Optimizing Location Quality in Privacy Preserving Crowdsensing

Yuhui Zhang Ming Li Dejun Yang Jian Tang Guoliang Xue

***Abstract*—Crowdsensing enables a wide range of data collection, where the data are usually tagged with private locations. Protecting users' location privacy has been a central issue. The study of various location perturbation techniques for protecting users' location privacy has received widespread attention. Despite the huge promise and considerable attention, the location perturbation operation causes inevitable location errors, which can diminish the location quality of the crowdsensing results. Provable good algorithms that consider location quality in privacy preserving crowdsensing from optimization perspectives are still lacking in the literature. In this paper, we investigate the problem of location quality optimization in privacy preserving crowdsensing, which is to minimize the location quality desegregation, while protecting all users' location privacy. We present an optimal algorithm OLoQ for this problem. Extensive simulations demonstrate that OLoQ significantly outperforms an existing algorithm in terms of the location quality and SSE.**

I. Introduction

Over the last decade, there has been an explosion of smart devices, e.g. smartphones and tablets. In 2015, there were available 3.2 billion smartphone subscriptions, with 6.2 billion predicted to be available in 2021 [14]. Current smart devices are embedded with increasingly powerful processors and a multitude of sensors (e.g., GPS, thermometer, microphone, camera). The ubiquity of mobile devices into everyday life can provide sufficient geographic coverage, especially in densely populated areas. The crowdsensing paradigm has been proposed to take advantage of the widely distributed mobile devices for sensing and collecting ubiquitous data, such as P-Sense to monitor air pollution [13], Nericell to sense road and traffic conditions [20], and Ear-Phone to construct urban noise maps [23]. The sensing data are usually tagged with locations to form a database or map for information release.

It is essential to achieve location privacy protection, since mobile users' locations are tightly correlated with their identities and vulnerable to malicious attacks. Upon preserving location privacy in crowdsensing, various methods are proposed including information caching [25], spatial cloaking [29], data perturbation with noise [35] and microaggregation [32]. The goal is to prevent the servers or platforms from inferring users' actual locations. However, these privacy preserving methods need to hide the users' actual locations, which usually degrade the location (information) quality [7].

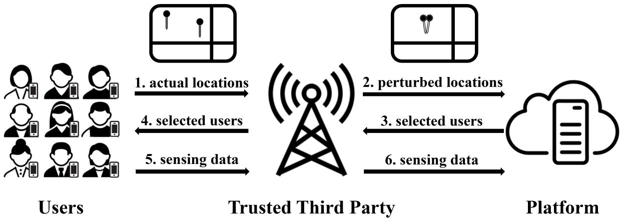


Figure 1: Location privacy preserving crowdsensing system

Location privacy and location quality are two conflicting concerns in crowdsensing. On the one hand, disclosing users' actual locations to the platform may severely discourage their participation, because users are increasingly wary of location privacy. On the other hand, the platform desires the actual locations of users to ensure the location quality. Therefore, it is essential to optimize location quality in location privacy preserving crowdsensing systems.

To quantify the impact of location privacy protection on location quality, we define the *location quality degradation* as the maximum distance between users' actual locations and their corresponding perturbed locations. The summation of squared location errors (SSE) [26] has also been used to measure location quality in the literature. Although minimizing the SSE is not our objective, our simulation results demonstrate that a small location quality degradation also implies a small SSE.

In this paper, we study the location quality optimization in privacy preserving crowdsensing. Specifically, we focus on the **Location Quality Degradation Minimization (LQDM)** problem: minimizing the location quality degradation, while protecting the location privacy for all users. We summarize **the main contributions** as follows:

- To the best of our knowledge, we are the first to consider the location quality optimization in location privacy preserving crowdsensing systems.
- We study the problem of location quality degradation minimization, while protecting the location privacy for all users.
- We design OLoQ, an **O**ptimal algorithm for **Lo**cation **Q**uality desegregation minimization, while protecting the location privacy for all users.

Zhang and Li are affiliated with Colorado School of Mines, Golden, CO 80401. Yang (corresponding author) is with Colorado School of Mines, Golden, CO 80401 USA and Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023 China. Tang is affiliated with Syracuse University, Syracuse, NY 13244. Xue is affiliated with Arizona State University, Tempe, AZ 85287. Email:{yuhzhang, mili, djyang}@mines.edu, jtang02@syr.edu, xue@asu.edu. This research was supported in part by NSF grants 1525920, 1704092, 1717197, and 1717315. The information reported here does not reflect the position or the policy of the federal government.

978-1-7281-0962-6/19/$31.00 ©2019 IEEE

- Extensive simulations demonstrate that OLoQ not only minimizes location quality degradation, but also outperforms an existing algorithm in terms of the summation of squared location errors (SSE).

