. Then, there are $N_{v_i,v}$ locations on this edge generated as follows:

$$\begin{aligned} \alpha &= (d_{tail} + k \times d_r^\tau) \ / \ d_r(v_i,v) \ , \\ x_k &= \alpha(x_{v_i} - x_v) + x_v \ , \\ y_k &= \alpha(y_{v_i} - y_v) + y_v \ , \end{aligned} \tag{11}$$

where x and y denote the latitude and longitude, $k \in \{1, ..., N_{v_i,v}\}$. After that, in line 10, the algorithm needs to detect if these locations are out of the distance threshold d_r^τ. If not, the algorithm adds locations into $\mathcal{W}$ as possible obfuscated locations. Finally, the algorithm updates the current distance d_{temp} by $d_{temp} + d_r(v_i,v)$ and reset the current vertex by v_i, and then performs recursion with the latest input. The algorithm will perform recursion until all suitable locations are selected, which compose the set $\mathcal{W}$ of possible obfuscated locations.

The time complexity of OLS. We assume that each recursion has N adjacent vertices on average and the depth of OLS is M. In OLS, the time complexity mainly includes recursion (the time complexity is $\mathcal{O}(N^M)$) and locations selection on adjacent vertices in lines 6-13 (the time complexity upper to $\mathcal{O}(N \times N_{v_i,v})$). Therefore, the time complexity is $\mathcal{O}((N \times N_{v_i,v})^M)$. Fortunately, M is at most 10 in practice, limited by the distance threshold d_r^τ and sample interval $\delta = d_r^\tau/10$. Moreover, the relationship between $N_{v_i,v}$ and M is inversely proportional. That is, the larger the M, the smaller the $N_{v_i,v}$, when M is 10, $N_{v_i,v}$ is basically 1. In addition, in the road network $\mathcal{G}$, N is 4 at most, 0 at minimum, and 1.5 on average. Therefore, in practice, the time complexity is less than $\mathcal{O}(4^{10})$, and the average complexity is $\mathcal{O}(1.5^{10})$, which is a small time cost.

7 Multi-task Assignment on Obfuscated Locations

With obfuscated locations of task requesters and workers, it is complicated to perform well on the multi-task assignment as far as to obtain high utility and efficiency. Most of existing works assume trusted third parties [2, 26] or the real location of the task requester accessible [1, 17, 18], which is not practical in real life. To implement a multi-task assignment on obfuscated locations, we design the region distance and propose a protocol to calculate it with the help of Bayesian inference (in Section. 7.1). Then, we formulate the multi-task assignment as two relevant problems. The first problem is a Binary Linear Programming problem solved by the Hungarian algorithm to obtain the assignment with optimal efficiency (in Section. 7.2). The second problem is formulated to improve ASR while keeping the increase of ATD within a threshold, solved by the ASR-aware Optimization algorithm proposed by us (in Section. 7.3).

7.1 Region Distance Model

Travel distance (introduced in Section. 3.3) plays an important role in multi-task assignment [1, 2, 12–15, 17, 18]. However, there are significant errors on distances among obfuscated locations compared to distances among real locations. To solve this problem, we design region distance as the distance between a worker's real location and a task requester's obfuscated location to better implement multi-task assignment. We firstly define as follows:

Definition 7 (Region Distance ($\mathcal{I}$)). Given the real location probability distribution π_i of a task requester $l_{o,i}^t$ and the real location $l_{r,j}^w$ of a worker, where π_i consists of possible real locations $l_{r,i,1}^{t,p}$, $l_{r,i,2}^{t,p}$,..., $l_{r,i,|\pi_i|}^{t,p}$ and their corresponding probabilities $\Pr_{r,i,1}^p$, $\Pr_{r,i,2}^p$,..., $\Pr_{r,i,|\pi_i|}^p$, we define the region distance $\mathcal{I}_{i,j}$ as the weighted distance between the real location $l_{r,j}^w$ of the worker and the probability distribution π_i of the task requester $l_{o,i}^t$:

$$\mathcal{I}_{i,j} = \sum_{k=1}^{N} \Pr_{i,k}^{'} d_r(l_{r,j}^w, l_{i,k}^{'}) \ , \tag{12}$$

where $d_r(l_{r,j}^w, l_{i,k}^{'})$ denotes the shortest distance between $l_{r,j}^w$ and $l_{i,k}^{'}$ on road networks $\mathcal{G}$, π_i is generated by Bayesian inference and the calculation of region distances is arranged on workers' device, where workers' real locations can be adopted without disclosing their location privacy.

Assume there are M unoccupied workers $l_{r,1}^w$, $l_{r,2}^w$,..., $l_{r,M}^w$ and N tasks waiting for assignment $l_{o,1}^t$, $l_{o,2}^t$,..., $l_{o,N}^t$, where $M \geq N$. Then, we design a protocol to calculate region distance without disclose workers' location privacy as follows.

1) The SC server calculates real location probability distribution $\pi_1, \pi_2, ..., \pi_N$ based on each real location of task requesters.
2) The SC server sends $\pi_1, \pi_2, ..., \pi_N$ to M workers.
3) Each worker generates a region distance vector $\mathcal{I}_j$=$[\mathcal{I}_{1,j}, \mathcal{I}_{2,j}, ..., \mathcal{I}_{N,j}]^T$, $1 \leq j \leq M$ based on Eq. (12) and its real location.
4) Each worker sends its region distance vector to the SC server.

Therefore, our protocol preserves the location privacy of workers while employing their real locations to calculate region distances. Finally, the SC server obtains M region distance vectors constructed as follows:

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_{1,1} & \mathcal{I}_{1,2} & \cdots & \mathcal{I}_{1,j} & \cdots & \mathcal{I}_{1,M} \\ \mathcal{I}_{2,1} & \mathcal{I}_{2,2} & \cdots & \mathcal{I}_{2,j} & \cdots & \mathcal{I}_{2,M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathcal{I}_{i,1} & \mathcal{I}_{i,2} & \cdots & \mathcal{I}_{i,j} & \cdots & \mathcal{I}_{i,M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathcal{I}_{N,1} & \mathcal{I}_{N,2} & \cdots & \mathcal{I}_{N,j} & \cdots & \mathcal{I}_{N,M} \end{bmatrix} \ , \tag{13}$$

where $\mathcal{I}$ denotes the region distance model organized for multi-task assignment.

Real Location Probability Distribution: In generating region distance, it is essential to prepare the probability distribution π_i of each task's possible real location in advance, which consists of two processes.

Firstly, the SC server needs to determine possible real locations of the task requester. To handle this issue, we enable several critical parameters of the task requester public to the SC server, that is, privacy budget ϵ, distance threshold d_r^τ and sample interval δ, where even though ϵ is public, the

location privacy of the task requester can be still guaranteed by ϵ-differential privacy [28, 32]. Holding these parameters with the obfuscated location $l^t_{o,i}$, the SC server executes OLS to generate possible real locations $\mathcal{W}^p$.

Thereafter, based on $\mathcal{W}^p$, the SC server employs Bayesian inference (explained in Section. 3.3) to calculate the probability of a possible real location $l^{t,p}_{r,i,k}$ in $\mathcal{W}^p$ as follows:

$$\Pr(l^{t,p}_{r,i,k}|l^t_{o,i}) = \frac{\Pr(l^t_{o,i}|l^{t,p}_{r,i,k})\Pr(l^{t,p}_{r,i,k})}{\sum_{l^{t,p}_{r,i,q}\in\mathcal{W}^p}\Pr(l^t_{o,i}|l^{t,p}_{r,i,q})\Pr(l^{t,p}_{r,i,q})}\ , \quad (14)$$

where $\Pr(l^{t,p}_{r,i,k}|l^t_{o,i})$ and $l^{t,p}_{r,i,k} \in \mathcal{W}^p$ construct the probability distribution π_i of the obfuscated location $l^t_{o,i}$ of a task requester.

