Geo-MOEA: A Multi-Objective Evolutionary Algorithm with Geo-obfuscation for Mobile Crowdsourcing Workers

S. Zhang, T. Zhang, Z. Chen, N. Xiong

Abstract—The rapid development of mobile Internet and sharing economy brings the prosperity of Spatial Crowdsourcing (SC). SC applications assign various tasks according to reported location information of task’s requesters and outsourced workers (such as DiDi, MeiTuan and Uber). However, SC-servers are often untrustworthy and the exposure of users’ locations raises privacy concerns. In this paper, we design a framework called Geo-MOEA (Multi-Objective Evolutionary Algorithm with Geo-obfuscation) to protect location privacy of workers involved on SC platform in mobile networks environment. We propose an adaptive regionalized obfuscation approach with inference error bounds based on geo-indistinguishability (a strong notion of differential privacy), which is suitable for the context of large-scale location data and task allocations. This enables each worker to report a pseudo-location that is adaptively generated with a personalized inference error threshold. Moreover, as a popular computational intelligence method, MOEA is introduced to optimize the trade-off between SC service availability and privacy protection while ensuring theoretically the most general condition on protection location sets for larger search space. Finally, the experimental results on two public datasets show that our Geo-MOEA approach achieves up to 20% reduction in service quality loss while guaranteeing differential and geo-distortion location privacy.

Index Terms—Differential privacy, multi-objective evolutionary optimization, geo-indistinguishability, inference error, spatial crowdsourcing.

I. INTRODUCTION

In recent years, due to the vigorous development of mobile networks (4G/5G) and dramatic proliferation of mobile devices, Spatial Crowdsourcing (SC) (a.k.a. mobile crowdsourcing) [1], [2] has emerged as a prevailing paradigm of crowdsourcing and has stimulated various applications such as taxi calling (e.g., Didi Chuxing and Uber), food delivery (e.g., MeiTuan) and social network platforms (e.g., Twitter and Microblogs) in the real world [1].

Compared with conventional crowdsourcing, SC based on location services has led to an exponential growth in data collection and sharing of intelligent terminals. For the scenario of large-scale SC, a larger number of tasks are requested and allocated at the SC-server side among a wider range of potential workers for providing faster and efficient services. However, the SC-server may be untrustworthy, and disclosing individual locations has serious privacy implications [3], [4], [5]. In recent years, many techniques appear to protect workers’ location information [6] while bringing about some more issues, like computational complexity and consuming time, to remedy.

Most of the current mechanisms utilize centralized differential privacy, which requests a third-party trusted entity to collect workers’ actual locations. To et al. [3] introduced a trusted Cellular Service Provider (CSP) to partition the domain of worker locations. Indeed, CSP has also the risk of being invaded, which destroys individual sensitive information. To avoid relying on trusted third-party entities, some authors pay attentions to obfuscation mechanisms with Local Differential Privacy (LDP) guarantee [7], [8], [9].

However, the existing solutions with LDP still have several limitations, which are summarized as follows:

1) Ignoring adversary’s inference attack. Some obfuscation mechanisms generate pseudo-locations with LDP guarantee via geo-indistinguishability [7], while not resisting against the Bayesian inference attack by prior knowledge [10].

2) Single-objective optimization. Generally, current privacy mechanisms adopt single-objective optimization [8]. In fact, we have to take into account the trade-off between two conflicting aspects: privacy level and service quality.

3) Small-scale location domain. Recent LDP protection mechanisms [8], [11] mainly adopt linear programming to improve the allocation efficiency that result in much higher computational complexity, while a previous scheme DPIVE [9] with Protection Location Sets (PLSs) can only be applied separately in small-scale disjoint domains.

To address the aforesaid issues, we design an adaptive obfuscation approach called ‘Geo-MOEA’ (Multi-Objective Evolutionary Algorithm with Geo-obfuscation) based on geo-indistinguishability, which achieves the distortion privacy guarantee on each worker location. Under our privacy framework, workers report their pseudo-locations with LDP directly to the server. The mechanism involving expected inference errors can effectively resist Bayesian attacks via prior knowledge. Moreover, the obfuscation scheme employs the genetic operator [12] to optimize the two conflicting metrics, expected inference error and service quality loss, for producing multiple optimized recommendations.

The main contributions are as follows:
1) We consider a large-scale SC scenario where the worker wants to protect individual true location without third-party trusted entities. For this, we propose “Geo-MOEA”, an obfuscation approach preserving differential and geo-distortion location privacy in all regions.

2) To the best of our knowledge, this is the first work to introduce a solution approach combining location privacy and MOEA to simultaneously optimize two conflicting metrics: service quality loss and average expected inference error.

3) We confirm theoretically the most general condition on qualified protection location sets for larger search space, by which service quality loss can keep mainly decreasing trend for ascending privacy budget.

4) A series of experimental results show that our proposal is more effective than the existing obfuscation mechanisms. Providing multiple recommendation mechanisms for practical applications can reduce workers travel distance while protecting workers’ location privacy in a global and scalable sense.

The remainder of this paper is structured as follows. In Section II, we conduct a survey of related work. Section III introduces system model, goal and adversary model. Section IV describes our designed privacy approach Geo-MOEA. Experimental evaluations are presented in Section V. Finally, we conclude this paper in Section VI.

II. RELATED WORK

We review related works in the direction of classical and differential privacy protection methods as well as multi-objective optimization.

Classical privacy protection methods. With the rapid development of mobile technologies, Spatial Crowdsourcing (SC) has attracted extensive attention from both the academia and the industry while the arising location privacy problem has also attracted extensive attention [11, 2]. For this, traditional privacy protection methods mainly include spatial anonymity and cryptography. Specifically, spatial anonymity technology often uses a spatial anonymity area to replace the user’s accurate location information. As a common kind of spatial anonymity models in SC, this model requires that each of the released locations be indistinguishable from at least $k-1$ other locations [13]. It has the fatal weakness that $k$-anonymity is vulnerable to attacks of background knowledge and reference [14]. Homomorphic encryption [15] is a good model to ensure the confidentiality of task’s location policy [16] while it has disadvantages of large key size and low calculation efficiency. Hence, it is not practical for large-scale scenarios [17].

Differential privacy protection. The Differential Privacy (DP) method [18] is a strong privacy concept that has been widely used for privacy protection recently. In particular, To et al. [3] develop centralized DP based Private Space Decomposition (PSD) to protect location privacy. Gong et al. [4] protect workers’ location and reputation privacy by using reputation-based DP method. Both schemes rely on a trusted third-party entity to protect location privacy and some attacks may lead to the disclosure of true locations.

Geo-indistinguishability and distortion privacy. Local Differential Privacy (LDP) can help users to report their pseudo-locations. Based on geo-indistinguishability [19], some LDP frameworks are proposed for SC scenario [17, 11, 20]. Expected inference error [21] is a concept of distortion privacy complementary to geo-indistinguishability, and can provide strict privacy protection against Bayesian attacks. Shokri [8] combines both concepts to construct the Joint mechanism and optimizes utility using linear programming. Yu et al. [22] formally study their relationship and combine them by adding personalized error lower bounds. Later, DPIVE mechanism is presented and solves the previous privacy theory problem of intersections among protection location sets [9]. Recently, Liu et al. [23] propose an obfuscation mechanism with the Gamma distribution and use a game-theoretic approach to maximize two-users’ utilities while preserving desired location privacy radius. However, these works focus mainly on the optimization of utility and have no concern with global trade-off between the conflicting utility and distortion privacy.

