Incentivizing Crowdsensing-Based Noise Monitoring with Differentially-Private Locations

Pei Huang, Xiaonan Zhang, *Student Member, IEEE*, Linke Guo, *Member, IEEE*, and Ming Li, *Member, IEEE*

Abstract—Mobile crowd sensing is a technique where a crowd sensing server outsources sensing tasks to the crowd for mobile data collection. In mobile crowd sensing, some tasks require location information to achieve their objectives, such as road monitoring, indoor floor plan reconstruction, and smart transportation. This required information incurs severe concerns on location privacy leakage and threatens workers' properties as well as public safety. In some cases, even sensing data itself can be used as auxiliary information resulting in location privacy breaches. Many existing works apply differential privacy mechanisms for location privacy preservation to tackle this problem, but they cannot efficiently fulfill privacy goals because each worker only considers his own privacy. As a consequence, the accumulated privacy budget will lower down the composed privacy level of all the workers' locations. In addition, deploying differential privacy is costly for workers and it will degrade the quality of data required in crowd sensing tasks. How to balance the cost and provide accurate aggregated data while fulfilling privacy objectives becomes a challenging issue. In this paper, we propose a group-differentially-private game-theoretical solution, which addresses these limitations in a privacy-preserving and efficient way. Our scheme enables the indistinguishability of workers' locations and sensing data without the help of a trusted entity while meeting the accuracy demands of crowd sensing tasks. The effectiveness and efficiency of our scheme are thoroughly evaluated based on real-world datasets.

Index Terms—Mobile crowd sensing, location privacy, game theory, differential privacy

1 Introduction

Mobile crowd sensing is an emerging sensing paradigm that outsources the collection of data to a crowd of participating workers with mobile devices. Mobile devices are equipped with a plethora of on-board sensors (e.g., compass, accelerometer, gyroscope, camera, GPS) to sense various types of data. Specifically, location data sensed by GPS has been applied to serve a wide spectrum of location-based mobile crowd sensing applications, including road monitoring [1], indoor floor plan reconstruction [2], and smart transportation [3]. Workers who participate in the location-centric mobile crowd sensing usually have to upload data containing their location information. For example, in a noise-monitoring mobile crowd sensing system, the server collects the noise level data at preset time slots near noise sources to estimate the noise exposure, and further provides references to noise control methodologies. McAlexander et al. in [4] set up a process for collecting and analyzing noise data throughout New York City from the crowd. Workers in the area near noise sources will provide their locations together with time and sensed noise levels to the crowd sensing server. However, the disclosure of actual locations that a worker visited with timestamps will compromise his location privacy and be used to infer his daily routines, personal interests, etc. Another aspect of privacy breaches resides in the noise level data itself. For a single, non-directional noise source, its noise power will diminish inversely to the square of the distance from the source because the sound energy is spreading over the increasing area of a sphere. The sound pressure levels of various random and uncorrelated noise sources can be added together [5]. Given this knowledge, since the server knows the exact locations of noise sources and even has the open access to estimated noise level online (e.g., Manhanttan noise map [6]), it can estimate the noise levels at locations with different distances to noise sources. Workers' actual distances to the noise centers can be reckoned, and finally, their actual locations can be inferred. Thus, noise level data should also be protected. To address these issues, every worker should incorporate a privacy-preserving mechanism to both their locations and sensing data. However, anonymizing workers' data alone is not helpful in the sense that anonymized data still can be used to infer daily patterns and other information [7], [8].

Differential privacy [9] is a mechanism applied in numerous systems to protect data privacy. In our noise-monitoring crowd sensing scenario, if everyone simply deploys a differential privacy mechanism to himself without cooperating

- *P. Huang, X. Zhang, and L. Guo are with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634 USA. E-mail: {peih, xiaonaz, linkeg}@clemson.edu.*
- *M. Li is with the Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington, TX 76019 USA. E-mail: ming.li@uta.edu.*

Manuscript received 29 Jan. 2019; revised 13 Aug. 2019; accepted 7 Oct. 2019. Date of publication 11 Oct. 2019; date of current version 7 Jan. 2021. (Corresponding author: Linke Guo.)
Digital Object Identifier no. 10.1109/TMC.2019.2946800

with others, the group privacy level of their locations at one time will be far lower than their individual privacy levels due to the composition property of differential privacy [9]. To achieve a low composed budget, the simplest way is decreasing each worker's privacy budget, but it will bring large variances in workers' perturbed data and hinder the completion of tasks. Differential privacy can also be embedded in the sensing data for location protection because noise level data possesses some location-related characteristics as mentioned above. Obviously, protecting privacy is bound to degrade data accuracy and service quality in mobile crowd sensing while consuming more computation resources and energy. Therefore, without a proper scheme designed for protecting the group location privacy and data privacy while guaranteeing task accuracy, workers will be reluctant to participate in location-based mobile crowd sensing tasks and the crowd sensing server cannot make desired profits from collected data nor provide sound data services.

Our Contributions. Our work enables location-related differential privacy for a group of workers. The Bayesian game [10] is deployed to model workers' behaviors and costs, giving a practical basis for formal decision making and algorithm development . We list our contributions as follows:

- Our scheme fulfills indistinguishability for locations without the help of a trusted entity. Moreover, locations cannot be inferred from workers' sensing data. Sensitive information leakage can be effectively restricted even for a group of workers.
- The server receives data with the desired accuracy when workers' data is protected.
- Our scheme enables workers to make proper choices for cost minimization while being aware of malicious workers during decision making.

The remainder of this paper is organized as follows. Section 2 introduces several preliminaries. Section 3 presents the system architecture and the adversary model. A new location inference attack is proposed and analyzed in Section 4. Then, we describe the detailed game in Section 5, followed by the privacy preservation scheme. The protocol is evaluated in Section 6. In the following Section 7, we briefly review some relevant works and their deficiencies. Finally, Section 8 concludes the paper.

2 Preliminaries

2.1 Physics of Sound and Noise Control

As introduced in the Noise Manual [5], the decibel notation of sound pressure level (SPL) is:

$$L_p = 20\ \log\left(\frac{p}{p_0}\right) = 10\ \log\left(\frac{p^2}{p_0^2}\right), \tag{1}$$

where p is the measured root-mean square (rms) sound pressure and p_0 is the reference rms sound pressure. The reference distance to the sound source is set to be r_0. The effectiveness of a noise control method is usually evaluated from the arithmetic difference between the SPLs measured at two locations as $L_{p_1} - L_{p_2}$, one on either side of a noise control device, or at the same fixed measuring location before and after a noise control method has been applied.

Sound levels can be added together in decibels when there are multiple noise sources. An example is the estimation of the total SPL due to the addition of a new machine of known sound output to an existing noise environment of known characteristics. The addition equation of the sound levels for N random, uncorrelated sounds is

$$L_P = 10\ \log\left(\frac{\sum_{i=1}^{N} p_i^2}{p_0^2}\right) = 10\ \log\left(\sum_{i=1}^{N} \frac{p_i^2}{p_0^2}\right).$$

From Equation (1), we find that $p_i^2/p_0^2 = 10^{\frac{L_{p_i}}{10}}$. Hence,

$$L_P = 10\ \log\left(\sum_{i=1}^{N} 10^{\frac{L_{p_i}}{10}}\right), \tag{2}$$

where L_P is the total SPL in decibels generated by N sources and L_{p_i} represents the individual SPLs to be added.

Many noise-control problems require knowledge of the relationship between sound fields and SPL. We only introduce the free field here for simplicity. A free field exists when sound radiates into space from a source and there is nothing to impede the sound energy as it flows from the source. Considering a small and nondirectional sound source that is radiating sound equally in all directions, the SPL will be the same at any point on the surface of a sphere centered on the source. The sound intensity diminishes inversely as the square of the distance, r, from the source since the sound energy is spreading over the increasing area of the sphere ($4\pi r^2$). Thus, the decrease of noise level with respect to the distance from a nondirectional noise source is proportional to $10\ \log\left|\frac{1}{r^2}\right|$.

2.2 Differential Privacy

Differential privacy is first introduced in the database, where a database can be viewed as a set of rows. Databases D_1 and D_2 differ in at most one element if one dataset is a proper subset of the other and the larger database contains just one additional row [9].

Definition 2.1 (Differential Privacy). *A randomized function $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq \mathrm{Range}(\mathcal{K})$,*

$$Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \times Pr[\mathcal{K}(D_2) \in S]. \tag{3}$$

The probability is taken over the coin tosses of $\mathcal{K}$.

Differential privacy can also be interpreted to be (α, β)-accurate as introduced in [9].

Definition 2.2 ((α, β)-Accuracy). *For two random variables Y_1 and Y_2 within the range, Y_1 is (α, β)-accurate to Y_2 if and only if $\Pr[|Y_1 - Y_2| \geq \alpha] \leq \beta$, where $\beta \in (0, 1)$.*

Geo-indistinguishability [11] is a formal notion of privacy for location-based systems to protect the actual location of a user (worker), while still allowing approximate information needed for a certain desired service to be released. The indistinguishability of locations is achieved so that an adversary cannot tell a random location from the actual location. Geo-indistinguishability is an extension of the generalized version of differential privacy.

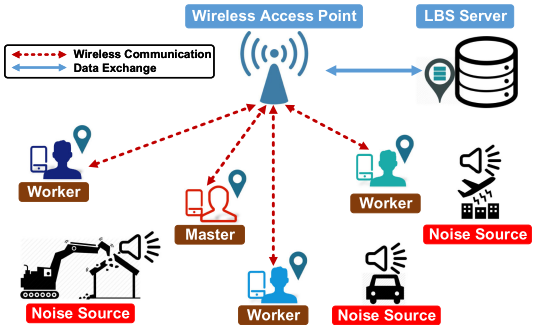


Fig. 1. System model.