The remainder of the paper is organized as follows. In Section II, we give a brief review of existing location privacy preserving mechanisms in the literature. In Section III, we formally introduce the system model and give a precise problem description. In Section IV, we present a polynomial-time optimal algorithm for **LQDM** and analyze its properties. Section V demonstrates the experimental evaluations. Section VI concludes this paper.

II. Related Work

A. Location Privacy Approaches

There is a rich collection of literature on location privacy in general frameworks. Surveys for location privacy-preserving methods can be found in [5, 10]. Following the discussions in [10], we classify location privacy-preserving techniques to three types: location generation [3, 16, 34, 36], cryptographic techniques [11] and differential privacy [15, 28]. Along the line of location generation, various methods are proposed including position dummies [16], mix zone [3], pseudonym [12], and k-anonymity [34].

B. Location Privacy Preserving Crowdsensing

Much effort has also been made to protect location privacy in crowdsensing systems [1, 9, 15, 19, 22, 30, 31]. This line of work aims at preventing location privacy leakage from sensing reports submitted by crowdsensing users. Gao *et al.* [9] designed a partner selection algorithm and construct several trajectories that are closer the user. Agir *et al.* [1] proposed a scheme which estimated the expected location-privacy level at the user-side locally in real-time, which satisfies each user's privacy requirement adaptively. Vu *et al.* [30] utilized Voronoi diagram to partition a space into cells that contain at least k users in each, without considering to minimize the cloaking area. Differential location privacy in the crowdsourced spectrum sensing was preserved in [15, 19, 31]. However, a significant problem neglected in these works is to optimize the crowdsensing platform's location quality, while protecting the users' location privacy.

C. Location Information Quality

As pointed by Krause *et al.* in [18], it is challenging to control location privacy with location privacy protection. Rodhe *et al.* [24] reconstructed the data distribution and investigated the impact of location privacy preserving mechanisms on the quality of information. Xiao *et al.* developed a directed-graph based cloaking algorithm for protecting location privacy in location-based service, while meeting user-specified quality of service requirements [33]. Murshed *et al.* proposed a subset-coding scheme to achieve almost lossless data integrity in [21].

Another related topic is the microaggregation problem: divide a data set into several disjoint subsets, such that the size of each subset is more than k and the sum of squared error is minimized. This problem aims to strike a balance between privacy protection and information loss reduction [6, 17, 26]. However, the location quality degradation minimization is not considered in the microaggregation problem.

III. System Model and Problem Formulation

In this section, we introduce the system model and give a precise problem description.

A. System Model

We consider a location-based crowdsensing system consisting of a set $\mathcal{U} = \{1, 2, \ldots, n\}$ of n users, a trusted third party [28, 32] (e.g., a cellular service provider) and a crowdsensing platform. Each user carries a mobile device with sensing capabilities and wishes to earn rewards by completing crowdsensing tasks. The user registers with the platform and communicates with the platform via an app installed on his mobile device. Developed by the platform, the app is assumed to pass the strict vetting process of the trusted app store and has no unauthorized access to the user's locations.

We assume that the platform is honest but curious, which is commonly used to characterize a reasonable crowdsensing platform. In particular, the platform is trusted to follow the protocol execution but is also interested in learning users' locations. We assume that the platform can have arbitrary prior knowledge for attempting to breach the users' location privacy.

A precision-aware location privacy preserving crowdsensing system is shown in Figure 1. The platform publishes crowdsensing tasks and collects location-aware sensing data from the users. The trusted third party, which is a cellular service supposed to protect the location privacy. The workflow of the system is as follows:

1) All the users report their actual locations $\mathcal{L} = \{l_1, l_2, \ldots, l_n\}$ to the trusted third party for location privacy protection.
2) The trusted third party processes the actual locations and reports a set of perturbed locations $\{h_1, h_2, \ldots, h_m\}$ to the platform, where a perturbed location h_j is tagged to at least k users.
3) The users tagged with perturbed locations are reported to the platform, and the rest users are discarded.

B. Problem Formulation

To formally formulate our studied problems, we introduce the following necessary concepts. In order to preserve location privacy, one solution is to make a user's location indistinguishable from at least $k-1$ others' locations. This property is proposed in [27] and called *k-anonymity*.

k-anonymity: To protect user's privacy, k-anonymity requires that at least k reports are combined together before releasing.

Location perturbation: Location perturbation is defined as deliberately degrading the quality of location information about a user's location in order to protect that user's location privacy.

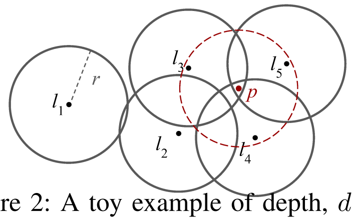


Figure 2: A toy example of depth, $d_{\mathcal{D}(\mathcal{L},r)}(p) = 3$.

Location quality degradation: The location quality degradation is the maximum distance between users' actual locations and their corresponding perturbed locations.