Multi-objective optimization. With the rapid development of scientific and engineering areas, a variety of optimization problems containing multiple conflicting objectives have appeared [24]. This means that there does not exist a single solution optimizing all of the objectives. It is natural to find the Pareto optimal set that contains multiple solutions trading off between all of the objectives, where the improvement of one objective cannot be achieved without the deterioration of some other objectives. A number of swarm intelligence algorithms have been proposed to search for well-converged solutions by using conventional reproduction operators [25, 26], such as the genetic algorithm [12] and Particle Swarm Optimization (PSO) [27, 28]. To achieve better privacy protection and higher service quality, Zhang et al. [29] provide a PSO anonymization method to accelerate the process of finding similar attributes. However, in many occasions the convergence of the canonical PSO is premature and it suffers from local minima [30]. Zhang et al. [31] present a Multi-Objective Evolutionary Algorithm (MOEA) framework to protect private information based on the hybrid elite selection strategy. In order to obtain optimal task allocation with the differential-and-distortion geo-obfuscation, Wang et al. [31] use linear scaling to execute optimization of two concordant objectives while ignoring local inference error bounds. In this paper, for the large-scale SC data scenario, the MOEA approach is adopted to design Geo-MOEA, which realizes the trade-off between conflicting distortion privacy and service quality for the first time.

III. SYSTEM MODEL AND DEFINITIONS

This section first presents our system model, goal and adversary model, then we provide the problem statement and the concepts of multi-objective optimization.

A. Spatial Crowdsourcing Model

Spatial Crowdsourcing (SC) [11] is a new type of platform for efficient and scalable data collection of online crowdsourcing in the era of mobile Internet and sharing economy, which
TABLE I: Summary of Notations

| Symbol     | Definition                                                                 |
|------------|---------------------------------------------------------------------------|
| ε₉, ε₈    | Total privacy budget and privacy level on Φₖ                               |
| f(x’|x)     | Probability of reporting location x’ for the actual x                     |
| d(x, y)   | Travel distance between the locations x and y                             |
| ε₉, θ    | Geo-indistinguishability parameter and its deviation                       |
| Φ        | Protection Location Set (PLS)                                             |
| D(Φ)     | Diameter of Φ (the largest distance between points)                       |
| | | |

B. Goal - Achieving Differential Privacy

Differential Privacy (DP) [32] provides provable privacy protection for users. Regardless of the adversary’s prior knowledge, it ensures that any adversary cannot confirm the presence of a particular individual in the processed data set. As for location privacy notion, geo-indistinguishability [19] has been widely used to achieve differential privacy. In order to achieve DP on each Protection Location Set (PLS), we adopt the loose definition on the PLS introduced by [32].

Definition 1 ((ε₉, θ)-Geo-indistinguishability within PLS [32]). Suppose the probability distribution f(·|·) for a mechanism A satisfies, for any x, y in PLS Φ with the reporting range Y ⊂ X,

\[
f(x'|x) \geq e^{\epsilon_9(d(x,y)+\theta)}, \quad x' \in Y,
\]

then A is (ε₉, θ)-geo-indistinguishable on Φ. If θ = 0, we say that A gives ε₉-geo-indistinguishability on Φ without deviation.

This means that any two locations geographically close to each other have similar probability distributions, so that they are theoretically indistinguishable to each other for the adversary. Particularly, the geo-indistinguishability parameter ε₉ is related to the privacy budget and the circular region usually centered at the user’s location. If all locations in a region have similar release distributions f, then the true location can be hidden in this region and the whole locations in this region are called Protection Location Set (PLS). Afterwards, we define ε = ε₉ · (D + θ), where D denotes the diameter of the PLS region. By Definition 1 the mechanism f satisfies ε-DP on PLS as follows.

Definition 2 (ε-DP on PLS [22]). A randomized location obfuscation mechanism f(·|·) satisfies ε-differential privacy on protection location set Φ, if for any locations x, y ∈ Φ, and any output x’ ∈ Y ⊂ X, we have

\[
\frac{f(x'|x)}{f(x'|y)} \leq e^{\epsilon}.
\]

As above, the local DP is achieved via geo-indistinguishability operating on a single dataset X, we use the notion ε-DP instead of ε-LDP as in the previous literatures [8], [11]. This paper continues to consider geo-indistinguishability and local DP on each PLS and some more general cases. Any two locations from the same PLS are regarded to be neighboring. Further, as a quite general approach preserving DP, the exponential mechanism usually involves a scoring function q : Φ × Y → R which assigns a real-valued score to each point-point pair, ideally such that each point x’ from the reporting range Y having better utility receives a higher real score for a given x ∈ Φ.

Definition 3 (Sensitivity on PLS [33]). Let x₁, x₂ ∈ X be any pair of neighboring locations (in PLS Φ). The sensitivity of the scoring function q on Φ is given by, its maximal change, with x’ ∈ Y,

\[
\Delta q = \sup_{x_1, x_2, x'} |q(x_1, x') - q(x_2, x')|.
\]

Definition 4 (Exponential Mechanism on PLS [34], [35]). Given a scoring function q on Φ × Y, the exponential mechanism M(x, q) outputs x’ ∈ Y with probability proportional to exp(q(x, x’)/2εq).

C. Bayesian Adversary Model

As all the Location-Based Service (LBS) providers require the access permission to users’ location data, the location privacy is potentially disclosed to untrusted entities. Knowing user’s locations, an adversary can perform a broad spectrum of attacks. Thus, ensuring location privacy is foremost for LBS applications. A common method is location perturbation, which generates a pseudo-location based on the true location and the user sends it to the server. Following [35], [22], we suppose that the discretized location set X represents the user’s possible locations. An obfuscation mechanism takes the user’s real location x from A as input and randomly chooses a pseudo-location x’ from O with the probability f(x'|x):

\[
f(x'|x) = \Pr(O = x'|A = x), \quad x, x' \in X.
\]
In general, the objective of obfuscation mechanisms is mainly to design suitable probability distribution \( f(\cdot | \cdot) \). Similar to [22], we assume that the adversary has prior knowledge about user’s location, which is regarded as background knowledge to perform inference attacks [10]. The adversary collects background knowledge by building a prior probability distribution \( \pi \) on \( \mathcal{X} \). The prior probability \( \pi \) can be obtained via population density, historical locations and so on. The adversary is also informed of the obfuscation mechanism \( f \).

In the current scenario, the adversary infers the user’s real location \( x \) under the Bayesian adversary model. After the user reports her/his pseudo-location \( x' \in \mathcal{X} \), the adversary computes the probability that each apriori location \( x \in \mathcal{X} \) is the true location in the condition of generating \( x' \), i.e., the posterior probability distribution \( \Pr(x|x') \), by

\[
\Pr(x|x') = \frac{\Pr(x,x')}{\Pr(x')} = \frac{\pi(x)f(x'|x)}{\sum_{x \in \mathcal{X}} \pi(x)f(x'|x)}. \tag{5}
\]

Afterwards, a Bayesian adversary can launch optimal inference attack [10] to get the estimated location \( \hat{x} \) which has the minimal expected distortion, i.e.,

\[
\hat{x} = \arg \min_{y \in \mathcal{X}} \sum_{x \in \mathcal{X}} \Pr(x|x')d(y, x). \tag{6}
\]

Now we come to the conflicting metrics. The distortion privacy level is measured by unconditional expected inference error [21],

\[
ExpErr = \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \pi(x)f(x'|x)d(\hat{x}, x). \tag{7}
\]

The service quality loss is defined by the unconditional expected distance between true and perturbed locations [8],

\[
QLoss = \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \pi(x)f(x'|x)d(x', x). \tag{8}
\]

D. Problem Statement

With the application of mobile devices, location-based SC services are spreading to many cities. The rapid growth of service data brings two serious challenges. Firstly, task allocation is often oriented to large-scale user location privacy data (such as DiDi, MeiTuan and Uber). The existing geo-distortion mechanisms (without a trusted third party) are only suitable to generate pseudo-locations with probability distributed in a small-scale domain [9], [8]. Adding the domain scale would increase greatly the quality losses while boosting the computational complexity.