Definition 2.3 (Geo-indistinguishability). *For any radius $r > 0$, a mechanism $\mathcal{K}$ satisfies ϵr-geo-indistinguishability iff the worker enjoys ϵr-privacy within r.*

The ϵr-geo-indistinguishability is mathematically defined as: a mechanism satisfies ϵr-geo-indistinguishability iff for all observations in a set of possible reported values S, the probability that the worker is assumed to be located at x and x' are bounded as

$$\frac{P(S|x)}{P(S|x')} \leq e^{\epsilon r}, \ \forall\ r > 0, \ \forall x, x' : \ d(x, x') \leq r. \tag{4}$$

One property of geo-indistinguishability is that the privacy level at a location is smaller when it is farther away from the worker's location. Within a small radius, for instance, $r = 1$ km, ϵr is also small, which guarantees that the attacker cannot infer the worker's location. When r increases, such as $r = 10$ km, ϵr becomes very large, and the attacker can infer that the worker is located within a city with high probability. Adding noise is a way to fulfill the requirements of geo-indistinguishability, which is the same as traditional differential privacy. However, since the location is a two-dimensional coordinate including longitude and latitude, we should add a planar Laplace distribution in the polar coordinate system to the location. For example, given an actual location $x = (s, t)$, we need to pick a point (r_c, θ) from the polar Laplacian. Then, the obfuscated point will be $z = (s + r_c \cos\theta, t + r_c \sin\theta)$.

3 System Model

3.1 System Overview

Our system shown in Fig. 1 consists of a LBS server and a set of participating workers, denoted as $\mathcal{W} := \{W_1, \ldots, W_w\}$.

- *LBS Server:* The LBS server publishes a set of K sensing tasks, denoted as $\mathcal{T} = \{T_1, \ldots, T_K\}$, and corresponding rewards, e.g., a fixed or flexible (determined by the workload of each worker) amount of financial compensation, for workers to participate. It is also responsible for distributing public parameters for privacy preservation to workers and aggregating data collected by workers. One practical scenario is a noise controlling one: the server is notified of several newly coming noise sources and wants to check how these sources impact the neighborhood. The workers' locations and noise data are not required to be exactly the same as ground truth, but should be bounded in a certain range (the locations are in the wanted neighborhood and the noise data can reflect the impact) to retain the task utility.
- *Workers:* A bunch of workers in one sensing region is required to report noise levels at their locations for each task. If a worker W_i accepts the task T_j, he will provide his location and noise level as $\{l_{i,j}, X_{i,j}\}$ to the server (cheating, e.g., accept a task but do not report sensing data, is not related to information privacy concerns in this paper, so it is not considered). However, directly reporting data to the server will impair workers' privacy. So, in our scheme, workers only report perturbed locations and falsified noise levels $\{l'_{i,j}, X'_{i,j}\}$ via differential privacy. In case workers have to collaborate to fulfill privacy goals, a master is selected to collect information from all members.
- *Wireless Infrastructures:* Workers communicate with the server and each other via cellular networks, such as 4G/LTE. The communication is done via network access points such as base stations, so they do not need to be in close physical proximity for direct communication. The server also relies on cellular networks to release tasks and collect data from the workers. The base station is assumed to be honest for relaying messages, and it does not reveal workers' location information from physical properties, such as signal strength, during the packet relay. However, it is not a centralized trusted third party and does not contribute to the privacy protection process except for packet relaying.

3.2 Adversarial Model

- *Server:* The server is assumed to be honest but curious. It will follow rules when announcing tasks, giving out rewards, and computing the aggregated data. However, it tries to get workers' private information and make profits from it. The server should not learn the worker's private information from uploaded locations and sensing data.
- *Workers:* Workers are curious about other workers' information and may misuse their location and data information, so a worker's actual location and sensing data should be kept confidential from others and hard to be inferred from his/her uploaded data. There are also malicious workers, who are going to cheat and make the privacy goal fail.

3.3 Notations and Definitions

For clarity, we summarize the important notations in the following Table 1 and list their corresponding definitions.

4 Location Inference Attack

In the example of noise monitoring crowd sensing, workers are required to report noise levels with location information, which further compromises their location privacy. Here, we identify a new location inference attack to illustrate this additional location privacy leakage in mobile crowd sensing.

TABLE 1
Notation Table

Notations	Definitions
T_j	the jth crowdsensing task in the task set $\mathcal{T}$
$W_i^{T_j}$	the ith worker in the group W^{T_j} for T_j
$l_{i,j}$	the location of worker i when carrying out task T_j
$X_{i,j}$	the sensing data reported by worker i for task T_j
$l'_{i,j}, X'_{i,j}$	the perturbed versions of $l_{i,j}, X_{i,j}$
ϵ_l, R_i	the privacy budget for location privacy protection
ξ	the accuracy requirement of the reported sensing data
G_{p_i}	the privacy gain for ith worker if participate in game
G_{d_i}	the privacy gain for ith worker if deceive in game
S_i	the strategy chosen by the ith worker in the game
S_{-i}	the strategies chosen by all workers other than i
ρ_i	the type assigned to the ith worker
U_i	the utility gained in the game for ith worker
$\mathcal{F}$	the distribution of privacy requirements
η_i	the expected **Cooperate** probability of worker i
f^o_{h-1}	the fan-out for parent nodes at height h in the tree
$c_{i,h-1}$	the value of the ith node at height $h-1$ in the tree
$D_{\epsilon_l}(x_0)(x)$	the planar Laplace noise that perturb location x_0 to x
$C_{\epsilon_l}(r)$	the cumulative function of $D_{\epsilon_l}(x_0)(x)$
$r_i^l.\theta_i^l$	the noises added to location $l_{i,j}$ in polar coordinates

4.1 Attack Description

This attack is similar to trilateration with auxiliary knowledge. Because the noise level is closely related to the distances between the noise sources and measuring devices, an entity that has knowledge of the noise sources' locations can deduce a noise map. Assume that the server broadcasts one sensing task in one round to the target area. We conduct a simulation with 100×100 grids as an example, in which exist three noise sources at the squared regions with coordinates [20,45], [42,40], and [23,37]. The SPLs of noise sources measured at the reference distances r_0 within the red regions are 100 dB, 90 dB, and 80 dB, respectively. The theoretic SPLs distributed in each grid without considering the specific environments (e.g., building, traffic, and etc.) can be computed from Equations (1) and (2). For a spot whose distances to the three red regions are r_1, r_2, r_3, the SPLs incurred by each individual noise source are $L_{p_1} = 100 - \left|20 \log \frac{r_0}{r_1}\right|$, $L_{p_2} = 90 - \left|20 \log \frac{r_0}{r_2}\right|$, and $L_{p_3} = 80 - \left|20 \log \frac{r_0}{r_3}\right|$. Therefore, the SPL after taking all three sources into account is $L_P = 10 \log \left(\sum_{i=1}^{3} 10^{\frac{L_{p_i}}{10}}\right)$. We simulate each grid's SPL on MATLAB and list parts of them in Fig 2, where grids containing noise sources are painted red. Suppose there is a worker resides at the yellow grid and reports his true noise level data 69.81427 dB. With the help of differential privacy and geo-indistinguishability, he intends to hide himself in a circled area colored in green. However, his noise level data is closer to those in the blue grids, indicating that he is more likely to appear in the blue region. Thus, his location information is further revealed. As a result, we need to randomize workers' sensing data to avoid location information leakage.

4.2 The Defending Capability of Differential Privacy

Suppose that the worker applies geo-indistinguishability to perturb his location l as l'. Then, we will analyze how the inference ability is restricted by geo-indistinguishability in traditional inference attacks and how the ability is enhanced in our proposed attack. The attacker's goal is to infer the user's actual location l. We assume that the adversary has prior knowledge of the location obfuscation mechanism, which is named as an informed adversary in [12].

x/y	36	37	38	39	40	41	42	43	44	45
19	70.13624	71.10184	72.18349	73.43516	74.92418	76.74026	79.0266	82.02641	86.00053	89.00923
20	70.32591	71.29677	72.35507	73.59476	75.11047	77.00862	79.48494	82.99468	89.00952	100.0001
21	70.59573	71.58633	72.47626	73.55685	74.9692	76.7567	79.03257	82.02844	86.00114	89.00947
22	71.31369	72.97867	72.83085	73.36147	74.53857	76.06975	77.90401	79.99485	82.02775	82.99406
23	72.30724	80.45777	73.40365	72.98703	73.89339	75.12527	76.50755	77.8993	79.02824	79.48168
24	70.94009	72.58731	72.1654	72.34005	73.10704	74.07916	75.10272	76.04688	76.73974	76.99719
25	69.71265	70.49131	71.00951	71.5578	72.25386	73.02795	73.79326	74.45705	74.91611	75.07875
26	68.97068	69.61052	70.19693	70.78247	71.3983	72.02251	72.60644	73.09023	73.41181	73.52126
27	68.40084	68.96509	69.51131	70.04839	70.57883	71.08683	71.54176	71.90548	72.13999	72.21625
28	67.90252	68.40325	68.89089	69.36318	69.81427	70.23056	70.59111	70.87114	71.04678	71.10033
29	67.44873	67.89314	68.32279	68.73236	69.11425	69.45715	69.74621	69.96496	70.09805	70.13433
30	67.03554	67.43102	67.80865	68.16257	68.4856	68.76874	69.00144	69.17265	69.27243	69.2937
31	66.66777	67.02255	67.35605	67.66259	67.93594	68.16925	68.35524	68.48676	68.55763	68.56359
32	66.35467	66.67841	66.977	67.2448	67.47639	67.66675	67.81136	67.90633	67.94859	67.93631
33	66.10902	66.41386	66.68883	66.92741	67.12429	67.27585	67.38028	67.4371	67.4467	67.40992
34	65.947	66.24956	66.51621	66.73757	66.90675	67.02086	67.08131	67.09241	67.05949	66.98728
35	65.888	66.21263	66.49396	66.71567	66.86559	66.94013	66.94543	66.89418	66.8003	66.67491
36	65.9533	66.33698	66.67018	66.92141	67.06473	67.09219	67.0186	66.8734	66.68688	66.48086
37	66.16166	66.6609	67.10862	67.44215	67.60138	67.56431	67.36432	67.06783	66.73707	66.41048
38	66.51945	67.21764	67.88701	68.40499	68.62594	68.4847	68.06208	67.51258	66.95865	66.45869
39	67.00224	68.01024	69.08121	69.99688	70.38762	70.04227	69.19099	68.21411	67.32994	66.59892
40	67.52873	68.95639	70.69925	72.49634	73.38327	72.51748	70.7621	69.09319	67.77201	66.77245
41	67.94487	69.80372	72.44624	76.17852	79.10139	76.18607	72.48137	69.89811	68.13098	66.88923
42	68.06453	70.13517	73.30604	79.09179	90.00688	79.09503	73.33027	70.20886	68.21761	66.85382
43	67.78543	69.69397	72.3835	76.15075	79.08664	76.15613	72.40882	69.76305	67.92407	66.61067

Fig. 2. Noise-Location table.