Perturbed group: A perturbed group is a set of users $\mathcal{S} \subseteq \mathcal{U}$ tagged with the same perturbed location, denoted by $(h, \mathcal{S})$, satisfying k-anonymity.

Apparently, the perturbation operation for protecting users' location privacy causes inevitable location errors, which can diminish the location quality of the crowdsensing results. Therefore, it is essential to control the location quality degradation while preserving users' location privacy. Towards this goal, we consider the following optimization problem **Location Quality Degradation Minimization (LQDM)** in this paper: Given a set of n users' actual locations and an integer $k \leq n$, form a set of perturbed groups, denoted by $\mathcal{H}$, including all users to minimize the location quality degradation.

Note that in the literature, the summation of squared location errors (SSE) [26] has been used to measure data quality. In this paper, we use the location quality degradation, because even if the SSE is small, some large errors are still detrimental to the crowdsensing application. Whereas, a small location quality degradation guarantees that none of the errors exceeds this value. Although we do not focus on minimizing the SSE, extensive simulations show that our algorithm achieves a lower SSE, compared to an existing k-anonymity location privacy preserving algorithm.

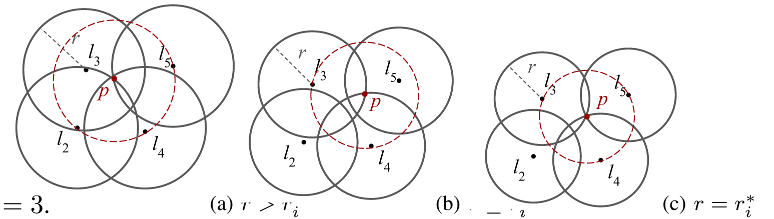
C. Geometric Problem Transformation

LQDM can be transformed into an equivalent geometric problem. Before the transformation, we introduce the following definition.

Let $\mathcal{P}$ denote a plane. For any two points $p \in \mathcal{P}$ and $q \in \mathcal{P}$, we use $||p, q||$ to denote the Euclidean distance between p and q. A disk centered at c of radius r is denoted by $D(c, r)$. We say $D(c, r)$ *covers* p, if $p \in D(c, r)$, i.e., $||p, c|| \leq r$. Let $B(c, r)$ denote the closed boundary of $D(c, r)$. Given a set $\mathcal{L}$ of n points, let $\mathcal{D}(\mathcal{L}, r)$ denote a set of disks of radius r centered at points in $\mathcal{L}$.

Definition 1 (k-enclosing Disk). *Let $\mathcal{L}$ be a set of n points on the plane $\mathcal{P}$. Given an integer $k \leq n$, a k-enclosing disk is a disk that covers at least k points in $\mathcal{L}$.*

The transformed **LQDM** problem is: Given a set $\mathcal{L}$ of n points on the plane $\mathcal{P}$ and an integer $k \leq n$, find a minimum r and a set of k-enclosing disks $\mathcal{D} = \{D(h_1, r), D(h_2, r), \ldots\}$, such that any $l_i \in \mathcal{L}$ is covered by at least one disk in $\mathcal{D}$.



(a) $r > r_i$
$d_{D(l_i,r)} = 4$

(b) $r = r_i$
$d_{D(l_i,r)} = 3$

(c) $r = r_i^*$
$d_{D(l_i,r)} = 3$

Figure 3: Comparison between r and r_i^* $(k = 3, i = 5)$

To solve these problems, we need the following definitions and claims from [8].

Definition 2 (Depth of a Point)**.** *Given a point $p \in \mathcal{P}$ and a disk set $\mathcal{D}(\mathcal{L}, r)$, the depth of p with respect to $\mathcal{D}(\mathcal{L}, r)$, denoted by $d_{\mathcal{D}(\mathcal{L},r)}(p)$, is the number of disks in $\mathcal{D}(\mathcal{L}, r)$ covering p.*

Figure 2 gives an example of calculating the depth of a point p.

Definition 3 (Depth of a Disk)**.** *Given a point $l_i \in \mathcal{L}$ and a disk set $\mathcal{D}(\mathcal{L}, r)$, the depth of $D(l_i, r)$, denoted by $d_{D(l_i,r)}$, is the maximum depth of all points $p \in D(l_i, r)$:*

$$d_{D(l_i,r)} = \max_{p \in D(l_i,r)} \{d_{\mathcal{D}(\mathcal{L},r)}(p)\}.$$

Claim 1. *Given two points $p, q \in \mathcal{P}$, $p \in D(q, r)$ if and only if $q \in D(p, r)$.*

Claim 2. *The depth of a point $p \in \mathcal{P}$ in $\mathcal{D}(\mathcal{L}, r)$ is the number of points in $\mathcal{L}$ covered by $D(p, r)$.*

In addition, we provide the following definition and lemma.