It is desired to develop a LDP framework to avoid the constraint of small-region boundaries during location reporting and task allocations. In detail, a new domain partition method for generating families of PLSSs is necessary for achieving LDP based on geo-indistinguishability and implementing distortion privacy against Bayesian attacks.

Secondly, most of existing privacy mechanisms ignore the balance between privacy protection and practical quality. Currently, there is no standard optimization of conflicting metrics for SC in a global sense when workers have personalized requirements of distortion privacy [9], [11].

This motivates us to combine location privacy with Multi-Objective Evolutionary Algorithm (MOEA) involving Geo-obfuscation. The problem formulation is as follows:

Given some discrete settings of two privacy knobs, differential privacy budget \( \epsilon_0 \) and inference error bound \( E_m \), each idle worker in the large-scale domain of discrete locations is required to select one privacy setting and submit it to the server/platform for recommendation of multiple mechanisms. After this, the worker chooses an obfuscation mechanism to generate a pseudo-location for possible allocation of SC task. We focus on how to generate the multiple solutions at the SC-server side, which should achieve (Pareto-)optimal trade-off between service quality loss and average expected inference error globally as well as the above differential and geo-distortion privacy requirements.

E. Multi-Objective Optimization

In order to address the trade-off issue between privacy protection and practical quality, we introduce some basic concepts and methods related to Multi-Objective Optimization (MOO) [24], which has attracted much attention in the computational intelligence community, see [25] for a survey.

The definition of MOO problems is given by

\[
\min \mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_m(x)), \tag{9}
\]

where \( x = (x_1, x_2, \ldots, x_d) \) denotes the \( d \)-dimensional decision vector of a solution from decision space \( \Omega \), and \( \mathbf{f}(x) \) is an objective vector involving \( m \) conflicting functions.

Definition 5 (Pareto Dominance [23]). For any two solutions \( x \) and \( y \), if \( f_i(x) \leq f_i(y) \) for all \( i = 1, 2, \ldots, m \) and \( f_i(x) < f_i(y) \) for at least one \( i \), then \( x \) Pareto dominates \( y \).

Given a Pareto dominance set \( S = \{P_1, P_2, \ldots, P_n\} \) consisting of \( n \) non-dominated two-objective vectors, the unary hypervolume indicator \( HV \) is a measure of the region which is dominated by \( S \) and bounded from above by a reference point \( r \in \mathbb{R}^2 \) such that \( r \geq (\max P_1, \max P_2) \), where \( P = (p_1, p_2) \in S \subseteq \mathbb{R}^2 \), and the relation \( \geq \) applies componentwise. Fig. 1 shows that this region consists of an orthogonal polytope and can be regarded as the union of \( n \) axis-aligned rectangles with a common vertex \( r \).

Definition 6 (Pareto Optimality [23]). A solution \( x \) is Pareto optimal, if there does not exist any solution \( y \) dominating \( x \) in the decision space \( \Omega \).

Obviously, all Pareto optimal solutions are non-dominated with each other. In the current SC scenario, we
will mainly consider two conflicting factors, service quality loss and expected inference error by using MOEA while in each privacy setting Pareto-optimal solutions are gradually improved by evolutions. The evolutionary process would be stopped when the HV value for Pareto-dominance set of solutions almost converges.

IV. OUR PROPOSED GEO-MOEA APPROACH

This section describes our privacy protection approach Geo-MOEA, including basic model, domain partition, local adaptive obfuscation, partition MOEA and task matching.

A. System Framework

We consider the privacy protection problem of Spatial Crowdsourcing (SC) worker locations in the SAT mode. Fig. 2 describes our proposed seven-stage system model in SC scenario that consists of three parts: SC-server, requesters and workers. The SC-server is assumed to be semi-honest and intentionally deduces sensitive information of locations from the workers which has to be protected locally at worker side.

Each idle worker downloads the public obfuscation model, generates pseudo-location with Local Differential Privacy (LDP) guarantee by inputting privacy parameters in private and sends the obfuscation matrix \( f \) and the false location to SC-server. Requesters submits tasks and exposed location information to SC-Server. After receiving a task request, the SC-server determines the allocation by querying the locations reported by workers and initiates the geocast process. It also plays the role of communications, and computations to determine task allocations are carried out at the SC-server part. Possible disclosure of worker location and identity after his/her consent to the task is outside our scope.

As summarized by Algorithm 1, the main steps of Geo-MOEA include:

1. **Step 1. Domain Partition.** The SC service domain is divided into \( 2^s \) cells for some integer \( s \) where each cell \( X_i \) has almost the same count of discrete locations. The binary partition method is carried out iteratively according to the widthwise and lengthwise directions (the longer distance is preferred). This ensures that each cell obtained covers a required amount of locations and their convexity and compactness are satisfied.

2. **Step 2. Initializing Local Adaptive Obfuscation Mechanism.** Each worker downloads in mobile networks the obfuscation model of the generation matrix from the server side, and inputs preferred privacy parameter values to initialize Protection Location Set (PLS) partitions. Each PLS is restricted in one \( X_i \) and meets the lower bound of inference error and geo-indistinguishability. For each PLS \( \Phi_j \), the mechanism adaptively generates a reporting range \( Y_j \) composed of multiple connected PLSs, with which several cells \( X_i \)'s may be associated. Such partitions are repeated randomly at each MOEA iteration.

3. **Step 3. Partition Optimizations by MOEA.** At each iteration the obfuscation mechanism utilizes genetic evolution to generate a new partition family and improves the Pareto-optimal solutions (graph). After achieving the convergence of graph’s HV value, the MOEA outputs the set of Pareto-optimal partitions. The worker selects one solution and inputs his/her actual location, then a pseudo-location is generated by the exponential mechanism to be reported to SC-server.

4. **Step 4. Task Matching.** According to the collected pseudo-locations, the SC platform informs the three nearest available (idle) workers around the task location. Finally, the task is assigned to the earliest responder.

As shown in Fig. 3, the privacy framework of Geo-MOEA is mainly composed of two components: the partitioning algorithm \( F_M \) (to determine a family of PLS partitions) and the differentially private mechanism \( K \) (to generate a pseudo-location). \( F_M \) has four inputs, prior distribution \( \pi \), inference error threshold \( E_m \), privacy parameter \( E_0 \) and location sets \( X \). The sets \( X_i \)'s are derived from the binary partition of \( X \). For the two privacy parameters, users are allowed to control the

![Fig. 2: The SC scenario of Geo-MOEA.](image)

![Algorithm 1 Basic steps of Geo-MOEA](image)

![Fig. 3: The privacy framework of Geo-MOEA.](image)
posterior information leakage via the provisioning of privacy budget $\epsilon_0$ and specify $E_m$ to bound the expected inference error in the worst case. Each PLS contains obviously at least two locations and ensures the lower error bound.