4.2.1 Defending against Traditional Inference Attacks

Since that the adversary knows the distribution of location obfuscation scheme and the set of possible locations after observing l', he can perform the optimal inference attack and Bayesian inference attack [13] by computing the highest posterior probability $\Pr(l|l')$ and minimizing the expected inference error as introduced in [14]. The posterior probability distribution follows:

$$\Pr(l|l') = \frac{\varphi(l) D_{\epsilon_l}(l'|l)}{\sum_{l \in S} \varphi(l) D_{\epsilon_l}(l'|l)},$$

where φ is the prior distribution of the set S of possible locations, D_{ϵ_l} is the distribution of the added noise in geo-indistinguishability, and ϵ_l is the privacy budget.

Based on the probability distribution, the expected inference error is formulated as the expected distance between the estimated location $\hat{l}$ and the actual location l

$$\sum_{l' \in S} \min_{\hat{l} \in S} \sum_{l \in S} \varphi(l) D_{\epsilon_l}(l'|l) d(\hat{l}, l). \tag{5}$$

Thus, the optimal inference attacker guesses a location $\hat{l}$ by minimizing the expected distance: $\hat{l} = \arg\min \sum_{l \in S} \Pr(l|l') d(\hat{l}, l)$. A Bayesian inference attacker chooses an estimated location by maximizing the posterior probability: $\hat{l} = \arg\max \Pr(l|l')$.

According to the analysis in [14], the capability of geo-indistinguishability for defending against optimal inference attack and Bayesian inference attack are bounded. The lower bound of inference error is $e^{-\epsilon_l} \min \sum \frac{\varphi(x)}{\sum \varphi(y)} d(\hat{l}, x)$ and the upper bound of posterior probability is $e^{\epsilon_l} \frac{\varphi(x)}{\sum \varphi(y)}$, where x and y are random locations in the protected region.

4.2.2 Noise-Level-Involved Inference Attack

Obviously, the defending capability is influenced by the size of the protected region and the privacy budget. Our noise-level-involved inference attack can impair the

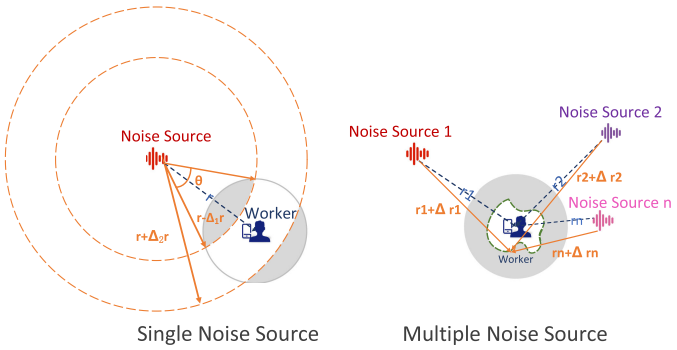


Fig. 3. Attack demonstration.

geo-indistinguishability's defending capability without changing the privacy budget, and this negative effect goes stronger with more noise sources. The adversary in our proposed attack model considers the noise level data X published by the worker and the locations of all noise sources provided by the crowd sensing server for deducing the worker's actual location. How to scale down the protected region is demonstrated in Fig. 3 and explained as follows:

Scenario 1: Single Noise Source. The worker is r meters away from one single noise source with noise level data X and his location is protected in the gray circle by geo-indistinguishability. The adversary assumes that the worker locates in the range where all noise levels are $X \pm \Delta X$. According to the relationship between SPL and distance introduced in Section 2.1, the protected region is cut by two orange circles, whose radiuses are $r - \Delta_1 r$ and $r + \Delta_2 r$, respectively. These radiuses are derived from

$$10 \log \left| \frac{r^2}{(r - \Delta_1 r)^2} \right| = 10 \log \left| \frac{(r + \Delta_2 r)^2}{r^2} \right| = \Delta X.$$

Thus, the protected region is restricted (the white part of the gray circle). The lower bound of inference error declines and the upper bound of posterior probability grows, implying that the attacker can deduce the actual location more easily.

Theorem 1. *When $R \ll r$, the lower bound of inference error is decreased by $\min\{\Delta_1 r, \Delta_2 r\}/R$ and the upper bound of posterior probability is increased in relationship with $2R^2/[\theta(\Delta_1 r + \Delta_2 r)(2r + \Delta_2 r - \Delta_1 r)]$, where R is the radius of original protected region and θ is the angle shown in Fig. 3.*

Proof. When $R \ll r$, the shrunk protected region can be approximated by the area difference between two sectors with the same angle θ, which is computed according to the law of cosines

$$\theta = 2 \arccos \frac{(r - \Delta_1 r)^2 + r^2 - R^2}{2r(r - \Delta_1 r)}.$$

Therefore, the size of the shrunk region is $\theta\pi[(r + \Delta_2 r)^2 - (r - \Delta_1 r)^2]/(2\pi)$, which is simplified as $\theta(\Delta_1 r + \Delta_2 r)(2r + \Delta_2 r - \Delta_1 r)/2$.

In the formulation of inference error's lower bound, the effects of the region size are canceled out due to the sum at the denominator and the overall sum, but the minimum distance between $\hat{l}$ and the random location x is reduced from R to $\min\{\Delta_1 r, \Delta_2 r\}$.

As for the upper bound of posterior probability, $\varphi(x)$ is not affected because the protect area size is not the prior knowledge obtained by the attacker. However, the sum of $\varphi(y)$ over the protected area is lowered due to the side information: reduction in protected area size. The ratio of area reduction is by $[\theta(\Delta_1 r + \Delta_2 r)(2r + \Delta_2 r - \Delta_1 r)]/2R^2$. Therefore, the upper bound turns out to be larger.

The changes in the bounds are proved. □

How the attack goes stronger for other R and r can also be derived by computing the area of shrunk protected region.

Scenario 2: Multiple Noise Sources. If this attack is extended to multiple noise sources, the border of the shrunk region is defined by the noise level as follows:

$$10 \log \left| \frac{\frac{1}{r_1^2} + \frac{1}{r_2^2} \cdots \frac{1}{r_n^2}}{\frac{1}{(r_1 + \Delta r_1)^2} + \frac{1}{(r_2 + \Delta r_2)^2} \cdots \frac{1}{(r_n + \Delta r_n)^2}} \right| = \Delta X$$

$$10 \log \left| \frac{\frac{1}{(r_1 - \Delta r_1)^2} + \frac{1}{(r_2 - \Delta r_2)^2} \cdots \frac{1}{(r_n - \Delta r_n)^2}}{\frac{1}{r_1^2} + \frac{1}{r_2^2} \cdots \frac{1}{r_n^2}} \right| = \Delta X.$$

Intuitively, the shrunk region observed by the adversary, where the noise levels are bounded by $X \pm \Delta X$, is smaller than that under the single-source scenario and the inference attack ability is further reinforced.

5 Scheme Details

5.1 Overview

In our scheme, the server first publishes a set of sensing tasks $\mathcal{T} = \{T_1, \ldots, T_K\}$ and corresponding rewards, i.e., money, for workers to participate. A privacy budget for location privacy ϵ_l and a sensing data accuracy requirement ξ are announced together by the server. Workers spot these tasks and decide whether they will accept them according to the rewards. Then, workers who reside in the task T_i's required area and accept this task will form a group $\mathcal{W}^{T_i}$. Assume there is a master worker in $\mathcal{W}^{T_i}$ who is responsible for collecting information from all members and computing individual's parameters to provide differential privacy in this group. Group members provide needed information to the master and receive noise parameters back from the master. Based on all parameters, the "falsified" locations and randomized noise levels are calculated and uploaded by workers.

The privacy budgets ϵ_l and accuracy requirement ξ require the server to utilize some existing methods to find out. Since each worker has his own privacy demands, the budget should satisfy more workers' privacy demands while guaranteeing the accuracy, which will enable the mobile crowd sensing system to collect high-quality data from the crowd. A game-theoretical approach is necessary to be incorporated into the design to let the group $\mathcal{W}^{T_i}$ choose a proper master, because how to effectively achieve the composed group privacy goals also needs more efforts on cooperation. After the game, workers collaborate to reach group geo-indistinguishability and differential privacy on the sensing data.

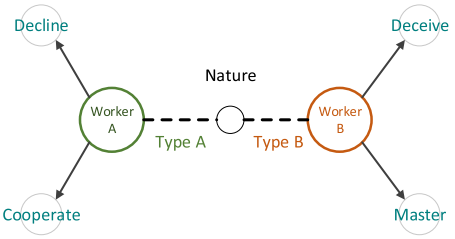


Fig. 4. Strategy set.

5.2 Private Crowd Sensing Game Model

In the game, for one task T_j, all workers who agree to perform this task are making decisions simultaneously without being informed of others' choices and game utilities. This activity is modeled as a non-cooperative Bayesian game denoted as the Private Crowd sensing Game (PCG). In accordance with the definition of a Bayesian game, the PCG has the following components:

5.2.1 The Player Set and the Finite Strategy Set

The player set $\mathcal{W}^{T_j} = \{W_i^{T_j} | i \in \{1, \ldots, N\}\}$ includes all workers currently present within the service area and accept privacy budgets when conducting the task T_j.

The strategy set S_i of a player refers to all available moves to the player. In our scheme, a malicious master cannot get the precise information of workers' locations, but he can refuse to return the needed values so that the privacy goal fails. Other workers can decide whether to cooperate with the master or decline the task to avoid loss. Therefore, strategies included in the set are shown in Fig. 4: **Master**, i.e., group management and data collection; **Deceive**, i.e., announce to be a master but do not fulfill the requests from other workers; **Cooperate**, i.e., to be a group member and send needed information to the master, do not interact with other cooperative workers; or **Decline**, i.e., be suspicious of a deceiving master and refuse to cooperate. The strategies for all players are represented as $\boldsymbol{S} = \{S_1, \ldots, S_N\}$. The strategy profile of $W_i^{T_j}$'s opponents is S_{-i}.

5.2.2 Game Utility

The game utility of a player depends on the strategies adopted by all players. When there is a master who helps preserve privacy, players' game utilities are their obtained privacy gains subtracting their costs severally. However, if none of the players chooses to be a master in the game, or if the master is a malicious one, the privacy objective fails. For a player $W_i^{T_j}$, we will discuss his privacy gains before introducing other factors.