7.2 Multi-task Assignment on Region Distance

Multi-task assignment on obfuscated locations is composed of two relevant problems, in which the first problem is multi-task assignment on region distance, aiming at minimizing ATD without considering ASR. *Note that travel distances have been replaced by region distances, and we can observe that minimizing the total region distances is equivalent to minimizing average region distance.* In addition, in Spatial Crowdsourcing, a task can be assigned to only one worker, and a worker can only accept a task at a time. Assume there are M workers and N task requesters at a time, and we formulated Problem 1 as a Linear Programming problem.

$$\begin{aligned} \min_{\mathcal{A}} \quad & \textstyle\sum_{i=1}^{N}\sum_{j=1}^{M}\mathcal{I}_{i,j}\mathcal{A}_{i,j} \\ \text{subject to} \quad & \mathcal{A}_{i,j} = 0 \text{ or } 1\ , \\ & \textstyle\sum_{j=1}^{M}\mathcal{A}_{i,j} = 1\ , \\ & \textstyle\sum_{i=1}^{N}\mathcal{A}_{i,j} \le 1\ , \end{aligned}$$

where $\mathcal{A}$ is a $N\times M$ matrix, $\mathcal{A}_{i,j} = 0$ or 1 means that the task l^t_i is assigned to ($\mathcal{A}_{i,j} = 1$) or not assigned to ($\mathcal{A}_{i,j} = 0$) the worker l^w_j, $\sum_{j=1}^{M}\mathcal{A}_{i,j} = 1$ constrains a task assigned to only one worker, and $\sum_{i=1}^{N}\mathcal{A}_{i,j} \le 1$ means that a worker can only accept at most one task. Hence, if a constrained matrix contributes to the minimum value of $\sum_{i=1}^{N}\sum_{j=1}^{M}\mathcal{I}_{i,j}\mathcal{A}_{i,j}$, this matrix is the assignment with minimum average region distance to assign multiple tasks.

However, for the constraint $\mathcal{A}_{i,j} = 0$ or 1, we observe that Problem 1 is a Binary Linear Programming (0-1 LP) problem, where $\mathcal{A}_{i,j}$ is constrained to have components equal to zero or one. This problem is not a convex problem, even though an optimal solution must exist. If adopting enumeration, we can find that even though the feasible set is finite, the computation complexity is $\mathcal{O}(2^{N\times M})$. Here, we implement the classic algorithm (i.e., Hungarian algorithm) to solve Problem 1 within polynomial-time ($\mathcal{O}(n^3)$). The algorithm is composed of four steps:

Step 1: The problem is needed to be balanced by adding $M-N$ dummy rows ($\mathcal{I}_{i,j}$, $N < i \le M-N$) into $\mathcal{I}$. Therefore, the problem is transferred as $min\sum_{i=1}^{M}\sum_{j=1}^{M}\mathcal{I}_{i,j}\mathcal{A}_{i,j}$ and the constrain item $\sum_{i=1}^{N}\mathcal{A}_{i,j} \le 1$ is reformed as $\sum_{i=1}^{M}\mathcal{A}_{i,j} = 1$;

Step 2: In each row of $\mathcal{I}$, the algorithm derive the minimum region distance $min(\mathcal{I}_i)$ and subtract it from all the elements in this row. Homoplastically, after alteration of rows, the algorithm subtract $min(\mathcal{I}_j)$ from all the elements in each column;

Step 3: The algorithm tries to cover all the zero entries ($\mathcal{I}_{k,q} = 0$) by recording multiple rows and columns with the minimum number of records and then determines whether the number of records is equal to M, if so, the optimal assignment $\mathcal{A}^*$ is completed, otherwise, the algorithm continues;

Step 4: The algorithm marks each row without records and then marks each column with records intersected by marked rows and finally marks rows with records intersected by marked columns. After that, the algorithm subtracts the minimum entry from all the entries intersected by marked rows and unmarked columns and adds this minimum entry to each entry intersected by unmarked rows and marked columns. Finally, back to Step 3.

Thereafter, we can obtain the optimal solution $\mathcal{A}^*$ of Problem 1 by minimizing total region distances.

7.3 Utility-aware Optimization

In order to both high utility and efficiency, we formulate Problem 2 as a utility-aware optimization problem to optimize the assignment $\mathcal{A}^*$ of minimum ATD in Problem 1. We have introduced the constraint that the increase rate of average region distance should be less than η^τ when increasing ASR. Thereby, assume C denotes total region distance with $\mathcal{A}^*$, Problem 2 aims at maximizing ASR subjected to $(C^+ - C)/C \le \eta^\tau$, where C^+ denotes the total region distances after the optimization of $\mathcal{A}^*$.

This is not a standardized mathematical problem, and not easy to obtain an optimal solution. Therefore, we propose the Algorithm 3 called ASR-aware Optimization (ASROpt) to resolve this complicated problem. The main idea of ASROpt is to exchange tasks between failed workers (i.e., workers unable to complete tasks on time.) and successful assigned workers while considering the rate of increase of the average region distance. More specifically, Assume that there are several workers and task requesters, as shown in Fig. 3. Based on Problem 1, the SC server obtains the assignment $\mathcal{A}^*$=$\{\{l^t_1, l^w_3, 3.1\}, \{l^t_2, l^w_2, 2.4\}, \{l^t_3, l^w_1, 1.3\}, \{l^t_4, l^w_5, 8.2\}, \{l^t_5, l^w_4, 0.8\}\}$ with the minimum total region distance equal to 15.8km. However, with $d^w_\tau = 8.0$km and $\mathcal{I}_{2,5} = 8.2$km, the worker l^w_5 can not complete the task l^t_4 on time for $\mathcal{I}_{2,5} > d^w_\tau$. Hence, l^w_5 is a failed worker, and the ASR=80%. To improve ASR, we want to exchange the task of the failed worker with other successful assigned workers, where we observe that there is just one worker l^w_3 that can exchange its task l^t_1 with failed worker l^w_4 and the region distance $\mathcal{I}_{4,3} = 6.0$km is within 8.0km. Thereafter, the assignment $\mathcal{A}^*$ is optimized as $\mathcal{A}^*$=$\{\{l^t_1, l^w_5, 6.2\}, \{l^t_2, l^w_2, 2.4\}, \{l^t_3, l^w_1, 1.3\}, \{l^t_4, l^w_3, 6.0\}, \{l^t_5, l^w_4, 0.8\}\}$ with total region distance equal to 16.7km (the increasing rate η^*=5.7%), while ASR has been increased to 100%.

There is only one failed worker in Fig. 3. However, we constantly suffer from situations where multiple workers fail on their tasks on time in practice. Therefore, we need to consider several issues. 1) We need to set reasonable metrics to select successful assigned workers for task exchange. 2) We need to formulate task exchange into a mathematical

Algorithm 3: ASR-aware Optimization (ASROpt).