By combining $F_M$ with MOEA, we conduct multi-objective optimization in each $X_i$ to generate a PLS partition family $\{\hat{F}^{(l)}_k\}$ with Pareto optimality, where the symbol $l$ stands for the partition candidates for users and the range of $k$ locally depends on $\hat{F}^{(l)}_k$. Obviously, the output of our algorithm $F_M$ is independent of the true location due to its no input. $F_M$ partitions each $X_i$ into Pareto-optimal disjoint PLSs with setting [13]. Then the mechanism $K$ utilizes the exponential mechanism to calculate the distribution matrices of reporting probability $\{f^{(l)}(x_j|x_i)\}$ where the diameter of its PLS is assigned locally as the sensitivity in each $x_i$’s row. After individual partition selection $\hat{F}^{(l)}_{i_0}$ that gives a public obfuscation matrix $\{f^{(l)}(x_j|x_i)\}$, a pseudo-location is produced with the input of true location.

Such a privacy framework supports Geo-MOEA to provide workers with differential and geo-distortion location privacy guarantee in all regions against Bayesian inference attacks. In particular, we develop a binary partition method adaptively producing cells, provide obfuscation mechanisms with locally adaptive reporting ranges and construct a novel multi-objective genetic algorithm with crossovers and mutations of PLS partitions.

B. Domain Partition

To ensure the local error bound $E_m$, we divide the domain into disjoint PLSs. However, the large-scale domain boosts greatly the computational complexity of searching optimal PLSs. For this, we first partition the domain into region $X_i$’s in each of which PLSs are obtained locally later.

First, we circumscribe the domain with a rectangle like urban roads according to the longitude and latitude spans. The added domain is naturally empty for locations. Next, we divide the domain with a rectangle like $X_i$ by binary partitions. At each iteration, each rectangle is divided, along the larger edges, into two smaller rectangles including equal number of locations. The recursive partitions end until each cell achieves the expected range of location number. Each final cell is denoted by the region $X_i$. Compared to the existing spatial decompositions [37], such a partition method ensures the convexity and compactness in each cell.

To be specific, assume that the total number of workers is $N$ and the range of expected number of locations in each cell is $[n_0, 2n_0]$, then there are $\lceil \log_2 \frac{N}{n_0} \rceil$ levels of partitions. Fig. 4 shows the partitions of $N = 400$ locations with $n_0 = 33$.

In the first level, since the rectangular length is larger than its width, it is divided into two rectangles, one left and the other right, equally covering 200 locations. The deeper partitions are processed recursively, which results in 8 cells.

C. Local Adaptive Obfuscation Mechanism

This section is the core part of Geo-MOEA, offering the initialization of obfuscation schemes involving adaptive reporting ranges. The flowchart is shown in Fig. 5 (PART 1).

Given the user-defined threshold, $ExpEr(x') \geq E_m$, for optimal inference attacks using any observed pseudo-location $x'$, the conditional expected inference error is

$$ExpEr(x') = \min_{x \in X} \sum_{x \in X} Pr(x|x')d(\hat{x}, x), \quad \text{for } x' \in X. \quad (10)$$

Given a partition $\{\Phi_k\}$ of $X$, we denote $z = \arg\min_{x \in X} \sum_{x \in X} Pr(x|x')d(\hat{x}, x), Pr(\Phi_k|x') = \sum_{y \in \Phi_k} Pr(y|x') \text{ and } \mathcal{Y} = \{x: Pr(x|x') > 0\}$. By normalization over each $\Phi_k$ (with $\epsilon_k$-DP) in $\mathcal{Y}$, we obtain, as in [32],

$$ExpEr(x') = \sum_{x \in \mathcal{Y}} Pr(x|x')d(z, x)$$

$$= \sum_k \sum_{x \in \Phi_k} Pr(x|x')d(z, x) \geq \sum_k \min_{x \in X} \sum_{x \in \Phi_k} Pr(x|x')d(\hat{x}, x)$$

$$= \sum_k \min_{x \in X} \sum_{x \in \Phi_k} \sum_{x \in \Phi_k} Pr(x|x')d(\hat{x}, x) = \sum_k \min_{x \in X} \sum_{x \in \Phi_k} \sum_{x \in \Phi_k} \pi(x)f(x'|x)d(\hat{x}, x)$$

$$= \sum_k \min_{x \in X} \sum_{x \in \Phi_k} \sum_{x \in \Phi_k} \pi(x)f(x'|x)d(\hat{x}, x) = \sum_k \min_{x \in X} \sum_{x \in \Phi_k} \pi(x)e^{-\epsilon_k}f(x'|x)$$

where

$$E'(\Phi) = \min_{x \in X} \sum_{x \in X} \sum_{x \in \Phi} \pi(x)e^{-\epsilon_k}f(x'|x). \quad (12)$$

Since $\sum_k \Pr(\Phi_k|x') = 1$, the condition that for all $\Phi_k$,

$$E'(\Phi_k) \geq e^{\epsilon_k}E_m, \quad (13)$$

implies the user-defined threshold, $ExpEr(x') \geq E_m$. Following this, we have the extended assertion:

Theorem 1. Given a domain partition $\{\Phi_k\}$ and an observed pseudo-location $x'$, suppose that an obfuscation mechanism satisfies $\epsilon_k$-differential privacy on each PLS $\Phi_k$. If $E'(\Phi_k) \geq e^{\epsilon_k}E_m$ for each $\Phi_k$, $ExpEr(x') \geq E_m$ for the optimal inference attack.

While allowing adaptive privacy specification $\epsilon_k$ on each PLS, the rule of [13] with using the identical privacy budget $\epsilon_k = \epsilon_0$ is so strict that some candidates are skipped for determining PLSs. This motivates us to search for optimized
Given the user’s location $x$ for minimizing quality losses with resistance to attacks. Due to (14), we have, for each $f(x) = \exp(-\theta|x|)$, then the optimal inference using any partition for disjoint PLSs $K$ can be rewritten shortly as

$$L(\epsilon_g) = \frac{\sum_j d_j \exp(-\epsilon_g d_j)}{\sum_j \exp(-\epsilon_g d_j)}.$$  

Then the first derivative is

$$L'(\epsilon_g) = \frac{\sum_j d_j e^{-\epsilon_g d_j} - \left(\sum_j d_j^2 e^{-\epsilon_g d_j}\right) \left(\sum_j e^{-\epsilon_g d_j}\right)}{\left(\sum_j \exp(-\epsilon_g d_j)^2\right)^2} \leq 0,$$  

where the last relation is due to Cauchy-Schwarz inequality with $a_j = d_j e^{-\epsilon_g d_j}/2 \geq 0$ and $b_j = e^{-\epsilon_g d_j}/2 \geq 0$. Hence, $L(\epsilon_g)$ is monotonically decreasing with $\epsilon_g$. 