Privacy Gains. Following the formulations in Equations (2) and (4), we define privacy gains for each worker $W_i^{T_j}$ as quantitative measures for the probabilistic ratio of indistinguishability benefits. With a privacy budget $\epsilon_l R_i$, the probability for an adversary to distinguish the true location from a falsified one is $\exp(\epsilon_l R_i)$. This probability decreases with the increase of budget value, which indicates that it becomes more possible for an attacker to guess out the actual location. Therefore, we formulate the location privacy gain as an inverse of the aforementioned probability. The gain for noise level data privacy follows the same fashion because it is also protected with differential privacy. The location privacy gain P_g^l and the noise level data privacy gain P_g^n are summed up to represent the privacy gain G_{p_i}

$$G_{p_i} = \frac{1}{\exp(\epsilon_l R_i)} + \frac{1}{\exp\left(\left|\frac{10 \log[\max(\zeta_1, \zeta_2)]}{\ln \xi}\right|\right)}, \tag{6}$$

where ϵ_l and R_i are the privacy budget and $W_i^{T_j}$'s radius requirement in geo-indistinguishability, ξ is the data accuracy requirement, $\zeta_1 = (1 + R_i/\kappa_i)^2$ and $\zeta_2 = (1 - R_i/\kappa_i)^2$. The first part on the right side reflects the inverse of posterior information gain. The posterior information gain is the ratio of posterior probability and prior probability, so the adversary benefits more and the worker enjoys less privacy with a larger posterior information gain. How we get the second part on the right side will be introduced in Section 5.3.2. Briefly speaking, it is the inverse of posterior information gain for noise level data.

Disruption Gains. The gain of malicious activities comes from disrupting the privacy goal and wasting the resources of regular nodes. If player $W_i^{T_j}$ is a malicious player, he gains G_{d_i} after disrupting the privacy procedure.

Energy Costs and Ratios. A master needs more energy to communicate and process data than ordinary members as he needs to exchange information with $N-1$ members via available networks and compute on these data. The communication cost is the energy spent on accessing network and sending/receiving data, so it depends on the data packet size, which is relevant to the number of workers. We denote the cost for a master as the fraction of energy costs to his remaining battery. To simplify the expression, we use the fraction of the expected sum of a group member's network connection cost, data sending/receiving cost, and data processing cost to the expected master energy consumption as the degradation factor δ. The unit costs are defined as the energy depleted by actions on one data packet. In all, the costs are formulated as follows:

$$\begin{aligned}
E(\textbf{Master}) &= \frac{3(N-1)E_{c1} + (N+3)E_{c2} + E_d}{E^{T_j}(\mathbf{B})}, \\
E(\textbf{Deceive}) &= \frac{(N-1)E_{c1} + E_{c2}}{2E^{T_j}(\mathbf{B})}, \\
E(\textbf{Cooperate}) &= \frac{E_{c1} + E_{c2} + E_d}{E^{T_j}(\mathbf{B})}, \\
E(\textbf{Decline}) &= 0,
\end{aligned}$$

where E_{c1} is the network connection cost, E_{c2} is the unit data sending/receiving cost, E_d is the unit energy consumed when processing data, and $E^{T_j}(\mathbf{B})$ is the remaining battery capacity when receiving task T_j. For simpler and clearer formulation, we do not go into the details of data processing complexity. In Section 5.3.3, we will interpret how we get these energy costs after presenting the complete scheme. The ratios of costs and gains $\gamma_i(\cdot)$ are derived and normalized from cost formulations.

Now that every worker knows the expected costs and gains, the utility functions for a regular player can be formulated as

$$U_i(\cdot;\rho_i,S_i,S_{-i})=\begin{cases} G_{p_i}[1-\gamma_i(S_i)], & S_i=\textbf{Master},\\ G_{d_i}[1-\gamma_i(S_i)], & S_i=\textbf{Deceive},\\ 0, & S_i=\textbf{Decline},\\ G_{p_i}[1-\gamma_i(S_i)], & S_i=\textbf{Cooperate},\\ & n_m(S_{-i})\geq 1,\\ & n_d(S_{-i})=0\\ -E_{c1}/E^{T_j}(\mathbf{B}), & S_i=\textbf{Cooperate},\\ & n_m(S_{-i})=0\\ -E(S_i)/2, & S_i=\textbf{Cooperate},\\ & n_m(S_{-i})\geq 1,\\ & n_d(S_{-i})\geq 1 \end{cases},$$

where $n_m(S_{-i})$ denotes the number of masters (excluding deceiving masters) other than $W_i^{T_j}$ himself and $n_d(S_{-i})$ denotes the number of deceiving masters. If there is no master, the privacy goal fails and the network connection energy is wasted. If the announced masters are all malicious, the cooperative player gains nothing and wastes his energy, which is less than the energy consumed when actually cooperating because there is no return data from the master.

5.2.3 Type ρ

A player does not have complete knowledge about other players' utilities, so we introduce Nature into the game as in [15]. Each player is assigned a type ρ_i by Nature, which is the privacy requirement for geo-indistinguishability, $\epsilon_l R_i$, and the malicious probability p_i. The privacy requirements are sampled independently from a distribution $\mathcal{F}$ with the probability density function f. In PCG, the strategy space, possible types, $\mathcal{F}$, and f are known to players, while ρ_i is $W_i^{T_j}$'s private information.

5.2.4 Best Response and Bayesian Nash Equilibrium

In the game, all rational players' choices are made to maximize their utilities. We introduce the concept of best response in [16] to capture this behavior:

Definition 5.1 (Best Response). *A player $W_i^{T_j}$'s best response $\hat{S}_i$ is the strategy that maximizes his utility U_i given S_{-i}*

$$\hat{S}_i(S_{-i})=\arg\max U_i(\rho_i,S_i,S_{-i}).$$

Our PCG is a Bayesian game on account of players with incomplete information. Here we refer to the Bayesian Nash Equilibrium to describe the steady state, where no player will get better utility by unilaterally changing his strategy.

Definition 5.2 (Bayesian Nash Equilibrium). *A strategy profile $S^*=S_1^*,S_2^*,\ldots,S_n^*$ is a Bayesian Nash Equilibrium (BNE) if strategy S_i^* for every player $W_i^{T_j}$ is the best response that maximizes their expected utilities. That is, given S_{-i} and players' beliefs about the types of other players ρ_{-i}, we have*

$$S_i^*(\rho_i)\in\arg\max_{S_i}\sum_{\rho_{-i}} f(\rho_{-i})\times U_i(\rho_i,S_i^*,S_{-i}^*),\ \forall\rho_i.$$

5.2.5 Mixed Strategy Bayesian Nash Equilibrium of PCG

To derive a Bayesian Nash Equilibrium, we assign p_i to each player $W_i^{T_j}$ as the probability that this player behaves maliciously. Then, we use the cooperate probability in [16] as a reference. Given a type ρ_i, a player $W_i^{T_j}$ has a probability of $d_i(\rho_i)$ to choose **Cooperate**. The expected probability that one player chooses **Cooperate** is

$$\eta_i=\int d_i(\rho_i)d\mathcal{F}(\rho_i). \tag{7}$$

Then, the expected utility for player $W_i^{T_j}$ is re-formulated as

$$\begin{aligned}
&\mathbb{E}[U_i(\textbf{Master};\rho_i,S_i,S_{-i},\mathcal{F})]\\
&\quad=(1-p_i)G_{p_i}[1-\gamma_i(\textbf{Master})],\\
&\mathbb{E}[U_i(\textbf{Deceive};\rho_i,S_i,S_{-i},\mathcal{F})]=p_iG_{d_i}[1-\gamma_i(\textbf{Deceive})],\\
&\mathbb{E}[U_i(\textbf{Cooperate};\rho_i,S_i,S_{-i},\mathcal{F})]\\
&\quad=G_{p_i}\times\left(1-\prod_{k\neq i}\eta_k\right)\times\prod_{j\neq i}(1-p_j)-\frac{E_{c1}}{E^{T_j}(\mathbf{B})}\prod_{k\neq i}\eta_k\\
&\qquad-E(\textbf{Cooperate})\left(1-\prod_{k\neq i}\eta_k\right)\left(\frac{1}{2}+\frac{1}{2}\prod_{j\neq i}(1-p_j)\right),\\
&\mathbb{E}[U_i(\textbf{Decline};\rho_i,S_i,S_{-i},\mathcal{F})]=0.
\end{aligned}$$

Since the goal of $W_i^{T_j}$ is to maximize his game utility, the player chooses to be a master, whether malicious or not, if

$$\begin{aligned}
&\mathbb{E}[U_i(\textbf{Master};\rho_i,S_i,S_{-i},\mathcal{F})]\\
+&\mathbb{E}[U_i(\textbf{Deceive};\rho_i,S_i,S_{-i},\mathcal{F})]\\
>&\mathbb{E}[U_i(\textbf{Cooperate};\rho_i,S_i,S_{-i},\mathcal{F})]\\
+&\mathbb{E}[U_i(\textbf{Decline};\rho_i,S_i,S_{-i},\mathcal{F})].
\end{aligned}$$

His further choice in line with game utilities becomes

$$s_i=\begin{cases}\textbf{Master}, & \text{if}\quad p_i\leq\frac{1}{1+\frac{G_{d_i}[1-\gamma_i(\textbf{Deceive})]}{G_{p_i}[1-\gamma_i(\textbf{Master})]}}\\ \textbf{Deceive}, & \text{if}\quad p_i>\frac{1}{1+\frac{G_{d_i}[1-\gamma_i(\textbf{Deceive})]}{G_{p_i}[1-\gamma_i(\textbf{Master})]}}\end{cases}. \tag{8}$$

Otherwise, he decides whether he will cooperate with the master

$$s_i=\begin{cases}\textbf{Cooperate}, & \text{if}\quad G_{p_i}\geq\frac{E_{c1}(1-\phi_1)+E(\textbf{Cooperate})\phi_1\left(1-\frac{1}{2}\phi_2\right)}{\phi_1\phi_2E^{T_j}(\mathbf{B})}\\ \textbf{Decline}, & \text{if}\quad G_{p_i}<\frac{E_{c1}(1-\phi_1)+E(\textbf{Cooperate})\phi_1\left(1-\frac{1}{2}\phi_2\right)}{\phi_1\phi_2E^{T_j}(\mathbf{B})}\end{cases}, \tag{9}$$

where $\phi_1=(1-\prod_{k\neq i}\eta_k)$ and $\phi_2=\prod_{j\neq i}(1-p_j)$. If he find out that the master behave maliciously during cooperation, he will report this to the crowd sensing platform.