Definition 4 (Critical Radius)**.** *Given any $l_i \in \mathcal{L}$, a radius r is a critical radius, if $d_{D(l_i,r)}$ decreases, when r is decreased by an arbitrarily small amount.*

Lemma 1. *A point $l_i \in \mathcal{L}$ is covered by a k-enclosing disk $D(p, r)$ for some $p \in \mathcal{P}$ and some r, if and only if there exists $p \in D(l_i, r)$, such that $d_{\mathcal{D}(\mathcal{L},r)}(p) \geq k$.*

Proof. Suppose $l_i \in \mathcal{L}$ is covered by a k-enclosing disk $D(p, r)$. Then we know $p \in D(l_i, r)$ by Claim 1, and $d_{\mathcal{D}(\mathcal{L},r)}(p) \geq k$. Suppose there exists $p \in D(l_i, r)$, such that $d_{\mathcal{D}(\mathcal{L},r)}(p) \geq k$. We can know at least k points in $\mathcal{L}$ are covered by $D(p, r)$, according to Claim 2. Thus $D(p, r)$ is a k-enclosing disk, and we know $l_i \in D(p, r)$ by Claim 1. This completes the proof. ■

At last, we have three geometrical facts as follows.

1) The point on $D(l_i, r)$ with maximum depth must be an intersection point on $B(l_i, r)$, if $B(l_i, r)$ intersects with the boundary of any other disk in $\mathcal{D}(\mathcal{L}, r)$. Then we only focus on the intersection points on $B(l_i, r)$ for computing $d_{D(l_i,r)}$.

2) Given any $l_i \in \mathcal{L}$, let r_i^* denote the minimum radius r, such that $d_{D(l_i,r)} \geq k$. We can locate r_i^* within a feasible range of r using the following criteria:
   - $d_{D(l_i,r)} < k \to r < r_i^*$;
   - $d_{D(l_i,r)} > k \to r > r_i^*$;
   - $d_{D(l_i,r)} = k \to r \geq r_i^*$.

   Figure 3 gives a example of how the value of $d_{D(l_i,r)}$ can derive the comparison between r and r_i^*.
3) A radius r can be a critical radius only if $B(l_i, r)$ is tangent to $B(l_j, r)$, or $B(l_i, r)$ is concurrent with $B(l_j, r)$ and $B(l_k, r)$, where $l_i, l_j, l_k \in \mathcal{L}$. In other words, a critical radius is either $\frac{1}{2}||l_i, l_j||$, denoted by r_{ij}, or a circumradius of a triangle with l_i, l_j and l_k as the vertices, denoted by r_{ijk}.

The main notations are summarized in Table 1.

Table 1: Main notations

Notation	Meaning
$\mathcal{U}$	a set of users $\{1, 2, \ldots, n\}$
$(h, \mathcal{S})$	a perturbed group, where all users in $\mathcal{S}$ are tagged with h
$\mathcal{H}$	a set of perturbed groups
$\mathcal{P}$	a plane
l_i	the actual location (a point on $\mathcal{P}$) of user i
$\mathcal{L}$	the set of actual locations (points on $\mathcal{P}$) of users in $\mathcal{U}$
$D(p, r)$	the disk of radius r, centered at $p \in \mathcal{P}$
$\mathcal{D}(\mathcal{L}, r)$	the set of disks with radius r, centered at points in $\mathcal{L}$
$d_{\mathcal{D}(\mathcal{L},r)}(p)$	the depth of a point p with respect to $\mathcal{D}(\mathcal{L}, r)$
$d_{D(l_i,r)}$	the depth of a disk $D(l_i, r)$
$B(p, r)$	the closed boundary of $D(p, r)$
r_i^*	the minimum radius of a k-enclosing disk to cover l_i
p_i^*	the center of the smallest k-enclosing disk to cover l_i

IV. An Optimal Algorithm for **LQDM**

In this section, we present an efficient optimal algorithm OLoQ for the **LQDM** problem.

A. Overview

Since r_i^* is the minimum radius, such that l_i is covered by a k-enclosing disk, the minimum radius in the optimal solution to the **LQDM** problem equals $\max_{l_i \in \mathcal{L}} r_i^*$, denoted by r^*. Thus the **LQDM** problem boils down to finding r_i^* for each $l_i \in \mathcal{L}$. Based on Fact 2) in Section III-C, it is necessary to determine a range in order to locate r_i^*. To locate the exact value of r_i^*, we need to discretize its range. By the definition of critical radius, r_i^* must be a critical radius of l_i. Thus we focus on critical radii and conduct a binary search among them for locating r_i^*. According to Fact 3), a critical radius of l_i can only be r_{ij} or r_{ijk}, where $l_j, l_k \in \mathcal{L}, i \neq j \neq k$. Once r_i^* is located, we find the point of maximum depth on $D(l_i, r_i^*)$, denoted by p_i^*. Then a set of k-enclosing disks $\mathcal{D} = \{D(p_i^*, r^*) \mid l_i \in \mathcal{L}\}$ can cover all $l_i \in \mathcal{L}$. However, not all disks in $\mathcal{D}$ are necessary. Thus we select a minimal $\mathcal{D}^* \subseteq \mathcal{D}$ covering all points in $\mathcal{L}$. The centers of the selected disks are the perturbed locations.