Input : $\mathcal{A}^*, \mathcal{I}, d^w_\tau, \eta$
Output: $\mathcal{A}^*_+$

1 $\mathcal{L}^{w,t}_f, \mathcal{L}^{w,t}_s = \emptyset, \emptyset$;
2 $\mathcal{A}^*_t$ = sort($\mathcal{A}^*, l^t$) ;
3 $\mathcal{A}^*_w$ = sort($\mathcal{A}^*, l^w$) ;
4 **for** $i = 1$ *to* M **do**
5 **if** $\mathcal{I}_{\mathcal{A}^*_w(l^w_i),i} > d^w_{\tau,i}$ **then**
6 Add l^w_i into $\mathcal{L}^{w,t}_f$;
7 **for** $j = 1$ *to* N **do**
8 **if** $\mathcal{I}_{j,i} < d^w_{\tau,i}, \mathcal{I}_{\mathcal{A}^*_w(l^w_i),\mathcal{A}^*_t(l^t_j)} < d^w_{\tau,i}$ **then**
9 Add $\mathcal{A}^*_t(l^t_j)$ into $\mathcal{L}^{w,t}_s$;
10 **end**
11 **end**
12 **end**
13 **end**
14 $\mathcal{S}$ is a $|\mathcal{L}^{w,t}_s| \times |\mathcal{L}^{w,t}_f|$ empty matrix ;
15 **for** $j = 1$ *to* $|\mathcal{L}^{w,t}_f|$ **do**
16 **for** $i = 1$ *to* $|\mathcal{L}^{w,t}_s|$ **do**
17 **if** $\{l^w_{f,j}, l^w_{s,i}\}$ *satisfies Eq. (15)* **then**
18 $\Delta\mathcal{I}^f_{j,i} = \mathcal{I}_{\mathcal{A}^*_w(l^w_{s,i}),l^w_{f,j}} - \mathcal{I}_{\mathcal{A}^*_w(l^w_{f,j}),l^w_{f,j}}$;
19 $\Delta\mathcal{I}^s_{j,i} = \mathcal{I}_{\mathcal{A}^*_w(l^w_{f,j}),l^w_{s,i}} - \mathcal{I}_{\mathcal{A}^*_w(l^w_{s,i}),l^w_{s,i}}$;
20 $\mathcal{S}_{l^w_{s,j},l^w_{f,i}} = \Delta\mathcal{I}^f_{j,i} + \Delta\mathcal{I}^s_{j,i}$;
21 **end**
22 **else**
23 $\mathcal{S}_{l^w_{s,j},l^w_{f,i}} = 0$;
24 **end**
25 **end**
26 **end**
27 # max $\sum_{i=1}^{|\mathcal{L}^{w,t}_s|}\sum_{j=1}^{|\mathcal{L}^{w,t}_f|}\mathcal{S}_{i,j}\mathcal{U}_{i,j}$;
28 $\mathcal{U}^*$ = Hungarian($\sum_{i=1}^{|\mathcal{L}^{w,t}_s|}\sum_{j=1}^{|\mathcal{L}^{w,t}_f|}\mathcal{S}_{i,j}\mathcal{U}_{i,j}$) ;
29 Obtain cost C^+ based on $\mathcal{U}^*$ and $\mathcal{A}^*$;
30 $\mathcal{U}^*_c$ = Sort($\mathcal{U}^*, \mathcal{S}(\mathcal{U}^*)$) ;
31 **for** $i = 1$ *to* $|\mathcal{U}^*_c|$ **do**
32 **if** $(C^+ - C)/C > \eta^\tau$ **then**
33 Delete $\mathcal{U}^*_{c,i}$ from $\mathcal{U}^*_c$;
34 **end**
35 **else**
36 Break ;
37 **end**
38 Obtain cost C^+ based on $\mathcal{U}^*_c$ and $\mathcal{A}^*$;
39 **end**
40 $\mathcal{A}^*_+$ = Adjust($A^*\mathcal{U}^*_c$) ;
41 Return $\mathcal{A}^*_+$.

expression for optimization. 3) We should adjust task exchange to keep the increase rate of total region distance within the threshold η^τ. Hereby, we design the ASR-aware Optimization algorithm (as shown in Algorithm 3, comprising three processes to resolve these three issues: *Initialization* (issue 1 resolved in Eq. (15)), *Problem Formulation* (issue 2 resolved in Eq. (20)) and *Exchange Adjustment* (issue 3).

Initialization (Lines 1-26): Firstly, $\mathcal{L}^{w,t}_f$ and $\mathcal{L}^{w,t}_s$ is initialized to two empty lists. Then, we set $\mathcal{A}^*_t$ and $\mathcal{A}^*_w$ as task-order assignment and worker-order assignment. For

tasks exchange

	l^w_1	l^w_2	l^w_3	l^w_4	l^w_5
l^t_1	8.1	∞	3.1	∞	6.2
l^t_2	∞	2.4	∞	4.5	10.4
l^t_3	1.3	∞	∞	10.2	∞
l^t_4	∞	5.7	6.0	∞	8.2
l^t_5	5.8	∞	∞	0.8	∞

	l^w_1	l^w_2	l^w_3	l^w_4	l^w_5
l^t_1	8.1	∞	~~3.1~~	∞	6.2
l^t_2	∞	2.4	∞	4.5	10.4
l^t_3	1.3	∞	∞	10.2	∞
l^t_4	∞	5.7	6.0	∞	~~8.2~~
l^t_5	5.8	∞	∞	0.8	∞

Fig. 3. Task exchange on the optimal solution of Problem 1.

example, $\mathcal{A}^* = \{\{l^t_1, l^w_3\}, \{l^t_3, l^w_2\}, \{l^t_2, l^w_1\}\}$ contributes to $\mathcal{A}^*_t = \{\{l^t_1 : l^w_3\}, \{l^t_2 : l^w_1\}, \{l^t_3 : l^w_2\}\}$ and $\mathcal{A}^*_w = \{\{l^w_1 : l^t_2\}, \{l^w_2 : l^t_3\}, \{l^w_3 : l^t_1\}\}$. Thereafter, based on workers' acceptable distance d^w_τ, we select failed worker-task pairs $v_f\{l^w_f, l^t_f\}$ and then filter successful assigned worker-task pairs $v_s\{l^w_s, l^t_s\}$ on the basis of optimal assignment $\mathcal{A}^*$, where the selection of $v_s\{l^w_s, l^t_s\}$ needs to subject to two constrains:

$$\mathcal{I}_{l^t_{s,j},l^w_{f,i}} \le d^w_{\tau,l^w_{f,i}} \ , \ \mathcal{I}_{l^t_{f,j},l^w_{s,i}} \le d^w_{\tau,l^w_{s,i}} \ . \tag{15}$$

These constraints mean that $l^w_{f,i}$ and $l^w_{s,j}$ can successfully carry out each other's tasks. We store $v_f\{l^w_f, l^t_f\}$ and $v_s\{l^w_s, l^t_s\}$ in $\mathcal{L}^{w,t}_f$ and $\mathcal{L}^{w,t}_s$, respectively. Finally, we set a $|\mathcal{L}^{w,t}_f| \times |\mathcal{L}^{w,t}_s|$ matrix $\mathcal{S}$ to store the degree of difference by exchanging tasks of each l^w_f and each l^w_s. We define $\mathcal{S}_{i,j}$ as follows:

Definition 8 (Region Distance Change $\mathcal{S}_{i,j}$). Assume $l^w_{f,j}$ and $l^w_{s,i}$ exchange their task $\mathcal{A}^*_w(l^w_{f,j})$, $\mathcal{A}^*_w(l^w_{s,i})$, we define $\mathcal{S}_{i,j}$ as the total change of their region distances of $l^w_{f,j}$ and $l^w_{s,i}$.

$$\mathcal{S}_{i,j} = \begin{cases} \Delta\mathcal{I}^f_{i,j} + \Delta\mathcal{I}^s_{i,j}, \text{ if satisfy Eq. (15)} , \\ 0, \text{ otherwise} , \end{cases} \tag{16}$$

$$\Delta\mathcal{I}^f_{i,j} = \mathcal{I}_{\mathcal{A}^*_w(l^w_{f,i}),l^w_{s,j}} - \mathcal{I}_{\mathcal{A}^*_w(l^w_{f,i}),l^w_{f,i}} \ , \tag{17}$$

$$\Delta\mathcal{I}^s_{i,j} = \mathcal{I}_{\mathcal{A}^*_w(l^w_{f,i}),l^w_{s,j}} - \mathcal{I}_{\mathcal{A}^*_w(l^w_{s,j}),l^w_{s,j}} \ , \tag{18}$$

$$1 \le i \le |\mathcal{L}^{w,t}_f|, \ 1 \le j \le |\mathcal{L}^{w,t}_s| \ , \tag{19}$$

where $\Delta\mathcal{I}^f_{i,j}$ and $\Delta\mathcal{I}^s_{i,j}$ represent the difference of $l^w_{f,i}$ and $l^w_{s,j}$'s region distances after the exchange, respectively.

