Discovering Truth in Mobile Crowdsensing with Differential Location Privacy

Tongqing Zhou*, Zhiping Cai*, Jingshu Su
College of Computer, National University of Defense Technology, Changsha, China
*Corresponding authors

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001292

***Abstract*—The combination of Mobile Crowdsensing (MCS) and truth discovery has benefited the ubiquitous monitoring and analysis of the physical world. To address the concerns alongside user data collection, the literature has partially studied data privacy protection for truth discovery. Yet, the threats of location leakage remain overlooked in such contexts. For joint accommodating privacy protection and truth elaboration, we propose to leverage differential privacy for distributed user location obfuscation and explore spatial correlation for corresponding observation's value calibration. We form this process into a truth estimation deviation minimization problem under differential privacy and obfuscation requirements. By theoretically transforming it into probabilistic calibration residual optimization, the problem can be solved via linear programming. Evaluation on real-world temperature and humid sensing data shows its effectiveness on providing significant location distortion distance and practically acceptable time consumption. Results also reveal an up to 53% truth discovery accuracy improvement compared to the SCP baseline.**

***Index Terms*—Mobile crowdsensing, truth discovery, location privacy, differential privacy.**

I. Introduction

The rich sensing equipment and pervasive communication capacity of mobile devices have promoted the recent development of Mobile crowdsensing paradigm [1]. With the participatory or opportunistic engagement of mobile users, MCS realizes large-scale real-world data sensing and collection, which has supported a broad spectrum of applications, including environmental monitoring, city sensing, and smart transportation. For example, the temperature of smartphone batteries is used in [2] as indirect thermometers of the operating environment for ubiquitous temperature monitoring.

Due to the noisy nature of user-contributed observations, truth discovery is considered an essential step of MCS tasks for resolving possibly conflicted reports [3]. Truth discovery approaches (e.g., CRH [4]) iteratively perform weighted aggregation on multiple observations and evaluate user weights (a.k.a., reliability) towards refined truth estimations. Such a process is carried out in different spatio-temporal regions independently for constant insights parsing (§ III).

User privacy is a general concern on the data-driven MCS applications and has been widely investigated [3] [5] [6] [7] for both user benefits and regulation purpose. In particular, the designs in terms of truth discovery rely on either secure computation or noise injection. Wherein, homomorphic encryption [7] and Gaussian perturbation [3] are adopted to protect user data and truth estimation performance simultaneously. Although effective in protecting data privacy, we point out that users' location privacy is seldom considered and not carefully taken care of during truth discovery on MCS. Simply encrypting locations with data protection techniques would ruin their basic property as spatial indices, because, unlike their data counterparts, locations should be explicitly and discretely represented for dividing observations into grids.

978-1-6654-3540-6/22 © 2022 IEEE

Existing location privacy-preserving methods for MCS are also ill-suited in the context of truth discovery. Along this line of work, recent research efforts have been primarily devoted to tailoring differential privacy to different sensing quality indicators, such as minimizing traveling distance [8], enlarging sensing coverage [9], and maintaining data recovery accuracy [10]. However, compared with these objectives, truth discovery is more sensitive to location perturbation and would experience significant accuracy degradation, or even wrong conclusions when the estimations are assigned to wrong locations. As far as we know, [11] is the only work that investigates location-preserving truth discovery for MCS. Yet, their work is designed on an infeasible assumption that there exists a trusted fog node that can pre-aggregate observations from multiple grids to hide their raw locations. The void of location obfuscation strategy for crowdsensing truth discovery motivates our work.

This paper explores the joint accommodation of location privacy protection and effective truth discovery in MCS. Basically, we propose to attain indistinguishable location obfuscation with differential privacy on behalf of the users (i.e., privacy requirement), and refer to the spatial correlation of different sensing units for dynamic observation calibration towards MCS server's benefits (§ IV-A). To achieve both, we formulate the truth deviation minimization problem with privacy constraints and prove its equivalent form as optimizing incurred residuals (§ IV-B). The main contributions of this work are:

- We present a probabilistic location obfuscation design for privacy-preserving truth discovery in crowdsensing based on differential privacy [10] and spatial correlation exploration.
- We formally model the location obfuscation problem as minimizing global truth estimation deviation under differential privacy constraints. We reduce the problem to inter-grids regression residual minimization and analyze the computation complexity.