A player will decide to be a master only when the expected utility of being a master is greater than not to maximize his utility. Then, he will behave maliciously if $\mathbb{E}[U_i(\textbf{Deceive};\rho_i,S_i,S_{-i},\mathcal{F})]>\mathbb{E}[U_i(\textbf{Master};\rho_i,S_i,S_{-i},\mathcal{F})]$, which gives out the threshold in Equation (8). The other condition in Equation (9) can be derived in a similar manner. These equilibriums indicate that when the additional cost to be a master is smaller and/or fewer players present in the game, a player is more likely to become a master. It satisfies the intuitive knowledge that players want to protect privacy with fewer costs.

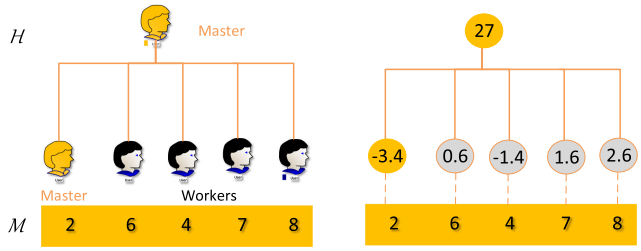


Fig. 5. Decomposition tree.

Theorem 2. *For $supp(\mathcal{F}) \subseteq [0,1]$, there exist N pure strategy equilibriums that exactly one player chooses* **Master** *or* **Deceive** *while all other players choose* **Cooperate** *or* **Decline**.

Proof. If all of $W_i^{T_j}$'s opponents choose **Cooperate** or **Decline**, $W_i^{T_j}$'s best response is **Master**. Otherwise, $W_i^{T_j}$ will receive a payoff of no more than 0. Because players are symmetric, it does not matter which player is the one that chooses **Master** or **Deceive**. Hence, there exist N pure-strategy equilibriums. □

The privacy goal fails only if the only master is deceiving while the other players are cooperating with the master. A new round of game will start without this master if all other players choose to decline.

Theorem 3. *The probability of failure is smaller with larger the privacy gain G_{p_i}.*

Proof. The probability of failure is

$$\Pr[\text{failure}] = \left(1 - \prod_{k \neq i} \Omega_i\right) \times \Omega',$$

where $\Omega_i = \Pr\left[G_{p_k} > \frac{E_{c1}(1-\phi_1)+E(\textbf{Cooperate})\phi_1\left(1-\frac{1}{2}\phi_2\right)}{\phi_1\phi_2 E^{T_j}(\mathbf{B})}\right]$ and $\Omega' = \Pr\left[p_i > \frac{G_{p_i}[1-\gamma_i(\textbf{Master})]}{G_{p_i}[1-\gamma_i(\textbf{Master})]+G_{d_i}[1-\gamma_i(\textbf{Deceive})]}\right]$. Ω_i and Ω' are one-sided p-values given by the complementary cumulative distribution functions.

The observed value $\frac{G_{p_i}[1-\gamma_i(\textbf{Master})]}{G_{p_i}[1-\gamma_i(\textbf{Master})]+G_{d_i}[1-\gamma_i(\textbf{Deceive})]}$ in Ω' is monotonically increasing w.r.t. G_{p_i}, so Ω' is negatively correlated with privacy gain. $(1 - \prod_{k \neq i} \Omega_i)$ becomes smaller with larger gain because Ω_i is a p-value increasing monotonically w.r.t. G_{p_i}. Therefore, the increasing-decreasing characteristics of $\Pr[\text{failure}]$ are proved. □

5.3 Group Location Privacy and Sensing Data Perturbation

In this subsection, we discuss the details of how privacy goals are fulfilled. Our scheme proceeds to privacy preservation after the master is chosen. In one group, locations of all members satisfy a group geo-indistinguishability related to the maximum radius requirement for geo-indistinguishability in the group. The noise level data of each individual also confronts to differential privacy w.r.t. locations.

5.3.1 Group Geo-Indistinguishability

As mentioned in the introduction, decreasing each worker's privacy budget for a low composed one endures large variances in workers' perturbed data. In our design, the master and his group members deploy the differential privacy mechanism via wavelet transformation [17] so that the achieved group geo-indistinguishability has much lower privacy bound than simply applying geo-indistinguishability to each worker without bringing about great variances. We first decompose a location in polar coordinates to wavelet coefficients and add noises to these coefficients. Then the perturbed location can be reconstructed from noisy coefficients.

Tree Construction. Since radius and angle components of a location are independent in polar coordinates, the wavelet transform of a location can be viewed to be separate for its radius and angle. For radius decomposition, each worker initially reports a random integer referring to the radius component of his location in polar coordinates to the master. The master collects integers from workers and arranges them as a vector **M**. A group hierarchy H connects the master and cooperative workers as shown in Fig. 5, which indicates that the master can receive information from all workers and is responsible for processing the received information. Based on H, a nominal wavelet decomposition tree R is constructed from H by attaching a child node Leaf_{c_i} to each leaf node Leaf_i in H. The value of Leaf_{c_i} is the same as the corresponding entry in **M**. The nodes except leaf children are wavelet decomposition coefficients, which are computed differently: The wavelet coefficient for the root node (referred to as the base coefficient) is set as the sum of all leaves in its subtree (the leaf-sum of the node), while the coefficient for any other internal node equals its leaf-sum minus the average leaf-sum of its parent's children. We illustrate this step with an example in Fig. 5. The left side is the hierarchy H and random numbers in **M**. The master occupies both the root node and one leaf node to manage group data and fulfill privacy, while members take the remaining $N-1$ leaf nodes. The decomposition tree constructed from H is to the right of H. We will continue using this example in the following sections.

Obfuscation. Next, a planar Laplace noise [11] is added with parameter ϵ_l to each decomposition coefficient of a location

$$D_{\epsilon_l}(x_0)(x) = \frac{(\epsilon_l W_{\text{Nom}}(c_{i,j}))^2}{2\pi} e^{-\epsilon_l r W_{\text{Nom}}(c_{i,j})}, \tag{10}$$

where r is the distance between x and x_0, which is, in other words, the noise to perturb location x_0 to x. $W_{\text{Nom}}(c_{i,j}) = 1$ if $c_{i,j}$ is the base coefficient, otherwise $W_{\text{Nom}}(c_{i,j}) = f_j^o/(2f_j^o - 2)$, where f_j^o is the fan-out of $c_{i,j}$'s parent in the decomposition tree (e.g.,the fan-out f_0^o of the root node is 5 in Fig. 5). Due to the independence of radius and angle, the marginal probabilities of $D_{\epsilon_l}(x_0)(x)$ are also independently, whose probability density functions are $D_{\epsilon_l,R}(r)$ and $D_{\epsilon_l,\Theta}(r)$ respectively. The angle is randomly selected in $[0, 2\pi)$, so we only describe how each worker computes his radius noise and adds it to his occupied leaf coefficient. In that sampling a random number from $D_{\epsilon_l,R}(r)$ is complicated, the inverse transform sampling is deployed to draw a value z uniformly in $[0,1)$ and the noise is computed as $r = C_{\epsilon_l}^{-1}(z)$, where $C_{\epsilon_l}(r)$ is the cumulative function of $D_{\epsilon_l,R}(r)$ and $C_{\epsilon_l}^{-1}(z)$ is its inverse function

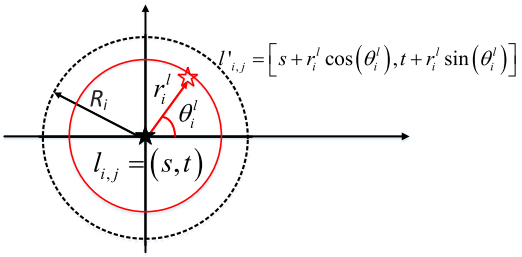


Fig. 6. Location obfuscation.

$$D_{\epsilon_l,R}(r) = \int_0^{2\pi} D_{\epsilon_l}(r,\theta)d\theta = (\epsilon_l W_{\text{Nom}}(c))^2 re^{-\epsilon_l r W_{\text{Nom}}}$$
$$C_{\epsilon_l}(r) = 1 - (1 + \epsilon_l r W_{\text{Nom}})e^{-\epsilon_l r W_{\text{Nom}}}.$$

Then, the noisy coefficients $\hat{c}_{i,h-1}$ are produced with its dedicated noise r_i as $\hat{c}_{i,h-1} = r_i + c_{i,h-1}$. Apart from the leaf coefficient noise, the master computes one more noise for the base coefficient and announces the noisy base coefficient $\hat{c}_0$ to other workers. All noises are normalized with $W_{\text{Nom}}(c)$. The computation and annunciation of noisy angle coefficients follow the same fashion. After that, workers reconstruct their perturbed components from noisy coefficients following the equation:

$$v_i = \hat{c}_{i,h-1} + \sum_{j=0}^{h-2}\left(\hat{c}_{i,j} \cdot \prod_{k=0}^{h-2} \frac{1}{f_k^o}\right) = \hat{c}_{i,1} + \hat{c}_{i,0}/f_0^o, \tag{11}$$

where h is the height of H, $c_{i,h-1}$ is the ancestor of v_i in the h level of the decomposition tree, and f_k^o is the fan-out of $c_{i,k}$.

A worker subtracts his originally reported number from v_i and gets his radius noises r_i^l. The angle noise θ_i^l is available in the same way. A new location is mapped from the actual location by adding component noises to coordinates as illustrated in Fig. 6: suppose a worker $W_i^{T_j}$'s actual location when conducting task T_j is $l_{i,j} = (s,t)$, his falsified location is $l'_{i,j} = [s + r_i^l \cos(\theta_i^l), t + r_i^l \sin(\theta_i^l)]$. Only the worker himself knows the total noise, so both the master and the server cannot learn his actual location from the obfuscated one even if they collude with each other.

Theoretical Analysis. We analyze some mathematical properties of group geo-indistinguishability here.

Theorem 4. *The group geo-indistinguishability based on wavelet transform and Laplace noise has a generalized sensitivity of $2R_{\max}$ with respect to W_{Nom}. It can achieve $2\epsilon_l R_{\max}$-geo-indistinguishability.*

Proof. According to Lemma 4 in [17], the nominal wavelet transform has a generalized sensitivity of h with respect to W_{Nom}, where h the height of the hierarchy associated with the input frequency matrix. If we expand the concept of generalized sensitivity from matrix differs only at one tuple to locations in a circle with radius R, the generalized sensitivity is also expanded to hR. In the decomposition tree, the maximum sensitivity that can be generated by a leaf node is $R_{\max}$. Thus the decomposition tree can have a sensitivity Δf of $2R_{\max}$, where 2 is the height of the tree in our paper. Therefore, the achieved privacy level is $\Delta f/\lambda = 2R_{\max}/\lambda = 2\epsilon_l R_{\max}$. □

Starting from a tuple of locations $\boldsymbol{l} = (l_1 \ldots, l_n)$, if we independently apply $\epsilon_l R_i$ geo-indistinguishability to l_i, the achieved privacy level is $\sum \epsilon_l R_i$ due to composition property. From Theorem 4, the geo-indistinguishability level achieved here is far lower.