B. Algorithm Design

OLoQ includes one key algorithm to find the smallest k-enclosing disk covering $l_i \in \mathcal{L}$, illustrated in Algorithm 1.

In Algorithm 1, we narrow the range where r_i^* can lie and locate r_i^*. To narrow the range where r_i^* can lie, we collect the $n-1$ values of r_{ij} and sort them in a non-decreasing order. Note that each r_{ij} is corresponding to a tangent point p_{ij} of $B(l_i, r_{ij})$ and $B(l_j, r_{ij})$, which is the midpoint of line $l_i l_j$. Then the range can be narrowed to $(\underline{r}_{ij}, \bar{r}_{ij}]$.

Then we collect $\frac{(n-1)(n-2)}{2}$ values of r_{ijk} and only keep the values of r_{ijk} within the range $(\underline{r}_{ij}, \bar{r}_{ij}]$. If there is no r_{ijk} within this range, then r_i^* is $\bar{r}_{ij}$ and its corresponding p_i^* is $\bar{p}_{ij}$. Otherwise, we sort the values of r_{ijk} within the range $(\underline{r}_{ij}, \bar{r}_{ij}]$ in a non-decreasing order. Note that each r_{ijk} is corresponding to a point p_{ijk}, which is the circumcenter of the triangle with l_i, l_j and l_k as the vertices. Using binary search, we further restrict the range to $(\underline{r}_{ijk}, \bar{r}_{ijk}]$, which is the smallest range such that $d_{D(l_i,\underline{r}_{ijk})} < k$ and $d_{D(l_i,\bar{r}_{ijk})} \geq k$. Therefore r_i^* is $\bar{r}_{ijk}$, and p_i^* is the $\bar{p}_{ijk}$ corresponding to $\bar{r}_{ijk}$.

At the end, Algorithm 1 outputs (r_i^*, p_i^*), which forms the smallest k-enclosing disk $D(r_i^*, p_i^*)$ that covers l_i. We shall run Algorithm 1 for each $l_i \in \mathcal{L}$. Then the minimum radius in the optimal solution to the **LQDM** problem is $\max_{l_i \in \mathcal{L}} r_i^*$.

Algorithm 1: Find-k-enclosing-Disk($l_i, k, \mathcal{L}$)

```
Sort all values in $\{r_{ij} \mid l_j \in \mathcal{L} \backslash \{l_i\}\}$ in a non-decreasing
  order and obtain a sorted list $R_{ij}$;
Run a binary search in $R_{ij}$ to find two consecutive
  values of $r_{ij}$, denoted by $\underline{r}_{ij}$ and $\bar{r}_{ij}$, such that
  $d_{D(l_i,\underline{r}_{ijk})} < k$ and $d_{D(l_i,\bar{r}_{ij})} \geq k$;
Sort all values in $\{r_{ijk} \mid r_{ijk} \in (\underline{r}_{ij}, \bar{r}_{ij}], l_j, l_k \in \mathcal{L} \backslash \{l_i\}\}$
  in a non-decreasing order and obtain a sorted list $R_{ijk}$;
if $R_{ijk} = \emptyset$ then
    $r_i^* \leftarrow \bar{r}_{ij}$; $p_i^* \leftarrow \bar{p}_{ij}$ ;
else
    Run a binary search in $R_{ijk}$ to find two consecutive
      values of $r_{ijk}$, denoted by $\underline{r}_{ijk}$ and $\bar{r}_{ijk}$, such that
      $d_{D(l_i,\underline{r}_{ijk})} < k$ and $d_{D(l_i,\bar{r}_{ijk})} \geq k$;
    $r_i^* \leftarrow \bar{r}_{ijk}$; $p_i^* \leftarrow \bar{p}_{ijk}$;
end
return $(r_i^*, p_i^*)$
```

Next, we generate a set of k-enclosing disks $\mathcal{D}^* = \{D(h_1, r^*), D(h_2, r^*), \ldots\}$, such that any $l_i \in \mathcal{L}$ is covered by at least one disk in $\mathcal{D}^*$. By the previous steps, we can obtain a set of k-enclosing disks $\mathcal{D} = \{D(p_1^*, r^*), D(p_1^*, r^*), \ldots, D(p_n^*, r^*)\}$ covering all points in $\mathcal{L}$. However, not all of them are necessary. So we design Algorithm 2 to select a minimal $\mathcal{D}^* \subseteq \mathcal{D}$ covering all points in $\mathcal{L}$. The idea is to select disks iteratively. In each iteration we select a disk covering as many points as possible. Thus we sort n values of r_i^* for all $l_i \in \mathcal{L}$ in a non-increasing order and select disks sequentially according to the sorted list. If l_i has not been covered, we add $D(p_i^*, r^*)$ to $\mathcal{D}^*$. For all users whose actual locations are covered by $D(p_i^*, r^*)$, we form a perturbed group with p_i^* as their perturbed location.