- We evaluate the proposal based on real-world environment sensing datasets. The results demonstrate its effectiveness w.r.t. location distortion and acceptable time overhead and show truth discovery accuracy advantages over the baseline of up to 53%.

II. Related Work

A. Preserving Location Privacy in MCS

Existing location privacy protection techniques in MCS can be roughly categorized as: obfuscation-based approaches that perturb locations to insensitive places, cloaking-based approaches that hide the precise location behind a coarse area, and encryption approaches that encrypt location together with the data [12]. Among them, obfuscation is the most popular strategy recently for the introduction of differential privacy in providing theoretical privacy guarantees.

The general design principle in obfuscation-based approaches is to perturb location and maintain sensing quality simultaneously. For example, the quality of user recruitment is considered in [8], where users manage to minimize the expected traveling distances after changing their location tags. Azhar et al. [9] investigate the sensing coverage maximization problem under differential-obfuscated locations. In [1] and [10], authors attempt to reduce data recovery error during location obfuscation.

However, these approaches are not effective in truth discovery tasks on MCS as estimating truth on observations from arbitrary locations would degrade the estimation accuracy.

B. Privacy-preserving Truth Discovery

The concerns and protection of data privacy in truth discovery are also widely studied, either based on secure multi-party computation or noise injection. For the former category, Zheng et al. [7] design a secure sum protocol for privacy-aware truth discovery in MCS, which is criticized for requiring frequent interaction between users and the server. Furthermore, the privacy of individual weights (reliability) is considered in [5], and a homomorphic cryptosystem is used to protect reliability and data security. To avoid the intensive secure computation in these approaches, the latter category [3] [6] introduces dedicated perturbation (e.g., differential privacy noise) on the data for efficient data protection.

Unfortunately, these approaches fail to handle users' location privacy during MCS tasks. Truth discovery of MCS is conducted on the location granularity, i.e., performed in each grid independently. Different from data protection, the location information cannot be disclosed even after the aggregation, as aggregation can hide the original data, but the location property is inherited in this process.

III. Preliminaries

This section depicts the workflow of MCS and truth discovery. We list the frequently used notations in Table I.

TABLE I
Frequently Used Notations.

Notations	Description
l,l^*	Original location and its obfuscated location
$\mathbf{V}_l$,$\mathbf{V}_{l^*}$	The set of observations at location l and l^*
N_l,N_{l^*}	# users at location l and l^*, $N_l = \|V_l\|$
$\mathbf{L}$	The set of all interested locations
v_l,v_{l^*}	The truth estimation for location l and l^*
$P(l^*\|l)$	The probability of obfuscating location l to l^*
ω_k	The weight of the k-th user at specific location
$\mathbf{V}(k)$	The set of observations of the k-th user

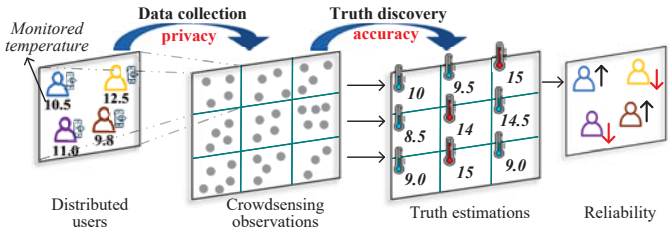


Fig. 1. Real-world truth discovery based on MCS systems. Temperature monitoring is used as an example here and multiple co-located observations are aggregated for an ultimate estimation in each grid.

A. MCS System Overview

MCS has been adopted in applications, such as environmental monitoring, city sensing, and smart transportation for distributed phenomena sensing and data collection of interested regions [2] [13]. A typical MCS task consists of a user-oriented data collection phase and a server-centric truth discovery phase, as shown in Fig. 1.

During data collection, the targeted region is divided into grids and users submit observations to each grid according to their locations. Exposing locations during this phase constitutes significant privacy concerns for users, which may make them reluctant to participate in the sensing tasks. During truth discovery, MCS server attempts to elaborate on the happening truth of a grid based on multiple observations from different users. The accuracy of the deduced truth is of utmost importance for the server as deviated estimations would mislead decision-making and application quality.