Problem formulation (Lines 27,28): After initialization, the algorithm obtains the region distance change matrix $\mathcal{S}$. We need to hold the principle to increase ASR by absorbing a certain degree of deficiency on the total region distance C. Therefore, we formalize this conception as a maximization problem to increase ASR as follows:

$$\begin{aligned} \max_{\mathcal{U}} \quad & \textstyle\sum_{i=1}^{|\mathcal{L}^{w,t}_s|}\sum_{j=1}^{|\mathcal{L}^{w,t}_f|}\mathcal{S}_{i,j}\mathcal{U}_{i,j} \\ \text{subject to} \quad & \mathcal{U}_{i,j} = 0 \text{ or } 1 \ , \\ & \textstyle\sum_{j=1}^{|\mathcal{L}^{w,t}_f|}\mathcal{U}_{i,j} = 1 \ , \\ & \textstyle\sum_{i=1}^{|\mathcal{L}^{w,t}_s|}\mathcal{U}_{i,j} \le 1 \ . \end{aligned}$$

In order to solve it, we reformulate this problem as follows:

$$\min_{\mathcal{U}} \quad \sum_{i=1}^{|\mathcal{L}_s^{w,t}|}\sum_{j=1}^{|\mathcal{L}_f^{w,t}|}(max(\mathcal{S}) - \mathcal{S}_{i,j})\mathcal{U}_{i,j} \ . \tag{20}$$

In addition, when the successful assigned workers is fewer than the failed workers, we can simply adjust the constrains to $\sum_{j=1}^{|\mathcal{L}_f^{w,t}|}\mathcal{U}_{i,j} \leq 1$ and $\sum_{i=1}^{|\mathcal{L}_s^{w,t}|}\mathcal{U}_{i,j} = 1$. Then, with the help of Hungarian algorithm, we obtain the optimal exchange $\mathcal{U}^*$.

Exchange adjustment (Lines 29-41): We have obtained the optimal exchange $\mathcal{U}^*$ so that ASR is closest to 100%. However, the total region distance C^+ of the latest assignment, generated based on $\mathcal{U}^*$ and $\mathcal{A}^*$, may exceed the threshold η^τ, where the SC system accepts the increase on C satisfying $(C^+ - C)/C \leq \eta^\tau$. The algorithm firstly sorts $\mathcal{U}^*$ from largest to smallest based on region distance changes. Then, the algorithm loops through the optimal exchange $\mathcal{U}_c^*$ and determines whether the latest region distance C^+ is out of range (i.e., $(C^+ - C)/C > \eta^\tau$). If so, the exchange with the largest change will be deleted from $\mathcal{U}_c^*$ and generates the latest region distance C^+. Otherwise it jumps out of the loop and gets the final task exchange $\mathcal{U}_c^*$. Thereafter, based on $\mathcal{U}_c^*$, we adjust $\mathcal{A}^*$ and obtain the final assignment $\mathcal{A}_+^*$.

8 PERFORMANCE EVALUATION

In this section, we conduct extensive experiments on a real-world dataset to evaluate the performance of our framework. We first provide details of our experiment setup consisting of the dataset, baselines, and metrics. Then, the performance of our framework is analyzed regarding several critical parameters.

8.1 Experimental Setup

These simulations are implemented on a taxi trajectory dataset in Roma [24] in Python 3.8 platform and performed on macOS with an 8-core Apple M1 CPU, 8GB memory.

8.1.1 Dataset

The dataset is collected in Roma, containing 21817851 GPS records of 316 taxis collected over 30 days from February 1st, 2014, to March 2nd, 2014. In our experiments, we extract most records (approximate 95%) from the dataset and divide the coverage area of these records into 8×8 regions as shown in Fig. 4. Then, we count the record proportion of each region and select region A (59.8%), region B (1.5%), region C (4.7%), and region D (5.1%) to show the performance downtown (i.e., the worker-dense region A) and in suburbs (i.e., worker-sparse regions B, C, and D).

8.1.2 Baselines

We introduce three representative differential privacy-based frameworks as baselines:

- *GO Function (CG) [17].* CG only focuses on the location privacy of workers and considers the road network in SC by involving a Linear programming problem to maximize expected estimation error subject to minimum ATD and geo-indistinguishability [25].

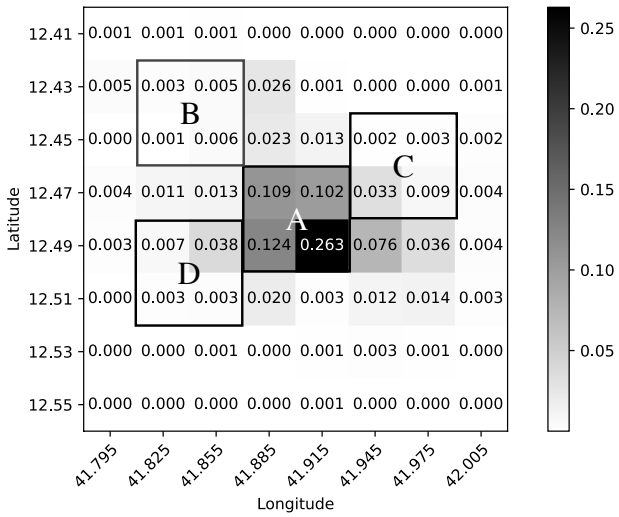


Fig. 4. Heat map of GPS records over Roma.

- *Framework on 2D (BD) [18].* BD protects workers' location privacy by perturbing locations on two dimensions in SC.
- *Laplacian Mechanism and Multi-task Assignment.* Laplacian mechanism is a de facto classic privacy-preserving mechanism for location privacy on two dimensions. To compare with our framework, we design this framework as the combination of Laplacian mechanism and our procedure of the multi-task assignment demonstrated in Section. 7.2.

8.1.3 Metrics

We need to evaluate three aspects: location privacy, efficiency and utility of multi-task assignment.

- *Location Privacy.* Expected Estimation Error (E3) represents the location privacy of task requesters and workers in SC, defined as the distance between real locations and locations inferred by adversaries in Section. 3.3.1. The higher the E3, the better the location privacy.
- *Efficiency of Multi-task Assignment.* Average Travel Distance (ATD) represents the efficiency of multi-task assignments, defined as the average value of multiple distances from assigned workers to corresponding task requesters in Section. 3.3.2. The lower the ATD, the better efficiency.
- *Utility of Multi-task Assignment.* Assignment Success Rate (ASR) represents the utility of multi-task assignments, defined as the percentage of success assignments in all assignments in Section. 3.3.3. The higher the ASR, the better utility.