Theorem 5. *The variance of the added noise increase when the fan-out f augments, but the variance of the noise in the answer is always less than $8\lambda^2$.*

Proof. In Equation (10), the new ϵ'_l equals to $\epsilon_l W_{\text{Nom}}$. Then, for the scale in Laplace noise

$$\lambda' = \frac{1}{\epsilon_l W_{\text{Nom}}} = \frac{\lambda}{W_{\text{Nom}}}. \tag{12}$$

As W_{Nom} is monotonically decreasing to f, the variance $2\lambda'^2$ increases.

The variance of one reconstructed element following Equation (11) is at most

$$\left[(1-1/f)^2 + 1/f^2(f-1)\right] \cdot 4(1-1/f)^2 \cdot 2\lambda^2$$
$$= 4(1-1/f)^3 \cdot 2\lambda^2,$$

whose approximation is $8\lambda^2$ for $f \to \infty$. □

5.3.2 Sensing Data Perturbation

In this step, the privacy goal is: a worker $W_i^{T_j}$'s reported noise level cannot be distinguished from others collected within distance R_i with probability bounded by a budget related to ξ. The total sound pressure level generated by N sources follows Equation (2) and the SPL for each source follows Equation (1). We assume for simplicity that in the noise monitoring task for one noise source, the sound field is a free field. Thus, for one location which is κ km away from the source and another random location whose distance to the noise source differs from κ by r_{rand}, the difference of measured noise levels shall be $\left|10\log\frac{4\pi(\kappa \pm r_{\text{rand}})^2}{4\pi\kappa^2}\right| = \left|10\log\left(1 \pm \frac{r_{\text{rand}}}{\kappa}\right)^2\right|$. Again, the noise controlling application is considered. The server may announce some known noise sources and encourage workers to measure the noise levels around the sources. If κ is unknown to the worker, it is possible for the worker to guess the distance by sight if the noise source is close enough. Otherwise, the worker chooses a κ larger than the protection range r to confine the accuracy loss rising from imprecise distance.

Since the worker $W_i^{T_j}$ wants to protect his location (κ_i km from the source) within R_i, he assigns his original data $X_{i,j}$ with a Laplacian noise $\text{Lap}\left(0, \left|\frac{10\log[\max(\zeta_1,\zeta_2)]}{\ln\xi}\right|\right)$ to get $X'_{i,j}$, where $\zeta_1 = (1 + R_i/\kappa_i)^2$, and $\zeta_2 = (1 - R_i/\kappa_i)^2$. $X'_{i,j}$ can achieve $|\ln\xi/10\log[\max(\zeta_1,\zeta_2)]|$-differential privacy in the dataset of possible noise levels collected within distance R_i.

Theorem 6. *The data perturbation mechanism given above satisfies the privacy goal and accuracy demand*

$$\Pr\left[|X'_{i,j} - X_{i,j}| \geq |10\log[\max(\zeta_1,\zeta_2)]|\right] \leq \xi. \tag{13}$$

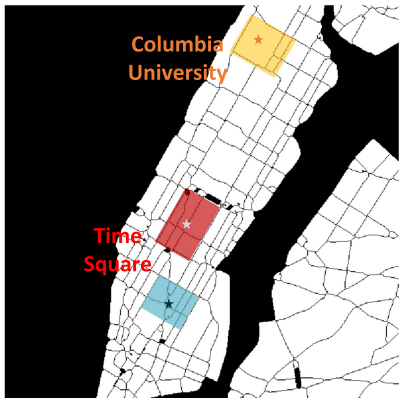


Fig. 7. Manhattan Map segments and experiment regions.

Proof.

$$|10\ \log\left[\max(\zeta_1,\zeta_2)\right]| = \alpha,\ N_i \sim \mathrm{Lap}\left(0, \frac{\alpha}{|\ln \xi|}\right),$$

$$\Pr\left[\left|X'_{i,j} - X_{i,j}\right| \geq \frac{\alpha}{|\ln \xi|}\right] = \Pr\left[|N_i| \geq \frac{\alpha}{|\ln \xi|}\right]$$
$$= 2\int_{\alpha}^{\infty} \frac{|\ln \xi|}{2\alpha} \exp\left(-\frac{|\ln \xi| x}{\alpha}\right) dx = \xi.$$

□

5.3.3 Energy Analysis

In the tree construction step, every member only exchanges information with the master once, whereas the master needs to receive/deliver $N-1$ packets from/to members and compute values of $N+1$ nodes in the decomposition tree. In the perturbation step, every worker spends energy on computing noises. The master's computation burden is as twice as a member's since he occupies two nodes. Also, the master has additional communication cost originated from sending noisy roots to $N-1$ members. The last part of energy consumption stems from sensing data perturbation, which is the same for all workers. Hence, we can derive the ratio of energy consumed by a member and the master after summing them up as

$$\frac{E(\textbf{Cooperate})}{E(\textbf{Master})} \approx \frac{E_{c1} + E_{c2} + E_d}{3(N-1)E_{c1} + (N+3)E_{c2} + E_d},$$

which corresponds to what we proposed in Section 5.2.2.

6 Evaluation

6.1 Experiment Settings

We implement our scheme on a PC with an Intel Core i5 CPU running at 3.20 GHz and 8 GB RAM and use the real-world datasets, noise collected by Lubin Liu from Microsoft [18], [19], [20] and Yelp [21] to evaluate our scheme. In [18], [19], [20], authors provide noise level data at 36 locations collected by six users on Manhattan and draw noise heat maps for weekdays and weekends. The fine-grained noise situation is reflected. Usually, areas with more points of interest (venues in a physical world, like a shopping mall or theatre, having a name, address, coordinates, category, and other attributes [18]) expose to more severe noise situations, and thus, demand more intense noise monitoring. Therefore, we choose regions near Time Square (red), Columbia University (yellow), and Washington Square (blue), divided by streets and avenues, as task regions in Fig. 7. Each location entry in regions is considered to be reported by an individual worker.

Besides this small crowd sensing dataset, we implement another location dataset to evaluate the performance of our scheme on a large number of workers. The Yelp [21] dataset contains 70,745 user check-in locations, where check-ins are logged incontinuously at a relatively low frequency. It fits our scenario where workers carry out tasks and only report their locations when presenting at specific spots.

According to [22] regarding smartphones' energy consumptions, the energy consumed under IEEE 802.11 when transmitting data at 700 kb/s and staying in a dynamic connection cost 31 mA and 2 mA respectively. The power consumptions of active CPU can be between 100 and 200 mA. The total battery capacity $E(\mathbf{B})$ is 3000 mAh. Usually one wireless data packet is sized at 8 kB. We suppose that a worker has equal probability to use LTE and WiFi and each round of computation has around 1,000 float operations, so the settings of unit energy deduced from above statistics are: $E_{c1}/E(\mathbf{B}) = 6.67e-4$ for one second, $E_{c2}/E(\mathbf{B}) = 1.18e-5$, $E_d/E(\mathbf{B}) = 0.067$ for one second. The distribution $\mathcal{F}$ of privacy demands follows Beta distribution $\mathrm{B}(0.149, 0.109)$ according to the analysis of social privacy demand distribution in [16]. For the clarity of analysis, we confine ϵ_l to the range from 0.3 to 10 [23], under which most existing algorithms are evaluated.

6.2 Evaluation Metrics

We define the following metrics in our experiments to quantify the performance from aspects of PCG, location obfuscation, and data perturbation.

Game Failure Ratio (GFR). It is the ratio describing how many failed PCG happen after performing certain rounds of games. It shows the effectiveness of PCG.

Drift Distances. The drift distance D_{dist} is the difference between a released location $l'_{i,j}$ and its corresponding actual location $l_{i,j}$. Both the mean and standard deviation (STD) are computed to measure the usefulness and stability of the location obfuscation. The mathematical formulation of its mean is defined as

$$D_{\text{dist}} = \frac{1}{N}\sum_{W_i:i=1}^{N} d(l_{i,j} - l'_{i,j}), \tag{14}$$

where $d(l_{i,j} - l'_{i,j})$ is the distance between $l_{i,j}$ and $l'_{i,j}$.

Satisfaction Ratio (SR). This metric calculates the ratio of the number of workers whose new locations are outside their radius requirements to the total number of workers.

Out-of-Range Ratio (ORR). This ratio is to measure the possibility that workers' perturbed locations are out of the required task region while they are actually residing inside.

Data Trustfulness. Data trustfulness is the probability that the shift from perturbed data to truth exceeds the bound claimed by (α, ξ)-accuracy. $\Pr[|X'_{i,j} - X_{i,j}| \geq \alpha]$ should not be greater than ξ.

We construct two application scenarios for the evaluation. One scenario is a crowd sensing application only containing

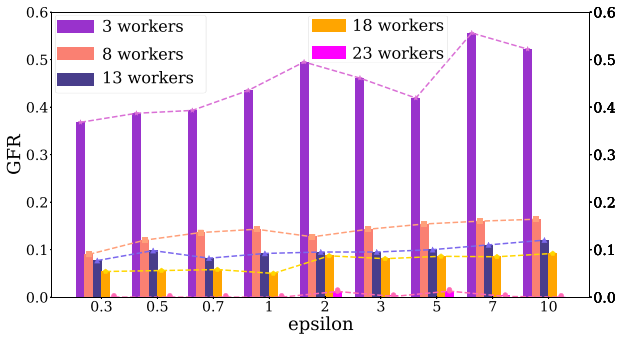


Fig. 8. Game failure ratio.