B. Truth Discovery Formulation

Given multiple noisy observations $\mathbf{V}_l$ at location l, truth discovery approaches can infer user weights (reliability) ω_k and estimate truth v_l via weighted observation aggregation. For example, in the temperature monitoring task of Fig. 1, truth estimation (i.e., 10) is obtained based on the observations of four users in the upper left grid, and the reliability of each user is in turn evaluated with the estimation.

Specifically, it initially assigns random weights for users and iteratively performs data aggregation and weight update towards convergence [5].

Aggregation. Given the current user weights $\{\omega_k\}_{k=1}^{N_l}$ of location l, the estimation of l is obtained by weighted aggre-

gating the observations:

$$v_l = \frac{\sum_{k=1}^{N_l} \omega_k \cdot \mathbf{V}_l(k)}{\sum_{k=1}^{N_l} \omega_k}. \quad (1)$$

Obviously, the estimation leans more on those reliability users with higher weights.

Weight update. Then the newest estimations are regarded as the ground truth. User weights can thus be updated by measuring the distance between their observations and the current truth. In particular, we use the CRH model [4] here:

$$\omega_k = -log(\frac{d(\mathbf{V}_l(k), v_l)}{\sum_{k'=1}^{N_l} d(\mathbf{V}_l(k'), v_l)}). \quad (2)$$

As a monotonically decreasing function, CRH assigns users with small differences to the truth estimation a high weight. These two steps are repeated until the estimation between two iterations is smaller than a threshold (set to 0.1 in § V).

Note that a user can temporally report multiple observations in different grids, given that it may move in the target region. The weights assessment in Equ. 2 is based on observations in a certain time interval (according to the time granularity requirement of truth discovery), during which we assume a user can only make observations in one grid. For a successive sensing and estimation process, we will assign user weights according to its reliability evaluated in the previous rounds.

IV. Optimizing Truth Discovery under Location Obfuscation

In this section, we first introduce our privacy-preserving design for truth discovery on MCS, and then formulate the obfuscation problem with a discussion on its solution and complexity.

A. Our Privacy-preserving Design

As shown in Fig. 2, we propose to protect user location privacy by obfuscating it according to differential privacy [14] and maintain truth discovery quality with data calibration.

1) Obfuscation for differential privacy: We follow the adversary assumption of [14] that an adversary has side information of location distribution $P(l)$ for inferring user location (probabilistic model). By observing a user's obfuscated location l^*, s/he can induce its posterior distribution with $P(l|l^*) = \frac{P(l^*|l)\cdot P(l)}{\sum_{l' \in \mathbf{L}} P(l^*|l')\cdot P(l')}$.

Intuitively, if an obfuscated location has an indistinguishable probability of being mapped from the original location and any other locations, then the adversary cannot learn more about the posterior location with such an obfuscation. Formally, this requires the obfuscation to satisfy:

$$P(l^*|l) \leq e^{\epsilon} \cdot P(l^*|l') \ \forall l, l^*, l' \in \mathbf{L}, \quad (3)$$

where ϵ denotes the differential privacy factor and is practically set by the privacy budget. A smaller ϵ indicates a stricter privacy requirement that the differences between obfuscation probabilities of different locations should be more similar.

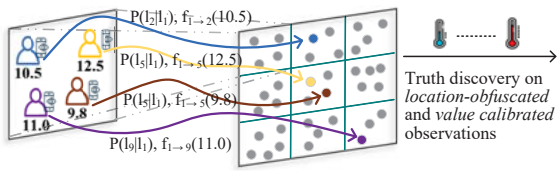


Fig. 2. Differential privacy-based location obfuscation and linear regression-based data adjustment.