Theorem 2 yields that for each PLS the larger $\epsilon_g$ produces the smaller quality losses. In the PART 1, based on the most general condition, $E'(\Phi) > E_m$, as shown in our designed Algorithm 2. We first execute QK-means algorithm from [9] on each cell $X_i$ to find its optimal number $k_i$ of disjointed PLSs with satisfying

$$E'(\Phi_k) \geq 6^0 E_m.$$  

The modifications include that Ret-C has no recursions on clustering centers and sets $\epsilon_k$ by Eq. (15) in each PLS $\Phi_k$ as shown in A.6

Any two PLSs do not intersect with each other. Each PLS is involved in only one cell $X_i$. The mechanism repeats such PLS partitions to initialize a population of $n$ solutions over $X$ (n families of partitions). For each PLS $\Phi_j$, in a partition, the mechanism construct a reporting range $Y_j$ by adding the nearest PLS to $\Phi_j$ recursively according to their centers’ distance, until that at least two PLSs are included and the number of locations is not less than 50. In this way, each reporting range $Y_j$ is not simply the involved cell $X_i$, but sometimes intersects with several cells. However,
Calculate $E'(\Phi) > E_n$.$E'(\Phi) \geq e^{\epsilon_0} \cdot E_n$.

**Fig. 6:** Adaptive privacy allocation procedure for PLS candidates.

such initialized solutions are not optimal in general for the tradeoff of privacy level and service quality. Step 3 is required for further optimization.

**D. Partition MOEA and Task Matching**

This section introduces the multi-objective genetic algorithm into optimizing the PLS partition population and generates a pseudo-location for task matching. Specifically, our innovated random combinations and variations of cluster centers in each cell lead to the crossovers and mutations of PLS partitions, respectively, which expands largely the solution space of the privacy problem.

The mechanism aims to generate higher unconditional expected inference error (ExpErr) and smaller service quality loss (Qloss) that conflict with each other. Following the line of MOEA, we describe the Pareto sets with regarding QLoss and minus ExpErr as two objective functions and use the unary Hypervolume (HV) indicator to measure the convergence and diversity of solutions. HV [26] is a usual performance metric for Multi-Objective Optimization (MOO) problems that is the only unary indicator strictly monotonic with Pareto dominance. In our setting, we first find the reference point $Q$ whose coordinates are defined as the maximum values of two objectives for each Pareto-dominated solution set, respectively. Then the HV value is defined as the area of the rectangular region surrounded by point $Q$ and each point in Pareto-dominated solution set with subtracting the overlapped area. We adopt the inverse number of ExpErr since two objectives in MOO prefer larger values (or both smaller).

The PART 2 in Fig. 5 shows the genetic evolution of PLS partition populations. After calculating the two metrics for each population, the mechanism makes fast non-dominated partition populations. After calculating the two metrics for each population, the mechanism selects a preferred solution with personalized threshold and carry out the exponential mechanism to produce a pseudo-location that is reported to SC-server. For each true location $x \in \Phi_j$, the utility (scoring) of output $x' \in \mathcal{Y}_j$ can be defined as $u(x, x') = -d(x, x')$. The sensitivity of $u$ for each PLS $\Phi_j$ is its diameter [9].

After a worker uploads the pseudo-location, the server improves the worker’s information. Once a task request is received, the server geocasts the task to the three idle workers closest to the task according to the collected pseudo-locations, see Fig. 4. The worker who approves first is qualified to execute the task. The task and the worker may be really located in different cells $X_i$ and even are far away from each other with a small probability. The metrics of worker travel distance (WTD) denoting their actual distance measures the service availability to a certain extent.

According to Theorem 4, the mechanism satisfies the $\epsilon_0$-differential privacy on each PLS $\Phi$. Moreover, for any two locations from different PLSs whose reporting ranges $\mathcal{Y}_j$’s intersect with each other, Geo-MOEA preserves weak differential privacy with a deviation factor on the coefficient as stated in Theorem 5.

**Theorem 4.** Assume disjoint PLSs $\{\Phi_j\}$, then the exponential mechanism $\mathcal{K}$ in Geo-MOEA satisfies $\epsilon_0$-differential privacy on each $\Phi_j$.

**Proof.** For any $x, y \in \Phi_j$ and $x' \in \mathcal{Y}_j$, we have by (15).

\[
\frac{f(x'|x)}{f(x'|y)} = \frac{w_{\Phi_j}(x)}{w_{\Phi_j}(y)} \exp \left( \frac{-\epsilon_j d(x, x')}{2D(\Phi_j)} \right) \\
\leq \frac{w_{\Phi_j}(x)}{w_{\Phi_j}(y)} \exp \left( \frac{-\epsilon_j d(x, y)}{2D(\Phi_j)} \right) \\
\leq \sum_{(x, y) \in \mathcal{Y}_j} \exp \left( \frac{-\epsilon_j d(x, y)}{2D(\Phi_j)} \right) \\
\leq \exp \left( \frac{\epsilon_j}{2} \right) \exp \left( \frac{\epsilon_j}{2} \right) \leq \exp(\epsilon_0).
\]

where $\epsilon_g = \epsilon_j/(2D(\Phi_j)).$

This shows that the mechanism $\mathcal{K}$ achieves $(\epsilon_g, D(\Phi_j))$-geo-indistinguishability within each PLS $\Phi_j$.

**Theorem 5.** Assume that $\mathcal{Y}_i$ and $\mathcal{Y}_j$ intersecting with each other are the reporting ranges for disjoint PLSs, $\Phi_i$ and $\Phi_j (i \neq j)$, respectively. Then the exponential mechanism $\mathcal{K}$
Fig. 7: Distribution charts with binary partitions.

satisfies \( \left( \frac{D(\mathcal{Y}_i)}{D(\Phi_i)} + \frac{D(\mathcal{Y}_j)}{D(\Phi_j)} \right) \leq \frac{\epsilon_0}{2} \) differential privacy on \( \Phi_i \cup \Phi_j \) with a deviation: For any \( x \in \Phi_i, y \in \Phi_j \) and \( x' \in \mathcal{Y}_i \cap \mathcal{Y}_j \),

\[
\frac{f(x'|x)}{f(x'|y)} \leq \frac{\left| \mathcal{Y}_j \right|}{\left| \mathcal{Y}_i \right|} \exp\left( \frac{\epsilon_0}{2} \left( \frac{D(\mathcal{Y}_j)}{D(\Phi_j)} + \frac{D(\mathcal{Y}_i)}{D(\Phi_i)} \right) \right).
\]

Proof.

\[
\begin{align*}
\frac{f(x'|x)}{f(x'|y)} & = \frac{\exp\left(-\frac{\epsilon_i d(x,x')}{2D(\Phi_i)}\right) \sum_{s \in Y_i} \exp\left(-\frac{\epsilon_j d(y,s)}{2D(\Phi_j)}\right)}{\exp\left(-\frac{\epsilon_j d(y,x')}{2D(\Phi_j)}\right)} \\
 & = \frac{\sum_{s \in Y_i} \exp\left(-\frac{\epsilon_i d(y,s)}{2D(\Phi_j)} + \epsilon_j (d(y,x') - d(y,t))\right)}{\sum_{s \in Y_i} \exp\left(-\frac{\epsilon_j d(y,s)}{2D(\Phi_j)} + \epsilon_i (d(y,x') - d(y,t))\right)} \\
 & \leq \frac{\left| \mathcal{Y}_j \right|}{\left| \mathcal{Y}_i \right|} \exp\left( \frac{\epsilon_j D(\mathcal{Y}_j)}{2D(\Phi_j)} + \frac{\epsilon_i D(\mathcal{Y}_i)}{2D(\Phi_i)} \right) \\
 & \leq \frac{\left| \mathcal{Y}_j \right|}{\left| \mathcal{Y}_i \right|} \exp\left( \frac{\epsilon_0}{2} \left( \frac{D(\mathcal{Y}_j)}{D(\Phi_j)} + \frac{D(\mathcal{Y}_i)}{D(\Phi_i)} \right) \right).
\end{align*}
\]

The assertions obtained above for variable reporting ranges provide strictly theoretical support on the differential privacy protection guarantee in the large-scale SC scenario.