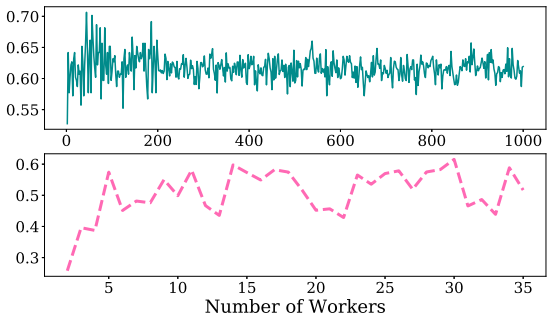


Fig. 10. Workers satisfaction ratio, $\epsilon_l = 1$.

location information and the other has both location and noise level data. The former is contributed by Yelp and Microsoft datasets and merely tested for two locations metrics. The latter is built on Microsoft dataset and provides extra performance assessment w.r.t. task completion and data accuracy.

6.3 Experimental Results

6.3.1 PCG Performance

Fig. 8 shows the GFR under varying ϵ_l and number of workers. Generally, it is climbing up with the growth of ϵ_l, which is in correspondence with Theorem 3. Since the probability that all workers choose to cooperate with a malicious master is smaller when more workers join in the game, GFR is declining if the number of participating workers increase. The GFR is almost 0 with over 20 workers, which proves the usefulness of PCG in practical applications.

6.3.2 Location Obfuscation Performance

The performance results of location obfuscation in the first scenario are shown in Figs. 9, 10 and 11. The group size of workers and the choice of ϵ_l can influence the performance.

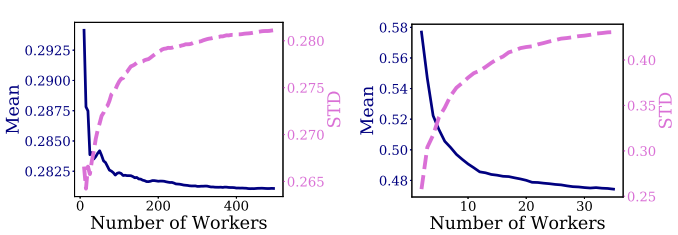
Impact of Group Size. Fig. 9 demonstrates how the group size influences the performance of drift distances when ϵ_l is fixed. In Fig. 9a, the drift distance averaging from a group of workers (10 to 500 workers) in Yelp approaches 0.3 km steadily, manifesting the stability of the PCG and the location obfuscation. On the contrary, the STD climbs up since the variances of added noises are growing with the group size. However, the STDs stay in the range confined by the upper bound, which indicates that the service quality regarding locations will not decline significantly even when there are many workers. The experiment on Microsoft dataset starts at the minimum number of workers (2 workers), which results in a much greater average drift in Fig. 9b than in Fig. 9a at the beginning. Nevertheless, it gradually approaches the similar mean value afterward. Though STDs are all slightly growing due to the accumulating variance, the increment is comparatively negligible to the group size. We can conclude that the PCG and the location obfuscation are stable with a relatively small drift.

As for the satisfaction ratio, more workers indicate a larger fan-out, which contributes to a larger but bounded variance according to Theorem 5. Thus, the satisfaction ratio in Fig. 10 fluctuates greatly due to the increasing variance, yet it will stay around 0.6 even when more workers participate in the task because of the upper bound of variances. In other words, around 60 percent of obfuscated locations from participating workers are located within the area required by their geo-indistinguishability: $\Pr[d(l_{i,j}, l'_{i,j}) \leq R_i] \approx 0.6$.

Impact of ϵ_l. The impact of ϵ_l is evaluate and summarized in Fig. 11 and the first row of Table 2. The increasing of ϵ_l means the downgrade of privacy demands, which is a straightforward explanation of the downward trends of average drift distances. A mathematical explanation for changes in the STD is that the noise variance $\frac{2}{(\epsilon_l W_{\text{Nom}})^2}$ turns to be smaller, so the added noise is closer to 0, indicating that the obfuscated location is more like to be in close physical proximity of the true location. The satisfaction ratio in Table 2 is a weighted average of results from two datasets since its trend is the same for two datasets. The influences from privacy demands and ϵ_l are combined: most workers' privacy demands are comparatively small or large as they follow the Beta distribution; the growing of ϵ_l scales down the drift distance but meanwhile reduces all workers' radius requirements. Therefore, the distribution of satisfaction ratio follows neither a complete Beta distribution nor the variety of ϵ_l.

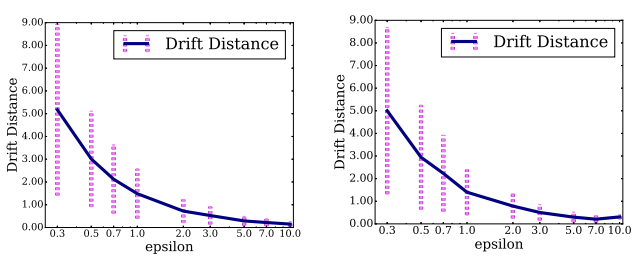
6.3.3 Data Accuracy and Task Completion

The second application is run for analyzing how our scheme can affect the completion of mobile crowd sensing tasks. Besides drift distance and satisfaction ratio that have



(a) Yelp (b) Microsoft

Fig. 9. Drift distances with different group size, $\epsilon_l = 3$.



(a) Yelp (b) Microsoft

Fig. 11. Drift distances under different ϵ_l.

TABLE 2
Impact of ϵ_l on Metrics

ϵ_l	0.3	0.5	0.7	1	2	3	5	7	10
Satisfactory Ratio	0.625616	0.807882	0.660099	0.605911	0.596059	0.522167	0.464532	0.423153	0.589655
ORR	0.972	0.722	0.671	0.556	0.364	0.176	0.111	0.090	0.092

already been discussed, we quantize the effects with other two metrics, ORR and data trustfulness. The group size is set as a constant here for valid evaluations. As revealed in the second row of Table 2, the higher possibility of close physical proximity stems from larger ϵ_l benefits the out-of-range ratio, implying that workers' reported false locations are less possible to be out of task region and thus the completion of a crowd sensing task is more assured.

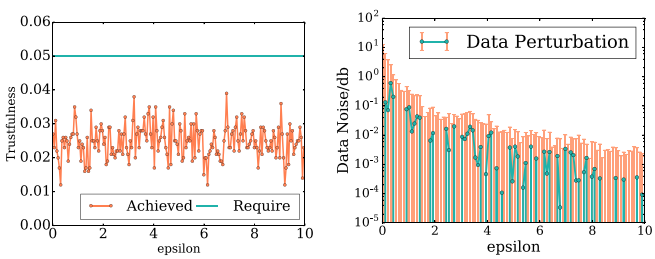
We also compute the trustfulness of data, which reflects the data accuracy after perturbation. More specifically, it is the probability that the distance between the perturbed data and the truth is smaller than ξ. Theoretically, the data trustfulness is bounded by Theorem 6. Here, ξ is set to be 0.05 in our simulation. The average trustfulness under each ϵ_l is about 0.25 ± 0.15 in Fig. 12a, which is much lower than the requirement ξ depicted by the green line. So the accuracy demand for crowd sensing tasks is fulfilled.

Fig. 12b shows the levels of noises added during data perturbation. They are the means and STDs of the numerical differences between perturbed data and estimated noise level data from Equation (2) under different settings of ϵ. We can tell from the graph that the average difference is only 1 dB with a choice of ϵ as small as 0.01, and the value goes even smaller with a larger privacy budget ϵ.

From both Figs. 12a and 12b we can tell that trustfulness is not evidently affected by ϵ_l while the accuracy of data is boosted with the growth of ϵ_l (lower privacy demand). In all, noises added to the sensing data can be lower than 3 dB when ϵ_l is greater than 0.7. To conclude from the analysis above, our scheme will not impede task objectives with a proper choice of ϵ_l to balance accuracy and privacy.

6.3.4 Time Efficiency

The time costs for PCG, location obfuscation, and data perturbation are listed in Table 3. Obviously, the time consumed only relates to the number of workers. However, since the individual computations in PCG and obfuscation processes are parallel, the time differences are relatively subtle. The time cost of noise data perturbation remains constant at around 2.65 ms. Overall, our scheme will not bring much burden to mobile crowd sensing in terms of time consumption.



(a) Data trustfulness (b) Data perturbation results

Fig. 12. Under different ϵ_l for $\xi = 0.05$.

TABLE 3
Time Costs

	3 Workers	13 Workers	23 Workers
PCG	91.398 ms	106.012 ms	116.531 ms
Location	5.771 ms	5.721 ms	5.970 ms
Data	2.657 ms	2.651 ms	2.613 ms

7 Related Work

Location privacy preservation is widely considered in various applications. The approaches can be categorized [24] as cryptographic methods, k-anonymity, and differential privacy.

7.1 Cryptographic Approaches

Cryptographic or private information retrieval (PIR) approaches use searchable encryption, asymmetric cryptography, and private proximity testing [25], [26] to provide location privacy for users. However, they do not address inference disclosures, and their schemes often incur high computation costs. Hence, they cannot be implemented directly on mobile devices due to their high memory and energy costs. Moreover, the curious service provider will still try to decrypt users' location coordinates even if they are encrypted, which will cause widespread network latency because of resource-consuming decryption operations [27].

7.2 k-Anonymity and Its Extensions

In [28], Sweeney et al. propose the notion of k-anonymity, which provides protection in a way that the information of each person contained in the released dataset cannot be distinguished from at least k individuals' information in the dataset. Nonetheless, an attacker can discover the values of sensitive attributes when there is little diversity in them. The k-anonymity approaches cannot protect against attackers with sufficient background knowledge. There are some extensions of k-anonymity to handle these deficiencies like l-diversity [29], p-sensitive [30], and t-closeness [31]. However, methods in the class of k-anonymity face plenty of shortcomings, such as relying on trusted intermediaries, offering limited privacy guarantees, and revealing approximate real-world locations to the servers in plaintext. In [32], authors combine the concepts of differential privacy and k-anonymity to propose the notions of query-indistinguishable k-anonymity and differentially private k-anonymity (DPkA) for query privacy in location-based service. In their algorithm, the privacy budget ϵ is minimized and the major breach of traditional k-anonymity among k queries reported to the service provider is overcome. However, this work only protects the query privacy and its application is limited.

7.3 Differential Privacy

Other than methodologies mentioned above, differential privacy [9] is a new way to bound the probability of distinguishing between two databases, which has been applied in location-based services to protect location privacy crowd sensing. Some works use differential privacy to protect workers' locations when performing tasks [33], [34], while some other works protect the aggregated data in a crowd sensing task [35]. In particular, existing differentially private crowd sensing usually unrealistically relies on a trusted entity. In [36], the authors propose a new definition, 'δ-location set' based differential privacy, to account for the temporal correlations in location data and a new notion, sensitivity hull, to bound the error of differential privacy. This work does not need a trusted entity, but it just protects a single trajectory. Our scheme effectively protects a group of workers' privacy in one region without a trusted third party.