2) Calibration for estimation quality: Discovering the truth of l^* with observations of l inevitably degrades the estimation accuracy. For example, in Fig. 2, the 4 users' observations are spatially mixed with reports that are very likely not sensed in the corresponding grid. We can relieve this tension by exploring spatial correlation [1] among the data in different grids. Following the assumption in [10], we propose to learn the correlation by performing linear regression on some observations of each location pair (l, l^*) to obtain $v_{l\to l^*} = f^{reg}_{l\to l^*}(v_l)$. In this way, a user chooses its obfuscated location according to $P(l^*|l)$, and meanwhile, calibrates its observation with f^{reg} to yield its report $< l^*, v_{l\to l^*} >$. Note that linear regression is performed in a centralized way by the MCS platform periodically, which is practically feasible as all the distributed observations are gathered to it.

As an empirical calibration, the fitted observations are still different from the field observations, which can be theoretically measured by the residual between regression and estimation:

$$R(l, l^*) = \sum |V_{l\to l'}(k) - V_{l^*}(k')|. \quad (4)$$

Combining it with the obfuscation probability, we can calculate the probabilistic residual between l and l^* as $P(l^*|l) \cdot R(l, l^*)$.

B. Building Obfuscation Matrix

We attempt to take a balance between privacy requirements and truth discovery accuracy that may be degraded by obfuscation. By setting the requirement as a constraint and the accuracy as an objective, we can formulate an optimization problem that tunes obfuscation strategy $P(l^*|l)$ towards the globally minimum truth estimation deviation.

$$\arg\min_{P} \ \frac{1}{|\mathbf{L}|^2} \sum_{l\in\mathbf{L}} \sum_{l^*\in\mathbf{L}} |v_{l\to l^*} - v_{l^*}| \quad (5)$$

$$s.t. \ P(l^*|l) \leq e^{\epsilon} \cdot P(l^*|l'), \ \forall l, l^*, l' \in \mathbf{L} \quad (6)$$

$$\sum_{l^*\in\mathbf{L}} P(l^*|l) = 1, \ \forall l \in \mathbf{L} \quad (7)$$

$$\sum_{l\in\mathbf{L}} P(l) \cdot P(l^*|l) = P(l), \ \forall l^* \in \mathbf{L} \quad (8)$$

$$P(l^*|l) \geq 0, \ \forall l, l^* \in \mathbf{L}, \quad (9)$$

where Equ. 6 is the differential privacy constraint under privacy budget ϵ, Equ. 7 and Equ. 9 are probability constraints, and Equ. 8 is used to assure the amount of observations in

each grid after obfuscation is probabilistically the same. By retaining observation distributions in the entire area, we intend to avoid biased truth estimation based on fewer observations in each grid.

Theorem IV.1. *The minimization of truth estimation differences between the original and obfuscated locations (Equ. 5) equals the minimization of the probabilistic residuals between all location pairs.*

We will use the Sum Reduction Inequality Lemma in [3] during the proof of Theorem IV.1. Briefly, it gives that if ω_k is a monotonically decreasing function of $\mathbf{V}(k)$, then $\frac{\sum_{k=1}^{N}\omega_k\cdot\mathbf{V}(k)}{\sum_{k=1}^{N}\omega_k} \leq \frac{\sum_{k=1}^{N}\mathbf{V}(k)}{N}$.

Proof. By replacing truth estimation in the optimization objective with Equ. 1, we have:

$$
\begin{aligned}
&\sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}|v_{l\to l^*}-v_{l^*}| \\
=&\sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}\Big|\frac{\sum_{k=1}^{N_l}\omega_k\cdot\mathbf{V}_{l\to l^*}(k)}{\sum_{k=1}^{N_l}\omega_k}-\frac{\sum_{k'=1}^{N_{l^*}}\omega_{k'}\cdot\mathbf{V}_{l^*}(k')}{\sum_{k'=1}^{N_{l^*}}\omega_{k'}}\Big| \\
=&\sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}\Big|\frac{\sum_{k=1}^{N_l}\sum_{k'=1}^{N_{l^*}}\omega_k\cdot\omega_{k'}\cdot(\mathbf{V}_{l\to l^*}(k)-\mathbf{V}_{l^*}(k'))}{\sum_{k=1}^{N_l}\sum_{k'=1}^{N_{l^*}}\omega_k\cdot\omega_{k'}}\Big| \\
\leq&\sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}\frac{\sum_{k=1}^{N_l}\sum_{k'=1}^{N_{l^*}}|\omega_k\cdot\omega_{k'}\cdot\mathbf{V}_{l\to l^*}(k)-\mathbf{V}_{l^*}(k')|}{\sum_{k=1}^{N_l}\sum_{k'=1}^{N_{l^*}}\omega_k\cdot\omega_{k'}} \\
\leq&\sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}\frac{\sum_{k=1}^{N_l}\sum_{k'=1}^{N_{l^*}}|\mathbf{V}_{l\to l^*}(k)-\mathbf{V}_{l^*}(k')|}{N_l\cdot N_{l^*}}\quad(Reduction) \\
=&\sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}\frac{1}{N_l}\cdot\sum_{k=1}^{N_l}\frac{1}{N_{l^*}}\cdot\sum_{k'=1}^{N_{l^*}}|\mathbf{V}_{l\to l^*}(k)-\mathbf{V}_{l^*}(k')|,
\end{aligned}
$$