8.1.4 Parameter Settings

In experiments, we need to consider the impact of several parameters: privacy budget ϵ and the number of tasks N. To evaluate location privacy, we set reasonable privacy budget $\epsilon \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3\}$. To evaluate the efficiency and utility of multi-task assignment, we set privacy budget ϵ and the number of tasks $N \in \{20, 30, 40, 50, 60\}$ (There are

TABLE 1
The Rate of Disclosure of Perturbing Behaviors.

	Region A	Region B	Region C	Region D
RoPriv	0.00%	0.00%	0.00%	0.00%
CG	0.00%	0.00%	0.00%	0.00%
BD	1.90%	71.4%	11.8%	30.9%
Laplacian	2.10%	68.3%	12.1%	28.6%

only approximately 80 taxis work on average in each day). We set the default value of distance threshold $d_r^\tau = 500m$, explained in Section. 8.2.2. In addition, we set $\eta^\tau = 5\%$ as the increase threshold of ATD, which is a marginal increase.

8.2 Experimental Results

In our experiments, we simulate the taxi dispatch system based on this dataset. The system starts at 8:00 and ends at 20:00 every day, in which a multi-task assignment is implemented every half hour. Thus, there are 25 multi-task assignments each day and 750 times totally. We use the average value of participants' ATDs and ASRs in each multi-task assignment to represent the performance of this multi-task assignment. Furthermore, we employ the average value of all participants' E3s in 30 days to show the performance in location privacy protection.

8.2.1 Effect of Privacy Protection

Location Privacy. As shown in Fig. 5, to evaluate location privacy, we set four experiments conducted on regions A, B, C, and D corresponding to a record-dense region (i.e., downtown) and three record-sparse regions (i.e., suburb). In addition, we set the range of $\mathcal{W}$ of our framework as $d_r^\tau = 500$m (called RoPriv-500) and $d_r^\tau = 1500$m (called RoPriv-1500), which represent the general location privacy protection and the strong location privacy protection, respectively. Considering that CG and BD can not provide privacy protection for task requesters, we adopt workers' location privacy of CG and BD.

Figs. 5(a), 5(b), 5(c), and 5(d) show the impact of privacy budget ϵ on location privacy E3 in region A, B, C, and D, respectively. In these experimental results, we can first observe that with the increase of privacy budget ϵ, the location privacy E3 decreases in all the frameworks. Therefore, if a participant needs strong location privacy protection, setting a small privacy budget is better. As shown in Fig. 5, under different ϵ, RoPriv-1500 performs best than any other frameworks in all the regions, where all the E3 of RoPriv-1500 exceed 800m. It means that the location, inferred by the adversary with side information, is more than 800m away from the real location of the participant with the help of RoPriv-1500. Moreover, even though with the set $d_r^\tau = 500$m, E3 achieved by RoPriv-500 is still more than 300m, which is able to guarantee the location privacy for all participants. Hence, our framework can provide sufficient location privacy protection for task requesters and workers in SC with considering road networks.

By comparing the performance in record-dense region and record-sparse region, we find that our privacy-preserving frameworks (i.e., RoPriv-500 and RoPriv-1500) provide stable location privacy protection regardless of the density of records. For example, RoPriv-1500 achieves E3 in the range of 800m-900m, whether downtown or in suburbs. However, CG cannot provide stable and proper location privacy protection considering the road network. E3 achieved by CG is less than 100m downtown in Fig. 5(a), which can not guarantee the location privacy of workers. In the remote suburb (i.e., region B), E3 achieved by CG even exceeds 800m in Fig. 5(b). The unstable performance of CG results from the density of workers, which heavily influences the location privacy protection of CG. Our framework just consider settings (i.e., d_r^τ and δ) of participants without involving any external factors. Thus, location privacy can be stably guaranteed with the help of our framework.

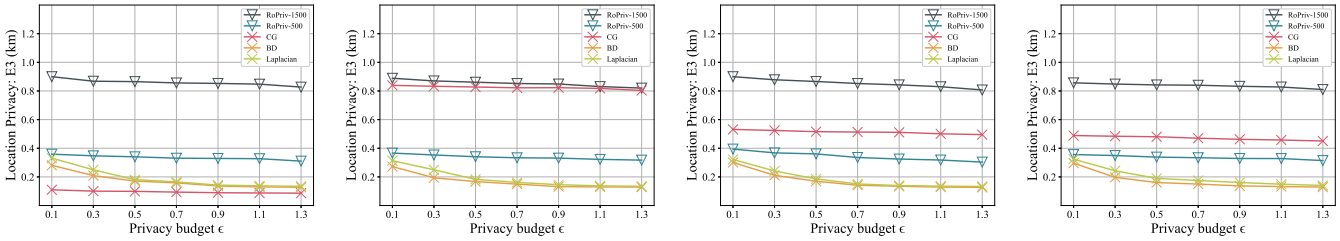
The Extensive Privacy. In addition to protect the location privacy of participants, the behavior of perturbing locations is needed to be preserved for protecting the extensive privacy of task requesters and workers. By investigating, we know that GPS-enabled mobile devices are typically accurate to within 4.9m under open sky [39]. Hereby, we assume that a participant will be considered perturbing its location if its location obtained by the SC server is 20m away from road networks or lies in a river or forest. We conduct several experiments on different regions to evaluate the effect of preserving the perturbing behavior, and use the rate of disclosure of perturbing behaviors to show the effectiveness of our framework, as shown in Tab. 1. It is obvious that our framework RoPriv well preserves the behavior of perturbing locations and moreover protects the extensive privacy. Even though in the remote suburb, the rate of disclosure of perturbing behaviors is still 0.0%. However, based on these frameworks without considering the road network (i.e., BD and Laplacian), participants suffer from potential extensive privacy disclosure, especially in the remote suburb with nearly 70% probability of being recognized. The experimental results verify the effect of our road network-aware framework for preserving the perturbing behaviors.

Therefore, our framework can provide sufficient and stable location privacy protection for task requesters and workers, whether downtown or in suburbs. Furthermore, our framework effectively preserves the behavior of perturbing locations for each participant.

8.2.2 Efficiency of Multi-task Assignment

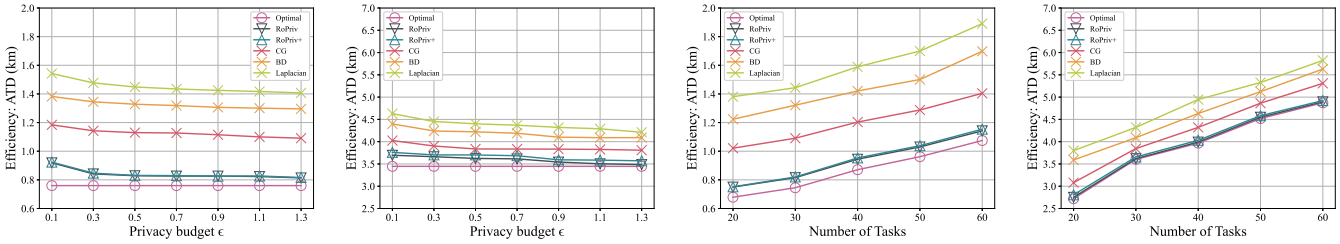
To evaluate the efficiency of multi-task assignment, we set the optimal assignment (called Optimal) by implementing multi-task assignment on real locations. We conduct our experiments downtown (i.e., region A) and in the remote suburb (i.e., region B). Considering the impact of the privacy budget ϵ and the number of tasks N, we obtain the experimental result as shown in Fig. 6, where RoPriv and RoPriv+ denote our framework without considering utility (i.e., ASR) and our framework after improving utility, respectively.