V. PERFORMANCE ANALYSIS

This section compares our Geo-MOEA approach with some previous mechanisms on the metrics of location privacy and service quality, then presents Pareto analysis, application analysis and visualization analysis. The results show that our mechanisms greatly improve service availability on the basis of ensuring individual location privacy.

A. Experimental Methodology

Datasets. Since SC services are mainly located in urban domains, we investigate its distribution characteristics in cities and sample two real datasets, NYTaxi is a dataset of taxi locations in New York. We extract 3000 positions picked up (longitude from \(-73.982738 \) to \(-73.982742 \), latitude from \(40.750562 \) to \(40.760048 \)) on May 21, 2013. Since most positions are distributed on the roads, we randomly and uniformly blur each position into a circle centered at the pickup position and with radius (Blur Radius, BR) of 80 meters. Gowalla is a dataset of social network check-in locations. We extract 1000 locations (longitude from \(-179.9910 \) to \(-179.9641 \), latitude from \(89.9232 \) to \(89.9330 \)). Gowalla involves more relatively sparse regions. Fig. 7 presents their distribution charts with binary partitions. In each chart, all cells \( X_i \) are marked in different colors. We simulate a prior distribution uniformly in which each value is sampled randomly in [0.00028, 0.00038] and [0.0005, 0.0015] on two datasets, respectively.

Parameter settings. The minimum expected number of locations in each cell is \( n_0 = 33 \). Geo-MOEA gives two privacy control knobs: 1) differential privacy parameter \( \epsilon_0 \in \{0.10, 0.20, \ldots, 1.50\} \) and 2) threshold of expected inference error \( E_{m} \in \{0.050, 0.075, \ldots, 0.300\} \). The default setting is given by \( \epsilon_0 = 1.0 \) and \( E_{m} = 0.1 \). Its maximum iteration times is 500. We sample randomly 200 tasks on the distributed locations with 2000 and 900 idle workers in each dataset, respectively. We consider two types of idle workers’ distributions: (1) 80% of idle workers are densely located in 20% space while the remaining 20% workers are distributed in the rest of 80% space (denoted by “1:4 mode”). (2) The idle workers are uniformly distributed in each dataset (denoted by “uniform mode”). We assume that different tasks have no spatio-temporal confliction to each other. This means that they can share a single worker. Their WTDs are computed on average. On comparisons of the metrics WTD, for the Pareto solutions of Geo-MOEA, only the extreme solution with the smallest QLoss is taken into account.

B. Privacy Protection Goals

The privacy protection goals of the experiments for applications are as follows. (1) preserving \( \epsilon_0 \)-DP inside each PLS based on geo-indistinguishability and weak DP in the large-scale data scenario. (2) achieving distortion privacy with lower bound \( E_{m} \) against Bayesian attack even on isolated locations. and (3) improving the data utility for applications.

We mention that our proposed Geo-MOEA is able to defeat Bayesian adversary attacks locally and globally. Locally, the lower error threshold \( E_{m} \) controls the (conditional) expected inference error ExpErr for each location involving the isolated regions. Globally, the unconditional expected inference error ExpErr quantifies the degree of resistance to Bayesian inference attack, which provides a method to pick a solution from the Pareto optimal recommendations.

Let us make comparisons on various requirements with DPIVE [9], Joint [8] and PSO [27] (the original Particle Swarm Optimization), see Table II. Only Geo-MOEA meets

|                | Local DP | \( E_m \) | Scalability | Genetic evolution |
|----------------|----------|-----------|-------------|-------------------|
| Joint [8]      | ✔        | ✔         | ✗           | ✗                 |
| DPIVE [9]      | ✔        | ✗         | ✗           | ✗                 |
| PSO [27]       | ✗        | ✔         | ✔           | ✗                 |
| Geo-MOEA       | ✔        | ✔         | ✔           | ✗                 |

TABLE II: Comparison of Geo-MOEA with other mechanisms on various requirements
all requirements. These mechanisms can satisfy the (local) lower inference error bound $E_m$ except Joint. DPIVE and Joint are not suitable for large-scale location domain scenarios while the PSO algorithm produces bad solutions due to its linear-weighted combination approach for multi-objective optimization. Geo-MOEA involving expected inference errors can effectively resist Bayesian attacks via prior knowledge, and meets the requirements of large-scale location domain. In essence, the three existing mechanisms adopt single-objective optimization while our proposed Geo-MOEA takes into account the trade-off between two conflicting aspects: distortion privacy level and service quality.

C. Pareto Analysis

In this section, we compare Geo-MOEA with DPIVE [9], Joint [8] and PSO [27], especially to verify the efficiency in balancing service quality loss and average expected inference error as well as some extreme solutions. Since the lower bound $E_m$ ensures the conditional expected inference error locally, we pay more attention to the global performance on the metrics, QLoss, Minus ExpErr, and HV.

The DPIVE carried out in each cell $X_i$ separately is a single-objective optimization algorithm that aims to minimize the quality loss while ensuring the lower bound $E_m$ of inference error. Joint is the first mechanism that uses linear programming to combine two privacy notions of expected inference error and geo-indistinguishability. We also carry out Joint in each cell $X_i$ separately and use the global expected inference error of DPIVE as the minimum desired distortion privacy level $d_m$, via adjusting $\epsilon_0$ to obtain the same expected inference error.

PSO converts our problem to single-objective optimization and proceeds without mutations and crossovers [27]. The fitness function defined by $F(\alpha) = \alpha \cdot \text{QLoss} - (1 - \alpha) \cdot \text{ExpErr}$ with $\alpha \in \{0.0, 0.1, \ldots, 1.0\}$ as in [11], which combines both objectives to optimize the obfuscation schemes.

Figs. 8 and 9 present a series of comparisons on the conflicting metrics of QLoss and ExpErr with the convergent Pareto dominance solutions on the two datasets. Each DPIVE solution is obviously dominated by some solutions of Geo-MOEA in the (a)-(i) settings. On the aspect of QLoss, DPIVE increases by 7.2% and 8.8% on average, and particularly 19.9% in case Fig. 8(c) and 24.9% (reversely, a 20.0% reduction) in case Fig. 9(h) respectively on the two datasets, compared to the extreme (left) mechanism with the smallest quality loss in the Pareto-recommendations generated by Geo-MOEA. The reasons include that, the reporting range of DPIVE is restricted in local small-scale cell while more PLSs are located in the corners, and DPIVE minimizes the average diameter of PLSs for optimizing QLoss while ignoring the fitness of ExpErr. Joint increases by 11.5% and 22.3% on average and a maximum increase of 24.9% and 52.4%, respectively on the
two datasets, compared to Geo-MOEA. Moreover, the Pareto solutions of Geo-MOEA are basically located at the lower left side of PSO solutions, which means that Geo-MOEA solutions can well Pareto dominate those of PSO globally. Indeed, the PSO algorithm also initializes its population, but lacks the crossover and mutation processes that help Geo-MOEA expand the solution space.