which equals the expectation of the differences between inferred observations and raw observations of l^*, i.e., the expectation of residual between location pair (l, l^*). With the residual definition in Equ. 4, we have:

$$
\frac{1}{|\mathbf{L}|^2}\sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}|v_{l\to l^*}-v_{l^*}| \iff \sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}P(l^*|l)\cdot R(l,l^*) \tag{10}
$$

□

According to Theorem IV.1, the optimization problem is then transformed into:

$$
\begin{aligned}
&\arg\min_{P}\ \sum_{l\in\mathbf{L}}\sum_{l^*\in\mathbf{L}}P(l^*|l)\cdot R(l,l^*) \\
&s.t.\ Equs.\ 6,\ 7,\ 8,\ 9.
\end{aligned} \tag{11}
$$

Interestingly, we note that Equ. 11 has a similar form with the optimization problem in [10], which attempts to maximize data recovery accuracy during differential privacy amplification. The rationale is that we both refer to the spatial correlation of crowdsensing data for mitigating negative effects of privacy protection.

C. Solution and Computation Analysis

Equ. 11 is a typical linear optimization problem with $|\mathbf{L}|^2$ variables from the obfuscation matrix. Given these variables, we can parse $|\mathbf{L}|^3$ constraints from Equ. 6, $|\mathbf{L}|^2$ constraints from Equ. 9, and $|\mathbf{L}|$ constraints from Equs. 7 and 8, respectively. We propose to use the simplex of linear programming to solve this optimization problem, whose computation complexity is proportional to the number of constraints $|\mathbf{L}|^3+|\mathbf{L}|^2+2\times|\mathbf{L}|$.

The obfuscation matrix is supposed to be deduced and updated by the honest-but-curious server and distributed to all potential participants in the target region. The differential privacy property [14] ensures that even if an adversary obtains the matrix, s/he possesses bounded location inference capacity that is irrelevant to its prior knowledge.

Finally, for u_k at location l, it will report its observation $\mathbf{V}_k$ at an obfuscated location l^* with probability of $P(l^*|l)$. Then the MCS server performs truth discovery on the location-obfuscated data collection. Wherein, u_k's reliability is estimated as:

$$
-log\frac{|\widehat{v_{l^*}}-v_{l\to l^*}(k)|}{\sum_{p=1}^{N_{l^*}}|\widehat{v_{l^*}}-v_{l\to l^*}(p)|},
$$

which indicates that obfuscating data to lower-residual locations yields higher weight. This provides flexibility for users to balance privacy requirements and their recorded reliability.

V. Evaluation

We conduct experiments on real-world environment sensing datasets by testing the proposal's and baselines' truth discovery and location obfuscation performance.

A. Setup

Our test computer is equipped with Intel(R) Core i7-1160G7@1.20GHz, 16GB RAM, and Windows10. We use Python Pulp[1] to solve the linear programming problem (Equ. 11) for obtaining the obfuscation matrix under performance target and privacy constraints.