In Fig. 6, it is obvious that ATD follows Optimal $<$ RoPriv $\approx$ RoPriv+ $<$ CG $<$ BD $<$ Laplacian, where RoPriv performs best than baselines no matter downtown (ATD $<$ 1.0km) or in the remote suburb (ATD $<$ 3.75km). Moreover, we can observe that the ATD achieved by these road network-aware frameworks (i.e., RoPriv, RoPriv+, and CG) is lower than the ATD achieved by BD and Laplacian,



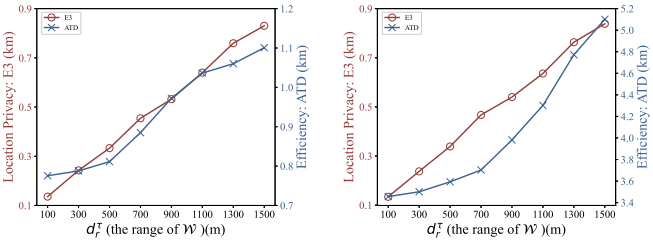
(a) In region A (downtown). (b) In region B (remote suburb). (c) In region C (suburb). (d) In region D (suburb).

Fig. 5. The impact of ϵ on location privacy downtown and in suburbs.



(a) Downtown. (b) In the remote suburb. (c) Downtown. (d) In the remote suburb.

Fig. 6. a) and b) The impact of ϵ on efficiency of multi-task assignment with $d_r^\tau = 500$m and $N = 30$. c) and d) The impact of the number of tasks N on efficiency of multi-task assignment with $d_r^\tau = 500$m and $\epsilon = 0.9$.



(a) Downtown. (b) In the remote suburb.

Fig. 7. The impact of d_r^τ on location privacy and multi-task assignment with $\epsilon = 0.9$, $N = 30$.

which proves the increase in efficiency by considering the road network. Especially, the ATD achieved by RoPriv is closed to the optimal ATD, and the difference between them is no more than 100m in most cases. Even though efficiency has been weakened for improving utility in RoPriv+, the performance of RoPriv+ is still extremely closed to RoPriv, indicating that the increase rate of ATD is constrained well by $\eta = 5\%$ in Utility-aware Optimization.

In addition, we find that the ATD in Figs. 6(b), 6(d) is typically higher than the ATD in Figs. 6(a), 6(c). That is because most taxis are concentrated downtown (i.e., region A), and most workers downtown have to travel a long distance for completing tasks in the remote suburb (i.e., region B).

Impact of privacy budget ϵ. Figs. 6(a) and 6(b) present the impact of privacy budget ϵ on the efficiency of multi-task assignment, where as ϵ increases, the ATD generally decreases. In addition, the ATD in $\epsilon = 0.1 \sim 0.7$ decreases more than the ATD in $\epsilon > 0.9$. We have known that the lower ATD represents the high efficiency of multi-task assignment. Thus, we set $\epsilon = 0.9$ as default value in our experiments.

Impact of the number of tasks N. Figs. 6(c) and 6(d) present the impact of the number of tasks N on the efficiency of multi-task assignment, where as N increases, the ATD significantly increases. That indicates N has a large negative impact on the efficiency of multi-task assignment. Comparing to the ATD in $N = 30 \sim 60$, the ATD increases more slowly in $N = 20 \sim 30$. Furthermore, with RoPriv and $N = 30$, ATD $\approx$ 800m downtown and ATD $\approx$ 3600m in the remote suburb are acceptable distances for workers. For that, we choose $N = 30$ as default value in our experiments.

Impact of the range of $\mathcal{W}$ d_r^τ. We set $d_r^\tau = \{100\text{m}, 300\text{m}, 500\text{m}, 700\text{m}, 900\text{m}, 1100\text{m}\ 1300m, 1500\text{m}\}$ and conduct multiple experiments. As shown in Fig. 7, we consider d_r^τ as horizontal axis, set location privacy (E3) and efficiency (ATD) as vertical axes to demonstrate the impact of d_r^τ on location privacy and multi-tasks assignment downtown (cf. Fig. 7(a)) and in the remote suburb (cf. Fig. 7(b)). Intuitively, we find that as d_r^τ raising, E3 increases and ATD increases, which indicates the improvement on location privacy and reduction on efficiency. Hence, to balance the location privacy and efficiency, we need to determine a value of d_r^τ as the default value in our experiments. We can observe that with $d_r^\tau > 500$m, ATD increases dramatically downtown and in the remote suburb. Moreover, when $d_r^\tau = 500$m, RoPriv provides sufficient location privacy exceeding than all the baselines downtown (cf. Fig. 5(a)). Therefore, we consider the default value of d_r^τ as 500m.

8.2.3 *Utility of Multi-task Assignment*

To evaluate the performance on improving utility of multi-task assignment, we conduct experiments downtown (cf. Fig. 8) and in the remote suburb (cf. Fig. 9), in which we

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2023.3255232

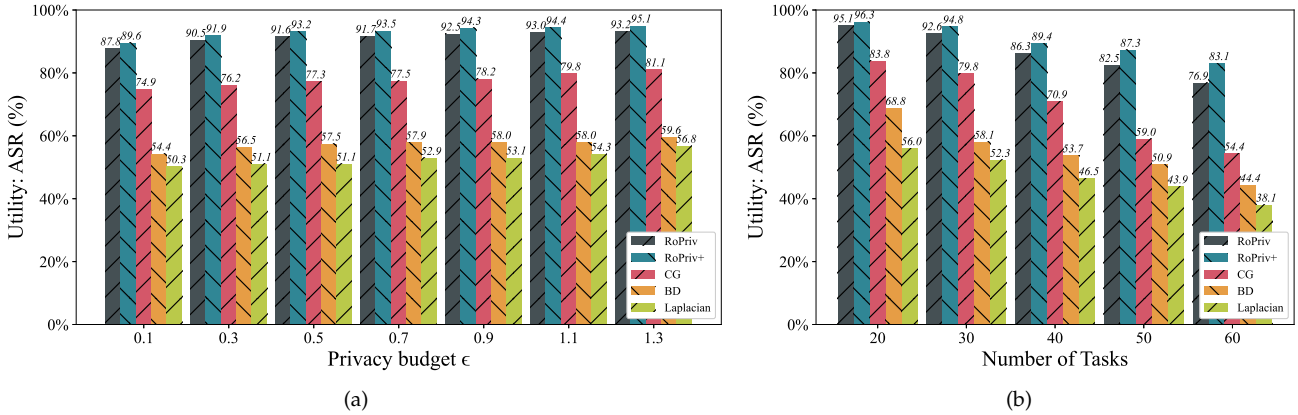


Fig. 8. a) The impact of ϵ on utility of multi-task assignment *downtown* with $d_r^\tau = 500$m and $N = 30$. b) The impact of the number of tasks N on utility of multi-task assignment *downtown* with with $d_r^\tau = 500$m and $\epsilon = 0.9$.