Though privacy-preserving schemes provide privacy gains to mobile workers, they are so costly that mobile workers may not want to pay for privacy preservation. Some works are inspired by the concept of game theory to tackle the conflict between costs and privacy gains [16], [35], [37]. Compared to their works, our scheme does not need complete knowledge and takes more factors like data trustfulness into consideration to achieve higher efficiency for mobile crowd sensing.

8 Conclusion

In this paper, we propose a game-theoretical approach that implements differential privacy in location-based crowd sensing services. It fulfills indistinguishability for both locations and sensing data, such that sensitive location information leakage can be effectively restricted. The Bayesian game is formulated with regard to the privacy gains and costs, and the Bayesian Nash Equilibrium is derived from the game. Then, crowd sensing workers' privacy is protected via differential privacy. Based on theoretical analysis and evaluations on the real-world dataset, we have shown that a sufficient privacy guarantee is achieved and demonstrated the efficiency and accuracy of our scheme.

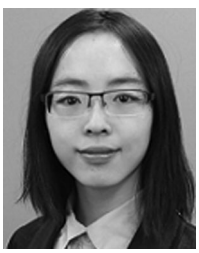
Acknowledgments

The work of L. Guo was partially supported by the National Science Foundation under grants IIS-1949640 and ECCS-1949639. The work of M. Li was partially supported by National Science Foundation under grants ECCS-1849860 and CNS-1924463.

References

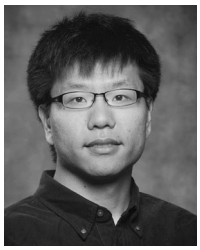
[1] X. Li and D. W. Goldberg, "Toward a mobile crowdsensing system for road surface assessment," *Comput. Environ. Urban Syst.*, vol. 69, pp. 51–62, 2018.
[2] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "Acoustic sensing based indoor floor plan construction using smartphones," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, 2017, pp. 519–521.
[3] C. Chowdhury and S. Roy, "Mobile crowdsensing for smart cities," in *Smart Cities: Found. Princ. Appl.*, ch. 5, pp. 125–154, 2017.
[4] T. P. McAlexander, R. R. Gershon, and R. L. Neitzel, "Street-level noise in an urban setting: Assessment and contribution to personal exposure," *Environmental Health*, vol. 14, no. 1, 2015, Art. no. 18.
[5] E. Berger, L. Royster, D. Driscoll, J. Royster, M. Layne, and D. Johnson , "The Noise Manual," *J. Acoustical Soc. America - J ACOUST SOC AMER*, vol. 112, pp. 1–796, Nov. 2002, doi: 10.1121/1.1506691.
[6] M. Charlton, S. Fotheringham, and C. Brunsdon, "Geographically weighted regression," White Paper, Nat. Centre for Geocomputation, Nat. Univ. Ireland Maynooth, 2009.
[7] X. Zhang *et al.*, "Incentives for mobile crowd sensing: A survey," *IEEE Commun. Surv. Tut.*, vol. 18, no. 1, pp. 54–67, 1Q 2016.
[8] X. Zhang, P. Huang, L. Guo, and M. Sha, "Incentivizing relay participation for securing IoT communication," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2019, pp. 1504–1512.
[9] C. Dwork, "Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.*, 2008, pp. 1–19.
[10] R. J. Aumann, *Lectures on Game Theory*. Boulder, CO, USA: Westview Press, 1989.
[11] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2013, pp. 901–914.
[12] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptography Conf.*, 2006, pp. 265–284.
[13] G. Theodorakopoulos, R. Shokri, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services," in *Proc. 13th Workshop Privacy Electron. Soc.*, 2014, pp. 73–82.
[14] L. Yu, L. Liu, and C. Pu, "Dynamic differential location privacy with personalized error bounds," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2017, pp. 1–15.
[15] J. C. Harsanyi, "Games with incomplete information," *The Amer. Econ. Rev.*, vol. 85, no. 3, pp. 291–303, 1995.
[16] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving k-anonymity in location based services," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2013, pp. 2985–2993.
[17] X. Xiao, G. Wang, and J. Gehrke, "Differential privacy via wavelet transforms," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 8, pp. 1200–1214, Aug. 2011.
[18] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang, "Diagnosing New York city's noises with ubiquitous data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 715–725.
[19] T. Liu, Y. Zheng, L. Liu, Y. Liu, and Y. Zhu, "Methods for sensing urban noises," Microsoft Research, New York, NY, Tech. Rep. MSR-TR-2014–66, pp. 1–4, May 2014. [Online]. Available: https://www.microsoft.com/en-us/research/publication/a-noise-map-of-new-york-city/
[20] Y. Wang, Y. Zheng, and T. Liu, "A noise map of New York city," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.: Adjunct Publication*, 2014, pp. 275–278.
[21] Yelp dataset challenge. [Online]. Available: https://www.yelp.com/dataset_challenge, Accessed: Oct. 5, 2017.
[22] Power profile for Android. [Online]. Available: https://source.android.com/devices/tech/power/values, Accessed: Dec. 22, 2017.
[23] J. Murtagh and S. Vadhan, "The complexity of computing the optimal composition of differential privacy," in *Proc. Theory Cryptography Conf.*, 2016, pp. 157–175.
[24] P. Jagwani and S. Kaushik, "Privacy in location based services: Protection strategies, attack models and open challenges," in *Proc. Int. Conf. Inf. Sci. Appl.*, 2017, pp. 12–21.
[25] G. Sun, Y. Xie, D. Liao, H. Yu, and V. Chang, "User-defined privacy location-sharing system in mobile online social networks," *J. Netw. Comput. Appl.*, vol. 86, pp. 34–45, 2017.
[26] P. Kotzanikolaou, C. Patsakis, E. Magkos, and M. Korakakis, "Lightweight private proximity testing for geospatial social networks," *Comput. Commun.*, vol. 73, pp. 263–270, 2016.
[27] Y. Mao, Y. Zhang, X. Zhang, F. Xu, and S. Zhong, "Location privacy in public access points positioning: An optimization and geometry approach," *Comput. Security*, vol. 73, pp. 425–438, 2018.
[28] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty Fuzziness Knowl.-Based Syst.*, vol. 10, no. 05, pp. 557–570, 2002.
[29] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, 2007, Art. no. 3.

[30] A. Solanas, F. Sebé, and J. Domingo-Ferrer, "Micro-aggregation-based heuristics for p-sensitive k-anonymity: One step beyond," in *Proc. Int. Workshop Privacy Anonymity Inf. Soc.*, 2008, pp. 61–69.
[31] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. IEEE 23rd Int. Conf. Data Eng.*, 2007, pp. 106–115.
[32] S. Wang, Q. Hu, Y. Sun, and J. Huang, "Privacy preservation in location-based services," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 134–140, Mar. 2018.
[33] X. Jin and Y. Zhang, "Privacy-preserving crowdsourced spectrum sensing," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1–9.
[34] J. Wang, Z. Cai, C. Ai, D. Yang, H. Gao, and X. Cheng, "Differentially private k-anonymity: Achieving query privacy in location-based services," in *Proc. Int. Conf. Identification Inf. Knowl. Internet Things*, 2016, pp. 475–480.
[35] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proc. 17th Int. Symp. Mobile Ad Hoc Netw. Comput. MobiHoc*, 2016, vol. 16, pp. 341–350.
[36] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Security*, 2015, pp. 1298–1309.
[37] R. Shokri, G. Theodorakopoulos, and C. Troncoso, "Privacy games along location traces: A game-theoretic framework for optimizing location privacy," *ACM Trans. Privacy Security*, vol. 19, no. 4, 2017, Art. no. 11.



Pei Huang (S'17) received the BSc degree from Xidian University, Xi'an, China, in 2015 and the MSc degree from State University of New York at Binghamton, Vestal, NY, in 2017. She is currently working toward the doctorate degree in computer engineering at Clemson University, Clemson, SC. Her research interests include the security and privacy in eHealth/mHealth system, wireless networks, and crowdsensing, with a focus on the security problems regarding physical layer properties in the Internet of Things recently.

Xiaonan Zhang (S'16) received the BE degree in communication engineering from the Beijing University of Chemical Technology, Beijing, China, in 2012, and the MS degree from Binghamton University, State University of New York, Vestal, NY, in 2017. She is currently working toward the PhD degree at Clemson University, Clemson, South Carolina. She was a research assistant with the Beijing University of Posts and Communications from 2012 to 2015. Her research interests include resource management and security problems in wireless networks and the Internet of Things. She is a student member of the IEEE.



Linke Guo (M'14) received the BE degree in electronic information science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008, and the MS and PhD degrees in electrical and computer engineering from the University of Florida, Gainesville, FL, in 2011 and 2014, respectively. From August 2014 to August 2019, he was an assistant professor with the Department of Electrical and Computer Engineering, Binghamton University, State University of New York, Vestal, NY. Since August 2019, he has been an assistant professor with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. His research interests include network security and privacy, social networks, and applied cryptography. He is currently serving as the editor of the *IEEE Transactions on Vehicular Technology*. He also serves as the poster/demo chair of IEEE INFOCOM 2020. He was the publication chair of IEEE Conference on Communications and Network Security (CNS), in 2016 and 2017. He was the symposium co-chair of Network Algorithms and Performance Evaluation Symposium, ICNC 2016. He has served as the technical program committee (TPC) members for several conferences including IEEE INFOCOM, ICC, GLOBECOM, and WCNC. He is the co-recipient of Best Paper Award of Globecom 2015, Symposium on Communication, and Information System Security. He is a member of the IEEE and ACM.

Ming Li (M'14) received the BE degree in electrical engineering from Sun Yat-sen University, Guangdong, China, in 2007, the ME degree in electrical engineering from the Beijing University of Posts and Communications, Beijing China, in 2010, and the PhD degree in electrical and computer engineering from Mississippi State University, Starkville, MS in 2014, respectively. She was an assistant professor with the Department of Computer Science and Engineering, University of Nevada, Reno, Nevada from Aug. 2014 to Aug. 2018. She is currently an assistant professor in the Department of Computer Science and Engineering, University of Texas at Arlington TX. Her research interests include cybersecurity, privacy-preserving data analysis, resource management and network optimization in cyber-physical systems, cloud computing, mobile computing, wireless networks, smart grid, and big data. She is a member of the IEEE.