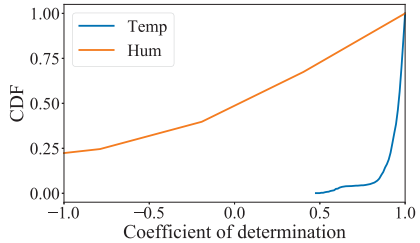
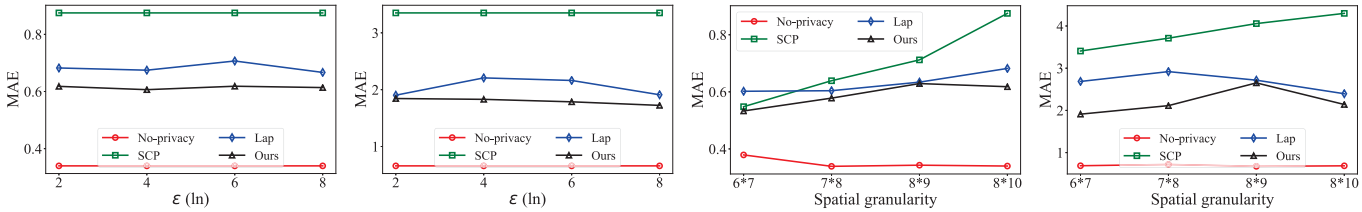


Fig. 3. CDF of the regression scores on two datasets.

[1]https://coin-or.github.io/pulp/



(a) varying ϵ on Temp. (b) varying ϵ on Hum. (c) varying granularity on Temp. (d) varying granularity on Hum.

Fig. 4. MAE on different datasets. When varying ϵ, spatial granularity is set to $10 * 8$; When varying the granularity, $\epsilon = ln2$.

1) Datasets: We use real-world sensing data from *SensorScope* [15] to evaluate our design. Specifically, the temperature and humidity data recorded in the EPFL campus ($260m \times 400m$) is adopted, termed as *Temp* and *Hum*. We take the data of one week with a sampling interval of 1 hour, which yields $|\mathbf{L}|*7*24$ truth discovery tasks, sufficient for our evaluation purpose. We divide them into equal-sized grids with different spatial granularity during evaluation (e.g., grid size $30m \times 50m$ will divide the area into $10*8$ grids). To simulate multiple users' reports for truth discovery, for each grid, we generate 6 users, whose data is obtained by adding Gaussian noise[2] (with deviations of 1 and 2 for *Temp* and *Hum*) on the corresponding field record in the grid.

2) Metrics: We measure the truth discovery performance with *Mean Absolute Error (MAE)* between the discovered truth v_l and the ground truth before adding noises. Lower *MAE* indicates better truth discovery accuracy.

Besides the differential privacy guarantee, we further test the caused location distortion of different obfuscation methods, calculated with the Euclidean Distance, to show the privacy-preserving performance.

3) Baselines: We compare the proposal's performance with Spatial Camouflage Participants (SCP) [1], the Laplacian noise-based method (Lap) [14], and the raw truth discovery method (No-privacy).

SCP is designed for location protection during data recovery in MCS. It permutes locations in the same row (column) to another row (column) simultaneously for preserving data correlation in compressive sensing. We generate random permutation for SCP in each sampling.

Lap refers to the typical location differential perturbation mechanism that adds Laplacian noise on the original location. Lap tends to perturb locations to nearby grids, so we set $P(l^*|l)$=$Norm_l(exp(-\epsilon \cdot \frac{d(l,l^*)}{d_{max}}))$ (d_{max} denotes the maximum distance between two locations), which is normalized for each original location. For comparison fairness, we also perform value calibration (§ IV-A2) for Lap after obfuscation.

B. Experimental Results

1) Regression performance: Recall that we propose to perform linear regression between each pair of grids for residual estimation during obfuscation. Here, we use the data of the 1st day for regression and measure the performance with coefficient of determination of the prediction, calculated as $1 - (\sum(\mathbf{V}_i - \mathbf{V}_j)^2 / \sum(\mathbf{V}_j - mean(\mathbf{V}_j))^2)$. A larger score ($\leq 1$) indicates a smaller residual, i.e., values in the tested two grids are similar. Specifically, we empirically neglect the intercept term during regression on *Hum* as its weak inter-grid correlation could easily over-fit this term towards extreme calibrations. As shown in the statistics of Fig. 3, most pairs of grids in *Temp* have relatively large scores, while the regression on *Hum* has nearly half pairs with scores smaller than 0 (arbitrarily worse). The modest regression performance of *Hum* causes larger residual and MAE than truth analysis on *Temp*, as we will show next.