As a result, the HV of Geo-MOEA solutions is twice as large as that of PSO solutions on average and even 3.3 and 2.4 times, respectively on the two datasets as shown in Fig. 10. Hence, Geo-MOEA brings more optimized solutions, faster convergence and greater diversity while achieving privacy protection goals.

Fig. 11 shows that, compared with the extreme Geo-MOEA, DPIVE has an average increase of 3.9% and 7.7%, and a maximum increase of 13.7% and 25.9%, respectively on the two datasets. The larger the $E_m$, the earlier the branch point of curve appears. Once (20) is satisfied, the gap between the curves expands soon due to the possible retreat for larger $\varepsilon_g$. When $E_m$ is very small, there are no possibilities to retreat so that both curves are close to each other. However, as $\varepsilon_0$ becomes large, it is not so easy to satisfy (20) and there is larger search space for Geo-MOEA with $\varepsilon < \varepsilon_0$. 

![Fig. 9: Comparison of Geo-MOEA with other mechanisms on Gowalla dataset.](image)

![Fig. 10: Comparisons of Geo-MOEA with PSO on HV.](image)

![Fig. 11: Comparisons on QLoss with varying $\varepsilon_0$.](image)
D. Engineering Applications

The distance WTD that the worker travels from the actual location to the allocated task determines the efficiency of mechanism application. For convenience, we compute WTD by the Euclidean distance simply while the actual travel distance depends on the road networks. We conduct comparative experiments on the two datasets with varying privacy parameter $\epsilon_0$. Under each parameter setting, we average the WTDs on random tasks for comparisons of four schemes. The notation Non-privacy means Geo-MOEA without privacy protection, that is, the SC-server geocasts the three idle workers closest to the task directly based on the real locations and their average WTD is referred to.

As shown by Fig. 12(a) and Fig. 12(b) for the 1:4 mode, on comparisons to the extreme mechanism with the smallest quality loss among the Pareto-optimality recommendations generated by Geo-MOEA, DPIVE has an average increase of 6.8% and 9.8%, and a maximum increase of 10.1% and 10.9%, respectively, on the two datasets, while PSO has an average increase of 23.8% and 15.4% and a maximum increase of 35.7% and 17.4%, respectively. In Fig. 12(c) and Fig. 12(d) for the uniform mode, DPIVE has an average increase of 11.6% and 9.8%, and a maximum increase of 16.3% and 11.2%, respectively, on the two datasets, while PSO has an average increase of 18.9% and 15.7% and a maximum increase of 27.2% and 17.2%, respectively. Geo-MOEA performs multi-objective genetic algorithms, which helps to generate extreme mechanisms reducing largely the quality losses. On the other hand, the privacy protection for Geo-MOEA leads to higher WTD (30.5%, 26.4%, 22.5% and 22.1% on average in the four sub-figures, respectively) than the Non-privacy case due to the privacy level and the spare distribution of idle workers particularly for the 1:4 mode. This shows that our mechanism can protect well worker’s location privacy while improving the availability of existing SC mechanisms in large-scale scenario.

Further, we make analysis in terms of visualization. For the extreme Geo-MOEA mechanism with the smallest quality loss, we make visualization analysis on the experiments under 110 settings of privacy knobs, $E_m = 0.050, 0.075, \ldots, 0.300$ and $\epsilon_0 = 0.1, 0.2, \ldots, 1.0$ for each dataset. We adopt surface fitting by the metric HV and mark the surface in variable colors by fitting on the metric WTD. The idle workers are located randomly in the uniform mode as mentioned above. As shown in Fig. 13 for fixed $E_m$, the larger $\epsilon_0$ leads to higher HV and the more expanded solution space while generating smaller WTD naturally due to the lower level of privacy requirement. When $\epsilon_0$ remains unchanged, WTD presents an upward trend for increasing $E_m$. For both datasets, the metrics of HV and WTD reach respective extreme values at the case of $E_m = 0.05, \epsilon_0 = 1.0$. The surfaces are marked in green in the middle part around $\epsilon_0 = 0.6$ and more green part are expressed for Gowalla dataset due to its local denser property. Each curved surface provides a state of high HV and low WTD, which indicates that the Geo-MOEA algorithm reaches an extreme state. In this setting, the PLSs are small as well as injected noises and the multi-objective solution space is large, which makes the Pareto-optimal solutions more diverse.

Moreover, in most crowdsourcing applications, workers are not located in fixed locations. That is, their locations for most change dynamically. For this, we can set up some discrete locations $\{x_i\}$ such as check-ins and bus stations. The closest $x_i$ to the worker is regarded as her true location for adopting Geo-MOEA. In addition, the workers can join the platform and leave the platform dynamically by reporting.

E. Efficiency Evaluation

In real life, besides the quality of task assignment, the time cost of generating mechanisms should also be taken into account. The execution of Geo-MOEA mainly includes two stages. The first stage (stage A) is the utilization of DPIVE to divide the domain into regions $X_i$ with generating its corresponding $k_i$. In the second stage (stage B), the Pareto-optimal solutions are produced by using MOEA.

The testing environment consists of Intel(R) Xeon(R) Silver 4210R CPU 2.40GHz, 132GB SSD, Red Hat 4.8.5-44, and python version 3.8. Based on this, the stage A takes about 160 seconds and 100 seconds, respectively on the two datasets, while the stage B takes about 200 seconds and 140 seconds in various settings of privacy parameters, respectively. Since the configuration of our running environment is relatively low, the whole running time for each privacy setting can be reduced to
20) seconds for some advanced platforms. In the mobile device applications, the Pareto-optimal solutions for various privacy settings can be renewed on the service platform every half or an hour for effective SC.

VI. CONCLUSION AND FUTURE WORK

This paper implements the geo-indistinguishable location privacy protection of Spatial Crowdsourcing (SC) workers in mobile networks. Our proposed Geo-MOEA achieves to generate a pseudo-location among an adaptive reporting range with distortion privacy guarantee in the large-scale data domain. Together with a binary partition method, a multi-objective genetic algorithm for Protection Location Set (PLS) partitions is introduced into the Local Differential Privacy (LDP) protection, which makes the first attempt to realize the optimized tradeoff between quality loss and inference error. Moreover, the most general condition for qualified PLSs is presented, which helps theoretically to design a clustering algorithm with retreats for more search space, particularly reducing quality loss. Experimental results confirm strongly that Geo-MOEA outperforms the existing schemes in terms of service quality. In the future work, we plan to explore multi-task assignment problems with LDP that involves higher computational complexity and various requirements on communication environments. The specific environments include complex road networks, workers’ reputation, task location privacy, and the wide applications in cyber-physical systems.

REFERENCES

[1] L. Kazemi and C. Shahabi, “Geocrowd: Enabling query answering with spatial crowdsourcing,” in Proceedings of the 20th International Conference on Advances in Geographic Information Systems, 2012, pp. 189–198.

[2] Y. Tong, Z. Zhou, Y. Zeng, L. Chen, and C. Shahabi, “Spatial crowdsourcing: A survey,” The VLDB Journal, vol. 29, no. 1, pp. 217–250, 2020.

[3] 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, 2017.

[4] Y. Gong, C. Zhang, Y. Fang, and J. Sun, “Protecting location privacy for task allocation in ad hoc mobile cloud computing,” IEEE Transactions on Emerging Topics in Computing, vol. 6, no. 1, pp. 110–121, 2018.