We now analyze the running time of OLoQ. The time complexity of Algorithm 1 is dominated by two times of

Algorithm 2: OLoQ $(\mathcal{L}, k)$

```
$\mathcal{H} \leftarrow \emptyset$; $\mathcal{D}^* \leftarrow \emptyset$;
for $l_i \in \mathcal{L}$ do
    $r_i^* \leftarrow$ Find-$k$-enclosing-Disk( $l_i, k, \mathcal{L}$ );
$r^* \leftarrow \max_{l_i \in \mathcal{L}} r_i^*$;
Sort points in $\mathcal{L}$ based on $r_i^*$ in a non-increasing order
  and obtain a sorted list $L$;
for $l_i \in L$ do
    if $l_i$ is uncovered by $\mathcal{D}^*$ then
        $\mathcal{D}^* \leftarrow \mathcal{D}^* \cup \{D(p_i^*, r^*)\}$;
        $\mathcal{H} \leftarrow \mathcal{H} \cup \{(p_i^*, \{i \mid l_i \in D(p_i^*, r^*)\})\}$ ;
    end
end
return $(\mathcal{H}, r^*)$
```

binary search. Each binary search needs to calculate disk depth, which takes $O(n^2)$ time. Hence the time complexity of Algorithm 1 is $O(n^2 \log n)$. The time complexity of Algorithm 2 is dominated by computing r_i^* for each $l_i \in \mathcal{L}$. Therefore, the overall time complexity is $O(n^3 \log n)$.

C. Algorithm Analysis

The following theorem guarantees OLoQ's optimality.

Theorem 1. OLoQ *returns an optimal solution to the* **LQDM** *problem.*

Proof. We first prove that each user i is tagged with the same perturbed location as at least $k-1$ other users and then prove that r^* is the minimum location quality degradation.

For each $l_i \in \mathcal{L}$, it guarantees that $d_{D(l_i, r_i^*)} \geq k$, based on Lines 2 and 7 in Algorithm 1. Since p_i^* is the point with maximum depth on $D(l_i, r_i^*)$, we have $d_{\mathcal{D}(\mathcal{L}, r_i^*)}(p_i^*) \geq k$. By Claim 2, $D(p_i^*, r_i^*)$ covers at least k points in $\mathcal{L}$. By Claim 1, we have $l_i \in D(p_i^*, r_i^*)$. With $r^* \geq r_i^*$, we know that $D(p_i^*, r^*)$ covers at least k points in $\mathcal{L}$ and $l_i \in D(p_i^*, r^*)$, as well. Thus there are at least k users in each perturbed group $(p_i^*, \{i \mid l_i \in D(p_i^*, r^*)\})$. Therefore each user i is tagged with the same perturbed location with at least $k-1$ other users.

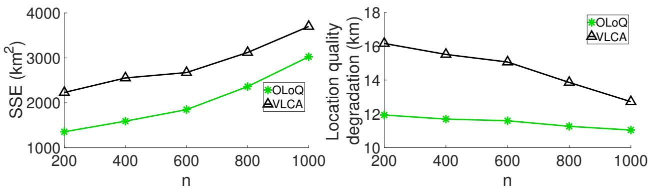
We know there must exist at least one point $l_i \in \mathcal{L}$, such that $r_i^* = r^*$. In addition, it guarantees that r_i^* is the minimum radius, such that $d_{D(l_i, r_i^*)} \geq k$, based on Lines 2 and 7 in Algorithm 1. We learned from the above proof that r_i^* is therefore the minimum radius, such that user i is tagged with the same perturbed location as at least $k-1$ other users. Therefore r^* is the minimum radius to satisfy k-anonymity for all users. ■

V. Performance Evaluation

In this section, we evaluate the performance of OLoQ by comparing it with an existing k-anonymity location privacy preserving algorithm [32].

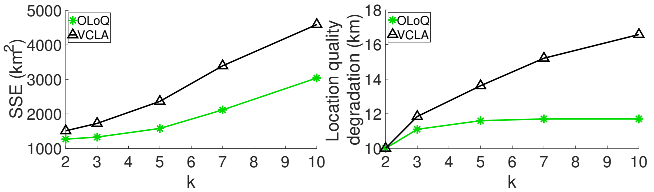
A. Evaluation Setup

As we surveyed in Section II, there is no existing algorithm that aims to minimize the location quality degradation. The most related work for k-anonymity location privacy is VCLA [32], which is a heuristic algorithm that uses the microaggregation approach to obtain anonymized locations and aims to minimize the summation of squared errors (SSE).