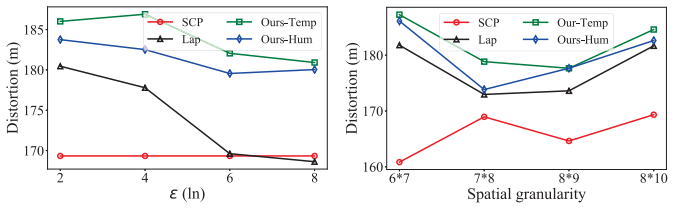
2) Truth discovery: We examine the proposal's truth discovery performance by comparing it with No-privacy and the SCP and Lap baselines. On one hand, Fig. 4(a) and Fig. 4(b) present the results under varying differential privacy level. No-privacy and SCP have static MAEs as not considering, so that not impacted by, the privacy budget/requirement. Our method outperforms the two baselines for all the test cases (outperforms SCP by a large margin). The advantage over Lap stems from the optimization for reducing residual deviation. Furthermore, we observe that our proposal causes larger MAE in truth discovery with a smaller privacy budget (i.e., ϵ) as stricter constraints will degrade the optimization value of truth discovery residuals. It effectively controls the truth deduction errors within certain bounds from No-privacy (i.e., 0.32 for Temp and 1.2 for Hum, both smaller than the deviation of the added Gaussian noises).

On the other hand, smaller grid size leads to denser grids with different observations, thus larger inter-grid data bias. As shown in Fig. 4(c) and Fig. 4(d), generally, the MAEs of both our proposal and SCP increase with growing spatial granularity, for having larger residual and differences between data of the original and obfuscated locations. Lap always attempts to obfuscate observations to nearby locations, so its performance becomes better for retaining data at nearer grids under denser granularity. Still, our proposal yields smaller MAE than both baselines because it is optimized to obfuscate locations of small residuals with high probability.

3) Location distortion: We test the averaged distortion levels between users' original location and obfuscated location under varying differential privacy level and spatial granularity. As shown in Fig. 5, both the baselines and our method incur significant distortion that can sufficiently camouflage

[2]It is reported that error of user observations follow standard Gaussian distribution [3].

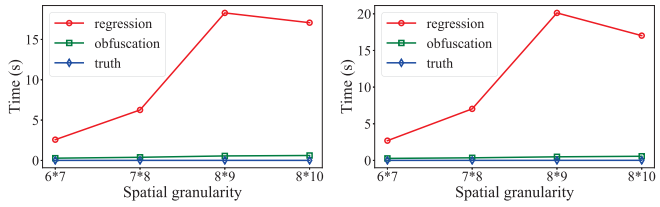
the original locations. In particular, our proposal yields larger distortions than the baselines on both datasets under all the test cases. Meanwhile, we can observe larger distortion gaps when the privacy level is smaller, as nearby grids may not satisfy the stricter privacy constraint, and thus are enforced to distant grids. As shown in Fig. 5(b), the distortion is sensitive to the granularity, which determines the unit distance in measuring the distortion. Note that with an average distortion of more than $100m$ in the $260*400$ region, the proposal still maintains well truth estimation performance.



(a) varying ϵ (granularity: $10 * 8$). (b) varying granularity (ϵ=$ln2$).

Fig. 5. Location distortion under different obfuscation strategies.

4) Computation cost: Finally, we test the time overhead of our proposal w.r.t., the linear regression, the obfuscation matrix computation, and the truth discovery stages. Results on two datasets are shown in Fig. 6(a) and Fig. 6(b). The overhead increases with finer-grained spatial division because there are more target grids for estimating obfuscation probability and the number of privacy constraints increases accordingly. As expected, calculating the obfuscation matrix occupies most of the resources for the optimization resolving process. Overheads on the two datasets are similar as the complexity is mainly in proportional to the number of grids. Note that the obfuscation matrix only needs to be computed once or infrequently updated for the target sensing area on MCS server, which makes the overhead practically acceptable.