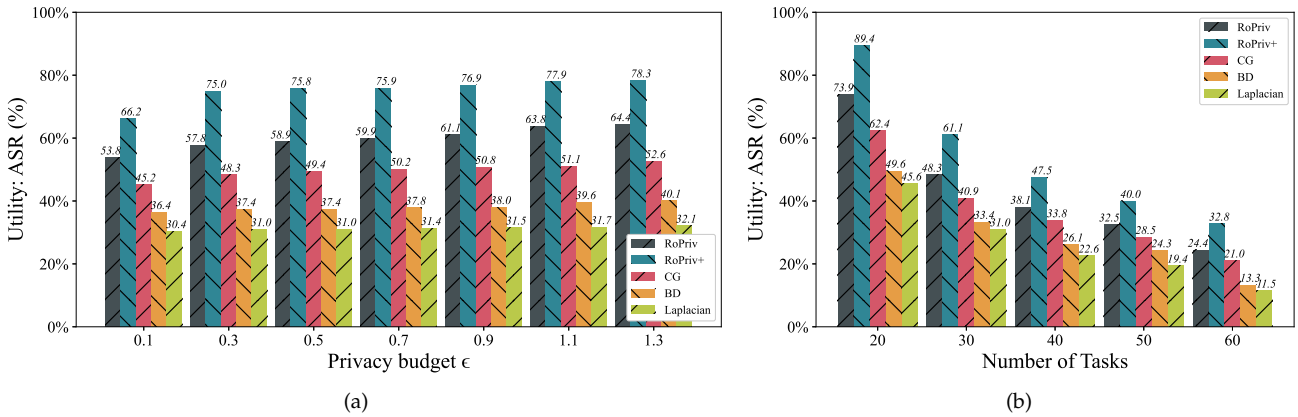


Fig. 9. a) The impact of ϵ on utility of multi-task assignment *in the remote suburb* with $d_r^\tau = 500$m and $N = 30$. b) The impact of the number of tasks N on utility of multi-task assignment *in the remote suburb* with with $d_r^\tau = 500$m and $\epsilon = 0.9$.

set $d_\tau^w = 800$m downtown and $d_\tau^w = 3600$m in the remote suburb.

As shown in Figs. 8 and 9, RoPriv+ achieves the highest ASR whether downtown or in the remote suburb. More specifically, the ASR achieved by RoPriv+ is in the range of 83.1%-96.3% downtown (cf. Fig. 8), in which the maximum ASR is closed to 100%. In the remote suburb (cf. Fig. 9), the range of ASR of RoPriv+ is 32.8%-78.3%. Even though there are 60 tasks in the remote suburb waiting for assignment, RoPriv+ still guarantees 32.8% task completed. Comparing to baselines, the ASR achieved by RoPriv+ is able to exceed baselines by up to 26.7% (in Fig. 9(a), $\epsilon = 0.3$). Hence, RoPriv+ performs well on the utility of multi-task assignment.

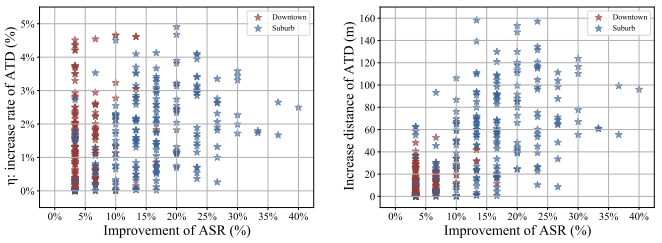
Comparing to the performance on utility between RoPriv+ and RoPriv, we can find that the ASR has been significantly improved after Utility-aware Optimization. Based on RoPriv+, the improvement rate of ASR ranges from 1.2% (in Fig. 8(b), $N = 20$) to 17.2% (in Fig. 9(a), $\epsilon = 0.3$). This indicates that our framework enables 17.2% workers to complete tasks on time after Utility-aware Optimization in the best case. Furthermore, we have demonstrated that there are no obvious increase in ATD of our framework after Utility-aware Optimization in Section. 8.2.2. Thus, our Utility-aware Optimization indeed improves the utility of multi-task assignment while keepging the efficiency tolerable.

In addition, we notice that the ASR downtown is improved better than the ASR in the remote suburb. Significantly, the maximum improvement rate of ASR is 6.2% downtown, and the minimum improvement rate of ASR is 7.5% in the remote suburb. That is because the ASR is a high degree of RoPriv downtown.

8.2.4 Negative Impact on Efficiency from Improvement in Utility

We have known that the improvement of ASR will inevitably reduce the efficiency of the multi-task assignment and our goal of the multi-task assignment is to obtain both high utility and efficiency. Therefore, we analyze representative experiments composed of experiments downtown and in the remote suburb ($\epsilon = 0.9$, $N = 30$).

As shown in Fig. 10(a), we take the improvement ΔASR of utility as the horizontal axis and take the increase rate of ATD as the vertical axis to show the relationship between ΔASR and η. Moreover, we show the relationship between ΔASR and the increase distance of ATD in Fig. 10(b). In Fig. 10(a), we can observe that the increase rate of ATD is well restricted by the increase threshold $\eta^\tau = 5\%$. Even



(a) Downtown. (b) In the remote suburb.

Fig. 10. Analysis on experimental results of ATD and ASR with $\epsilon = 0.9$, $N = 30$.

though there are 40% workers who are enabled to complete tasks by our Utility-aware Optimization, the increase rate of ATD is less than 3%, which is a minor increase. Furthermore, we can see plenty of assignments improved with $\eta \approx 0\%$, which indicates that our Utility-aware Optimization significantly improves the utility while just raising little influence on the efficiency of multi-task assignment. That is the trade-off between utility and efficiency.

Comparing the experiment results downtown and in the remote suburb, we find that most ΔASRs range from 3% to 10% downtown, and ΔASRs in the remote suburb are mainly distributed between 10% and 30%, which indicates that our Utility-aware Optimization performs well in the remote suburb. In Fig. 10(b), the experimental results downtown present that the improvement of ASR downtown cause the distance increase of ATD less than 40m in most cases. In the remote suburb, the distance increase of ATD is less than 160m, which is not a large value compared to $d_\tau^w = 3600$m.

In summary, our framework can provide stable and sufficient location privacy protection (cf. Fig. 5) for both task requesters and workers regardless of regions (e.g., downtown, in the remote suburb), meanwhile, protect their perturbing behaviors from disclosure (cf. Tab. 1). In multi-task assignment, our framework can obtain both high efficiency (i.e., low ATD) and utility (i.e., high ASR) by maximizing utility on the basis of ensuring a minor increase in efficiency (cf. Fig. 10).

9 Conclusion

In this paper, we proposed a road network-aware privacy-preserving framework to implement a multi-task assignment with both high utility and efficiency while protecting the location privacy of both task requesters and workers on road networks in Spatial Crowdsourcing. We firstly abstracted the road network into discrete locations and proposed an Obfuscated Locations Selection algorithm to generate possible obfuscated locations based on participants' real locations without disclosing their location privacy. Then, we designed a Road Network-aware Exponential Mechanism to perturb locations of task requesters and workers on the road network, in which the behavior of perturbing locations had been preserved. Based on obfuscated locations, we proposed region distance to replace the distance among obfuscated locations to implement multi-task assignments. Thereafter, with the basis of region distance, we decomposed multi-task assignment into a Binary Linear Programming problem and a Utility-aware Optimization problem to both high utility and efficiency. Our experimental results on real trajectory dataset indicated that our framework could provide sufficient and stable location privacy protection for both task requesters and workers downtown and in the remote suburb. Furthermore, our framework obtains both high utility and efficiency in multi-task assignments.

References

[1] C. Qiu, A. C. Squicciarini, C. Pang, N. Wang, and B. Wu, "Location privacy protection in vehicle-based spatial crowdsourcing via geo-indistinguishability," *IEEE Transactions on Mobile Computing*, 2020.

[2] H. To, G. Ghinita, L. Fan, and C. Shahabi, "Differentially private location protection for worker datasets in spatial crowdsourcing," *IEEE Transactions on Mobile Computing*, vol. 16, no. 4, pp. 934–949, 2016.

[3] B. Zhao, S. Tang, X. Liu, X. Zhang, and W.-N. Chen, "itam: Bilateral privacy-preserving task assignment for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 20, no. 12, pp. 3351–3366, 2020.