(a) SSE (b) Location quality degradation (km)

Figure 4: Impact of n on OLoQ and VCLA.



(a) SSE (b) Location quality degradation (km)

Figure 5: Impact of k on OLoQ and VCLA.

We use the CRAWDAD dataset roma/taxi [2, 4] for our simulations. The dataset contains the mobility traces of approximately 320 taxis collected over 30 days in Rome, Italy. Each mobility trace consists of a sequence of GPS coordinates collected roughly every seven seconds along with corresponding timestamps.

B. Performance Metrics

We are interested in the following performance metrics.

- *SSE:* Suppose a point set $\mathcal{L}$ is divided into m groups. The sum of squared errors of perturbed group j is defined as:

  $$sse_j = \sum_{p=1}^{n_j} [(x_{jp} - \bar{x}_j)^2 + (y_{jp} - \bar{y}_j)^2]$$

  where n_j is the number of users in j-th group satisfying $n_j \geq k$, (x_{jp}, y_{jp}) is the location of the pth user with $(\bar{x}_j, \bar{y}_j)$ the perturbed location of j-th group. The SSE is the sum of sse_j:

  $$SSE = \sum_{j=1}^{m} sse_j = \sum_{j=1}^{m} \sum_{p=1}^{n_j} [(x_{jp} - \bar{x}_j)^2 + (y_{jp} - \bar{y}_j)^2],$$

  where SSE describes the overall group homogeneity after group formation. When nearby points are grouped together, SSE will be small and the groups are more homogeneous.
- *Location quality degradation*: The location quality degradation is the maximum distance between users' actual locations and their corresponding perturbed locations.

In our evaluation, we show the impact of the number of users (n) and k on OLoQ and VCLA in terms of SSE and location quality degradation. For the impact of n, we vary it from 200 to 1000 with an increment of 200, fixing $k = 5$. For the impact of k, we set it to be $2, 3, 5, 7, 10$, fixing $n = 400$. All results are averaged over 100 independent runs.

C. Evaluation Results and Analysis

Figure 4 shows the impact of n. Figure 4(a) shows that OLoQ can always introduce lower SSE, which is essential to obtain accurate sensing data. Besides, SSE increases with n, because sparser location distribution leads to larger errors. In Figure 4(b), the location quality degradations of OLoQ and VCLA decrease with n. We also observe that OLoQ outperforms VCLA, especially with fewer users, because OLoQ minimizes the location quality degradation, while VCLA heuristically aggregate locations by choosing the farthest point and then aggregating the nearest points to it.

Figure 5 shows the impact of k. Figure 5(a) illustrates that the SSE gradually increases with more stringent privacy protection. To protect more users' locations in one perturbed group, it is inevitable to diminish the location quality to some degree. OLoQ has a lower SSE, because it minimizes the location quality degradation. From Figure 5(b), we observe that, OLoQ outputs perturbed groups with the minimum location quality degradation and significantly outperforms VCLA. The common trend is the location quality degradation increases with more stringent privacy protection .

VI. Conclusion and Future Work

In this paper, we consider the location quality optimization problem in location privacy preserving crowdsensing, Extensive simulations show that OLoQ not only achieves the minimum location quality degradation, but also outperforms an existing algorithm in terms of SSE.

References

[1] B. Agir, T. G. Papaioannou, R. Narendula, K. Aberer, and J.-P. Hubaux, "User-side adaptive protection of location privacy in participatory sensing," *GeoInformatica*, vol. 18, no. 1, pp. 165–191, 2014.

[2] R. Amici, M. Bonola, L. Bracciale, A. Rabuffi, P. Loreti, and G. Bianchi, "Performance assessment of an epidemic protocol in vanet using real traces," *Procedia Comput. Sci.*, vol. 40, pp. 92–99, 2014.

[3] A. R. Beresford and F. Stajano, "Mix zones: User privacy in location-aware services," in *Proc. of PerCom Workshops*, 2004, pp. 127–131.

[4] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "Crawdad data set roma/taxi (v. 2014-07-17)," http://crawdad.org/roma/taxi/20140717/.

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

[6] J. Domingo-Ferrer and J. M. Mateo-Sanz, "Practical data-oriented microaggregation for statistical disclosure control," vol. 14, no. 1, pp. 189–201, 2002.

[7] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," in *International Conference on Pervasive Computing*. Springer, 2005, pp. 152–170.

[8] A. Efrat, M. Sharir, and A. Ziv, "Computing the smallest k-enclosing circle and related problems," *Comput. Geometry*, vol. 4, no. 3, pp. 119–136, 1994.