(a) Time costs on Temp. (b) Time costs on Hum.

Fig. 6. Time consumption of different stages in our proposal.

VI. Conclusion

This work studies the location privacy-preserving problem in crowdsensing truth discovery. By identifying that existing efforts on data privacy protection in truth discovery is not effective with perturbed location indices, we present the design of location differential-obfuscation and observation correlation-guided calibration. We model a linear programming problem with truth estimation deviation as objective and privacy requirements as the constraints. Theoretical analysis on its inherent optimization form and computation complexity is also presented. We evaluate our design by comparing its performance with the No-privacy strategy and two state-of-the-art baselines on real-world sensing datasets.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No. 62102425, 62172155), the Science and Technology Innovation Program of Hunan Province (No. 2021RC2071), and the Natural Science Foundation of Hunan Province (No. 2022JJ40564).

References

[1] T. Zhou, Z. Cai, B. Xiao, L. Wang, M. Xu, and Y. Chen, "Location privacy-preserving data recovery for mobile crowdsensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, pp. 1–23, 2018.

[2] L. He, Y. Lee, and K. G. Shin, "Mobile device batteries as thermometers," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, pp. 1–21, 2020.

[3] Y. Li, H. Xiao, Z. Qin, C. Miao, L. Su, J. Gao, K. Ren, and B. Ding, "Towards differentially private truth discovery for crowd sensing systems," in *Proc. of the IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2020, pp. 1156–1166.

[4] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proc. of the ACM SIGMOD International Conference on Management of Data (Sigmod)*, 2014.

[5] C. Miao, W. Jiang, L. Su, Y. Li, S. Guo, Z. Qin, H. Xiao, J. Gao, and K. Ren, "Privacy-preserving truth discovery in crowd sensing systems," *ACM Transactions on Sensor Networks*, vol. 15, pp. 1–32, 2019.

[6] Y. Li, C. Miao, L. Su, J. Gao, Q. Li, B. Ding, Z. Qin, and K. Ren, "An efficient two-layer mechanism for privacy-preserving truth discovery," in *Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018.

[7] Y. Zheng, H. Duan, X. Yuan, and C. Wang, "Privacy-aware and efficient mobile crowdsensing with truth discovery," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, pp. 121–133, 2020.

[8] L. Li, X. Zhang, R. Hou, H. Yue, H. Li, and M. Pan, "Participant recruitment for coverage-aware mobile crowdsensing with location differential privacy," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[9] S. Azhar, S. Chang, Y. Liu, Y. Tao, and G. Liu, "Privacy-preserving and utility-aware participant selection for mobile crowd sensing," *Mobile Networks and Applications*, vol. 27, pp. 290–302, 2022.

[10] L. Wang, D. Zhang, D. Yang, B. Y. Lim, X. Han, and X. Ma, "Sparse mobile crowdsensing with differential and distortion location privacy," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2735–2749, 2020.

[11] J. Gao, S. Fu, Y. Luo, and T. Xie, "Location privacy-preserving truth discovery in mobile crowd sensing," in *Proc. of the International Conference on Computer Communications and Networks (ICCCN)*, 2020, pp. 1–9.

[12] A. Al-Anwar, Y. Shoukry, S. Chakraborty, B. Balaji, P. D. Martin, P. Tabuada, and M. B. Srivastava, "Proloc: Resilient localization with private observers using partial homomorphic encryption," in *Proc. of 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2017, pp. 41–52.

[13] M. T. Rashid, D. Y. Zhang, and D. Wang, "Socialdrone: An integrated social media and drone sensing system for reliable disaster response," in *Proc. of the IEEE Conference on Computer Communications (Infocom)*, 2020, pp. 218–227.

[14] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: differential privacy for location-based systems," in *Proc. of the ACM SIGSAC conference on Computer & communications security (CCS)*, 2013.

[15] F. Ingelrest, G. Barrenetxea, G. Schaefer, M. Vetterli, O. Couach, and M. B. Parlange, "Sensorscope: Application-specific sensor network for environmental monitoring," *ACM Trans. Sens. Networks*, vol. 6, pp. 17:1–17:32, 2010.