[5] Q. Hu, S. Wang, X. Cheng, L. Ma, and R. Bie, “Solving the crowdsourcing dilemma using the zero-determinant strategies,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1778–1789, 2020.

[6] K. Chatzikokolakis, E. ElSalamouny, C. Palamidessi, and A. Pazli, “Methods for location privacy: A comparative overview,” Foundations and Trends® in Privacy and Security, vol. 1, no. 4, pp. 199–257, 2017.

[7] H. To, C. Shahabi, and L. Xiong, “Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server,” in 2018 IEEE 34th International Conference on Data Engineering, 2018, pp. 833–844.

[8] R. Shokri, “Privacy games: Optimal user-centric data obfuscation,” Proceedings on Privacy Enhancing Technologies, vol. 2015, no. 2, pp. 299–315, 2015.

[9] S. Zhang, P. Lan, B. Duan, Z. Chen, H. Zhong, and N. Xiong, “DPIVE: A regionalized location obfuscation scheme with personalized privacy levels,” ACM Transactions on Sensor Networks, 2021, minor revision, arXiv preprint arXiv:2102.00654.

[10] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, “Protecting location privacy: Optimal strategy against localization attacks,” in Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS), 2012, pp. 617–627.

[11] 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, 2019.

[12] J. H. Holland et al., Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT press, 1992.

[13] M. Yang, B. Ye, Y. Chen, T. Li, Y. Yang, X. Qian, and X. Yu, “A trusted de-swarming k-anonymity scheme for location privacy protection,” Journal of Cloud Computing, vol. 11, no. 1, pp. 1–15, 2022.

[14] W. Lin and Z. Shen, “Embracing differential privacy for anonymizing spontaneous ad landing pages,” Journal of International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2015–2022, 2020.

[15] J. Zhou, Z. Cao, Z. Qin, X. Dong, and K. Ren, “LPRA: Lightweight privacy-preserving authentication from efficient multi-key secure outsourced computation for location-based services in VANETs,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 420–434, 2020.

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

[17] X. Song and Y. Wang, “Homomorphic cloud computing scheme based on hybrid homomorphic encryption,” in 2017 3rd IEEE International Conference on Computer and Communications, 2017, pp. 2450–2453.

[18] C. Dwork, “Differential privacy,” in International Colloquium on Automata, Languages, and Programming. Springer, 2006, pp. 1–12.

[19] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, “Geo-indistinguishability: Differential privacy for location-based systems,” in Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, 2013, pp. 901–914.

[20] L. Wang, D. Zhang, D. Yang, Y. Brian, and X. Xiao, “Sparse mobile crowdsensing with differential and distortion location privacy,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2735–2749, 2020.

[21] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, “Quantifying location privacy,” in 2011 IEEE Symposium on Security and Privacy. IEEE, 2011, pp. 247–262.

[22] L. Yu, L. Liu, and C. Pu, “Dynamic differential location privacy with personalized error bounds,” in NDSS, 2017, pp. 1–15.

[23] M. Liu, X. Zhou, and M. Sun, “Bilateral privacy-utility tradeoff in spectrum sharing systems: An axiomatic approach,” IEEE Transactions on Wireless Communications, vol. 20, no. 8, pp. 5144–5158, 2021.

[24] J. Branke, T. Kaußler, and H. Schmeck, “Guidance in evolutionary multiobjective optimization,” Advances in Engineering Software, vol. 32, no. 6, pp. 499–507, 2001.

[25] K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182–197, 2002.

[26] Y. Tian, L. Si, X. Zhang, R. Cheng, C. He, K. C. Tan, and Y. Jin, “Evolutionary large-scale multi-objective optimization: A survey,” ACM Computing Surveys, vol. 54, no. 8, pp. 1–34, 2021.

[27] R. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in MHS ’95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. IEEE, 1995, pp. 39–43.

[28] M. Clerc and J. Kennedy, “The particle swarm—explosion, stability, and convergence in a multidimensional complex space,” IEEE Transactions on Evolutionary Computation, vol. 6, no. 1, pp. 58–73, 2002.

[29] L. Zhang, S. Yang, J. Li, and L. Yu, “A particle swarm optimization clustering-based attribute generalization privacy protection scheme,” Journal of Circuits, Systems and Computers, vol. 27, no. 11, 2018, 1850179.

[30] S. Das, A. Konar, and U. K. Chakraborty, “Improving particle swarm optimization with differentially perturbed velocity,” in Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation, 2005, pp. 177–184.

[31] J. Zhang, F. Xue, X. Cai, Z. Cui, Y. Chang, W. Zhang, and W. Li, “Privacy protection based on many-objective optimization algorithm,” Concurrency and Computation: Practice and Experience, vol. 31, no. 20, 2019, e5342.

[32] S. Zhang, B. Duan, Z. Chen, and H. Zhong, “On the differential privacy of dynamic location obfuscation with personalized error bounds,” arXiv preprint arXiv:2101.12602, 2021.

[33] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” Theory of Cryptography Conference, pp. 265–284, 2006.

[34] F. McSherry and K. Talwar, “Mechanism design via differential privacy,” 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2007, pp. 94–103.

[35] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” Foundations and Trends® in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.
[36] K. Huguenin, I. Bilogrevic, J. S. Machado, S. Mihaila, R. Shokri, I. Dacosta, and J.-P. Hubaux, “A predictive model for user motivation and utility implications of privacy-protection mechanisms in location check-ins,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 4, pp. 760–774, 2017.

[37] X. Li, Y. Wang, X. Zhang, K. Zhou, and C. Li, “A more secure spatial decompositions algorithm via indefeasible laplace noise in differential privacy,” *International Conference on Advanced Data Mining and Applications*, pp. 211–223, 2018.

**Shun Zhang** was born in Anhui Province, China, in 1982. He received his PhD degree in applied mathematics from Beijing Normal University in 2012. He was a visiting scholar at Friedrich-Schiller-Universitat Jena, Germany, from 2014 to 2015. He is currently an associate professor at Anhui University. He has published more than 30 scientific papers. His research interests include privacy preservation, computational complexity, and quantum secure multiparty computation.

**Tao Zhang** was born in Anhui Province, China, in 1998. He is currently a master student in Anhui University. His main research interests include differential privacy, location privacy.

**Zhili Chen** was born in Fujian Province, China, in 1980. He received his PhD degree in computer science from University of Science and Technology of China in 2009. He is currently a professor and Ph.D. supervisor at East China Normal University. He has published more than 40 papers. His main research interests include privacy preservation, secure multiparty computation, information hiding, spectrum auction and game theory in wireless communications.

**N. Xiong** is currently a Distinguished Professor at National Engineering Research Center for E-Learning, Central China Normal University (CCNU), Wuhan, Hu Bei Province, 430079, China. He is also with the Department of Computer Science, Georgia State University, Atlanta, GA 30302, USA. He received his PhD degree in School of Information Science, Japan Advanced Institute of Science and Technology (JAIST) on March 1, 2008. His research interests include Deep Learning, Reliable Networks, Software Engineering, and Big Data Analytics. Dr. Xiong works in CCNU for many years, and obtained many research funding and many industrial projects. He published over 600 journal paper with 200+ IEEE journal papers. He also creates a company about design and analysis for complex reliable software systems, and obtains over 10 patents.