[4] H. Jiang, M. Wang, P. Zhao, Z. Xiao, and S. Dustdar, "A utility-aware general framework with quantifiable privacy preservation for destination prediction in lbss," *IEEE/ACM Transactions on Networking*, vol. 29, no. 5, pp. 2228–2241, 2021.

[5] Z. Wang, X. Pang, Y. Chen, H. Shao, Q. Wang, L. Wu, H. Chen, and H. Qi, "Privacy-preserving crowdsourced statistical data publishing with an untrusted server," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1356–1367, 2018.

[6] F. T. Islam, T. Hashem, and R. Shahriyar, "A privacy-enhanced and personalized safe route planner with crowdsourced data and computation," in *Proceedings of IEEE ICDE*, 2021, pp. 229–240.

[7] Y. Tong, L. Chen, and C. Shahabi, "Spatial crowdsourcing: Challenges, techniques, and applications," *Proceedings of the VLDB Endowment*, vol. 10, no. 12, pp. 1988–1991, 2017.

[8] W. Ni, P. Cheng, L. Chen, and X. Lin, "Task allocation in dependency-aware spatial crowdsourcing," in *Proceedings of IEEE ICDE*, 2020, pp. 985–996.

[9] Y. Zhao, K. Zheng, Y. Cui, H. Su, F. Zhu, and X. Zhou, "Predictive task assignment in spatial crowdsourcing: a data-driven approach," in *Proceedings of IEEE ICDE*, 2020, pp. 13–24.

[10] Y. Cheng, B. Li, X. Zhou, Y. Yuan, G. Wang, and L. Chen, "Real-time cross online matching in spatial crowdsourcing," in *Proceedings of IEEE ICDE*, 2020, pp. 1–12.

[11] P. Zhao, H. Jiang, J. Li, F. Zeng, X. Zhu, K. Xie, and G. Zhang, "Synthesizing privacy preserving traces: Enhancing plausibility with social networks," *IEEE/ACM Transactions on Networking*, vol. 27, no. 6, pp. 2391–2404, 2019.

[12] L. Wang, D. Yang, X. Han, T. Wang, D. Zhang, and X. Ma, "Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation," in *Proceedings of WWW*, 2017, pp. 627–636.

[13] Q. Tao, Y. Tong, Z. Zhou, Y. Shi, L. Chen, and K. Xu, "Differentially private online task assignment in spatial crowdsourcing: A tree-based approach," in *Proceedings of IEEE ICDE*, 2020, pp. 517–528.

[14] M. Li, J. Wang, L. Zheng, H. Wu, P. Cheng, L. Chen, and X. Lin, "Privacy-preserving batch-based task assignment in spatial crowdsourcing with untrusted server," in *Proceedings of the ACM Conference on Information and Knowledge Management*, 2021, pp. 947–956.

[15] J. Wei, Y. Lin, X. Yao, and J. Zhang, "Differential privacy-based location protection in spatial crowdsourcing," *IEEE Transactions on Services Computing*, vol. 15, no. 1, pp. 45–58, 2022.

[16] W. Huang, W. Tang, K. Zhang, H. Zhu, and Y. Zhang, "Thwarting unauthorized voice eavesdropping via touch sensing in mobile systems," in *Proceedings of IEEE INFOCOM*, 2022, pp. 31–40.

[17] C. Qiu, A. Squicciarini, Z. Li, C. Pang, and L. Yan, "Time-efficient geo-obfuscation to protect worker location privacy over road networks in spatial crowdsourcing," in *Proceedings of ACM Conference on Information and Knowledge Management*, 2020, pp. 1275–1284.

[18] L. Wang, D. Yang, X. Han, D. Zhang, and X. Ma, "Mobile crowdsourcing task allocation with differential-and-distortion geo-obfuscation," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 967–981, 2021.

[19] H. Li, Q. Song, G. Li, Q. Li, and R. Wang, "Gpsc: A grid-based privacy-reserving framework for online spatial crowdsourcing," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[20] F. Song, Z. Qin, D. Liu, J. Zhang, X. Lin, and X. Shen, "Privacy-preserving task matching with threshold similarity search via vehicular crowdsourcing," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 7161–7175, 2021.

[21] D. Yuan, Q. Li, G. Li, Q. Wang, and K. Ren, "Priradar: A privacy-preserving framework for spatial crowdsourcing," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 299–314, 2019.

[22] J. Shu, X. Jia, K. Yang, and H. Wang, "Privacy-preserving task recommendation services for crowdsourcing," *IEEE Transactions on Services Computing*, vol. 14, no. 1, pp. 235–247, 2018.

[23] X. Li, Y. Ren, L. T. Yang, N. Zhang, B. Luo, J. Weng, and X. Liu, "Perturbation-hidden: Enhancement of vehicular privacy for location-based services in internet of vehicles," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2073–2086, 2021.

[24] "Dataset of mobility traces of taxi cabs in rome, italy," https://crawdad.org/roma/taxi/20140717/, july, 2014.

[25] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of ACM CCS*, 2013, pp. 901–914.

[26] H. To, G. Ghinita, and C. Shahabi, "A framework for protecting worker location privacy in spatial crowdsourcing," *Proceedings of the VLDB Endowment*, vol. 7, no. 10, pp. 919–930, 2014.

[27] H. Ren, S. Ruan, Y. Li, J. Bao, C. Meng, R. Li, and Y. Zheng, "Mtrajrec: Map-constrained trajectory recovery via seq2seq multi-task learning," in *Proceedings of ACM SIGKDD*, 2021, pp. 1410–1419.

[28] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy." *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, 2014.

[29] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proceedings of IEEE FOCS*, 2007, pp. 94–103.

[30] A. Gadotti, F. Houssiau, L. Rocher, B. Livshits, and Y.-A. De Montjoye, "When the signal is in the noise: Exploiting diffix's sticky noise," in *USENIX Security Symposium*, 2019, pp. 1081–1098.

[31] N. Ashena, D. Dell'Aglio, and A. Bernstein, "Understanding ϵ for differential privacy in differencing attack scenarios," in *Conference on Security and Privacy in Communication Systems*, 2021, pp. 187–206.

[32] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proceedings of ACM CCS*, 2012, pp. 617–627.

[33] P. Zhao, H. Jiang, J. C. Lui, C. Wang, F. Zeng, F. Xiao, and Z. Li, "P 3-loc: A privacy-preserving paradigm-driven framework for indoor localization," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2856–2869, 2018.

[34] H. Jiang, P. Zhao, and C. Wang, "Roblop: Towards robust privacy preserving against location dependent attacks in continuous lbs queries," *IEEE/ACM Transactions on Networking*, vol. 26, no. 2, pp. 1018–1032, 2018.

[35] S. Takagi, Y. Cao, Y. Asano, and M. Yoshikawa, "Geo-graph-indistinguishability: Location privacy on road networks based on differential privacy," *arXiv:2010.13449*, 2020.

[36] T. Cunningham, G. Cormode, H. Ferhatosmanoglu, and D. Srivastava, "Real-world trajectory sharing with local differential privacy," in *Proceedings of the VLDB Endowment*, 2021, pp. 2283–2295.

[37] C. Ilvento, "Implementing the exponential mechanism with base-2 differential privacy," in *Proceedings of ACM CCS*, 2020, pp. 717–742.

[38] B. Weggenmann and F. Kerschbaum, "Differential privacy for directional data," in *Proceedings of ACM CCS*, 2021, pp. 1205–1222.

[39] "Official u.s. government information about the global positioning system (gps) and related topics," https://www.gps.gov/systems/gps/performance/accuracy/.