[9] S. Gao, J. Ma, W. Shi, and G. Zhan, "Towards location and trajectory privacy protection in participatory sensing," in *Proc. of MobiCASE*, 2011, pp. 381–386.

[10] G. Ghinita, "Privacy for location-based services," *Synthesis Lectures on Information, Security, Privacy, & Trust*, vol. 4, no. 1, pp. 1–85, 2013.

[11] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private queries in location based services: Anonymizers are not necessary," in *Proc. of SIGMOD*, 2008, pp. 121–132.

[12] X. Gong, X. Chen, K. Xing, D.-H. Shin, M. Zhang, and J. Zhang, "From social group utility maximization to personalized location privacy in mobile networks," *IEEE/ACM Tans. Networking*, 2017.

[13] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," *Mobile Sensing*, pp. 1–5, 2012.

[14] C. V. N. Index, "Cisco visual networking index: Global mobile data traffic forecast update, 2014–2019," *Tech. Rep*, 2015.

[15] X. Jin and Y. Zhang, "Privacy-preserving crowdsourced spectrum sensing," in *Proc. of INFOCOM*, 2016, pp. 1–9.

[16] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proc. of ICPS*, 2005, pp. 88–97.

[17] G. Kokolakis and D. Fouskakis, "Importance partitioning in microaggregation," *Comput. Stats. & Data Anal.*, vol. 53, no. 7, pp. 2439–2445, 2009.

[18] A. Krause, E. Horvitz, A. Kansal, and F. Zhao, "Toward community sensing," in *Proc. of IPSN*, 2008, pp. 481–492.

[19] S. Li, H. Zhu, Z. Gao, X. Guan, K. Xing, and X. Shen, "Location privacy preservation in collaborative spectrum sensing," in *Proc. of INFOCOM*, 2012, pp. 729–737.

[20] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *Proc. of SenSys*, 2008, pp. 323–336.

[21] M. Murshed, A. Iqbal, T. Sabrina, and K. M. Alam, "A subset coding based k-anonymization technique to trade-off location privacy and data integrity in participatory sensing systems," in *Proc. of NCA*, 2011, pp. 107–114.

[22] Y. Qiu and M. Ma, "A privacy-preserving proximity testing for location-based services," in *Proc. of GLOBECOM*. IEEE, 2018, pp. 1–6.

[23] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-Phone: An end-to-end participatory urban noise mapping system," in *Proc. of IPSN*, 2010, pp. 105–116.

[24] I. Rodhe, C. Rohner, and E. C.-H. Ngai, "On location privacy and quality of information in participatory sensing," in *Proc. of Q2SWinet*, 2012, pp. 55–62.

[25] R. Shokri, G. Theodorakopoulos, P. Papadimitratos, E. Kazemi, and J.-P. Hubaux, "Hiding in the mobile crowd: Location privacy through collaboration," *IEEE Trans. Dependable Secure Comput*, vol. 11, no. 3, pp. 266–279, 2014.

[26] A. Solanas, A. Martinez-Balleste, and J. Domingo-Ferrer, "V-mdav: a multivariate microaggregation with variable group size," in *17th COMPSTAT Symposium of the IASC, Rome*, 2006, pp. 917–925.

[27] L. Sweeney, "k-anonymity: A model for protecting privacy," *IJUFKS*, vol. 10, no. 05, pp. 557–570, 2002.

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

[29] I. J. Vergara-Laurens and M. A. Labrador, "Preserving privacy while reducing power consumption and information loss in LBS and participatory sensing applications," in *Proc. of GLOBECOM Workshops*, 2011, pp. 1247–1252.

[30] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for k-anonymous location privacy in participatory sensing," in *Proc. of INFOCOM*, 2012, pp. 2399–2407.

[31] W. Wang and Q. Zhang, "Privacy-preserving collaborative spectrum sensing with multiple service providers," vol. 14, no. 2, pp. 1011–1019, 2015.

[32] X. Wang, Z. Liu, X. Tian, X. Gan, Y. Guan, and X. Wang, "Incentivizing crowdsensing with location-privacy preserving," vol. 16, no. 10, pp. 6940–6952, 2017.

[33] Z. Xiao, X. Meng, and J. Xu, "Quality aware privacy protection for location-based services," in *Proc. of DASFAA*, 2007, pp. 434–446.

[34] T. Xu and Y. Cai, "Location anonymity in continuous location-based services," in *Proc. of GIS*. ACM, 2007, pp. 39:1–39:8.

[35] F. Zhang, L. He, W. He, and X. Liu, "Data perturbation with state-dependent noise for participatory sensing," in *Proc. of INFOCOM*, 2012, pp. 2246–2254.

[36] X. Zheng, Z. Cai, J. Li, and H. Gao, "Location-privacy-aware review publication mechanism for local business service systems," in *Proc. of INFOCOM*, 2017, pp. 289–297.