Differential Location Privacy for Crowdsourced Spectrum Sensing

Zonghao Huang
Oklahoma State University
Stillwater, Oklahoma 74078
Email: zonghao.huang@okstate.edu

Yanmin Gong
Oklahoma State University
Stillwater, Oklahoma 74078
Email: yanmin.gong@okstate.edu

***Abstract*—Dynamic spectrum access (DSA) enables secondary users (SUs) to access the underutilized licensed spectrum when the primary users (PUs) are absent and is a key solution to address the worldwide spectrum scarcity and improve the spectrum utilization. Database-driven DSA is a popular DSA paradigm and has been approved by FCC. In a database-driven DSA system, a spectrum service provider (SPP) accepts registrations from PUs and estimates the spectrum availability. To improve the accuracy of spectrum estimation in such a system, crowdsourced spectrum sensing (CSS) has been proposed, where the SPP recruits a large number of ubiquitous mobile users and outsources spectrum-sensing tasks to them. Although showing great potential, CSS requires the SPP to know the locations of mobile users for spectrum-sensing task allocation, which presents a serious privacy concern. In this paper, we propose an approach for protecting location privacy of mobile users while ensuring the sensing performance in CCS. The proposed approach is based on differential privacy and geocast, which allows the SSP to allocate tasks to mobile users without inferring their individual location information. Analytic models and task allocation strategies are developed in this paper to balance privacy, utility, and system overhead in CSS. Experimentation results based on real-world datasets show that the proposed approach can provide rigorous privacy protection to mobile users while providing effective services with low system overhead.**

I. Introduction

The proliferation of mobile devices and bandwidth-hungry applications has contributed to a explosive demand on wireless communications, which makes the wireless spectrum a scarce resource. As a key technique to address the worldwide spectrum scarcity and enhance the spectrum utilization, dynamic spectrum access (DSA) [1] has been proposed to allow secondary users (SUs) to opportunistically access the underutilized licensed spectrum when the primary users (PUs) are absent. As the PUs have a high priority to access the licensed spectrum, the harmful interference from the SUs to the PUs must be reduced or eliminated. Towards this goal, database-driven DSA [2] has been advocated by the FCC. In a database-driven DSA system, a spectrum service provider (SSP) accepts registrations from PUs and predicts the availability of the spectrum. All SUs are required to inquire the SSP first about the availability of the spectrum before using the spectrum. The spectrum availability is predicted by the SSP based on PUs' registered locations and transmission schedules together with radio propagation models. However, it has been shown in some recent measurement studies [3] that such database-driven DSA systems suffer from low estimation accuracy and are often overly conservative without considering the local environment.

With the emergence of the crowdsourcing paradigm [4], crowdsourced spectrum sensing (CSS) shows great promise in improving the accuracy of spectrum availability estimation in database-driven DSA system [5], [6]. In such an approach, an SSP recruits many ubiquitous mobile users and outsources the spectrum sensing tasks to them. These mobile users are equipped with mobile devices that are capable of performing spectrum sensing tasks, which is feasible due to the wide deployment of dynamic spectrum sensing and the relative low requirement for individual sensing accuracy. The advantages of using CSS lies in the following aspects. First, due to the ubiquitous penetration of mobile devices, they provide sufficient geographical coverage/spatial diversity for spectrum sensing requirement. Second, if secondary users are performing spectrum sensing, some secondary user may intentionally report false information in order to prevent other secondary users from using the idle spectrum, or just want to be "free riders" without contributing any sensing results [7], [8]. With crowdsourcing, sensing agents are recruited through an open call for the sensing task to collect rewards, and thus it is more difficult to launch targeted attacks against a specific spectrum sensing task.

Despite of the advantages, CCS poses several challenges. First, in order to allocate nearby spectrum sensing tasks to mobile users, the SSP needs to know the current location of mobile users. However, location information is considered to be sensitive, because it could reveal a lot of personal information such as daily activities, health status, social relationships, and even the identity of the mobile users [9]. Hence, location privacy of mobile users is a critical concern that needs to be addressed in order to encourage more participation. Second, spatial diversity of the set of the selected mobile users is also critical to the performance of a spectrum sensing task. Thus when selecting a group of mobile users for a specific spectrum sensing task, we need to consider whether the selected group provides sufficient diversity. This makes many previous location privacy mechanisms ineffective. For example, a popular way to protect location privacy is perturbation, however, it is difficult to guarantee diversity over a set of perturbed locations with reasonable privacy levels.

In this paper, we propose an approach that provides a privacy-preserving mobile crowdsourcing approach for spectrum sensing. Our approach provides differential privacy guar-

978-1-5386-0683-4/17/$31.00 ©2017 IEEE

antee for mobile users by only allowing the SSP to access sanitized location data of mobile users. We also enable the SSP to select mobile users based on their location diversity, which is critical to the performance of spectrum sensing. The key idea of protecting location privacy is exploiting the existing trust relationship between mobile users and the cellular service providers (CSP) that they subscribe to. Since the CSP already learns the location information of mobile users when providing the cellular service, they can serve as a trusted intermediate party for privacy protection. In this way, the CSP could sanitize location data from mobile users and transform it to the form that facilitates allocation of spectrum sensing tasks. Specifically, we adapt the private spatial decomposition approach proposed in [10] and utilize geocast routing protocols [11] to disseminate tasks without de-identifying mobile users in the sanitized location data. Our main contribution can be summarized as follows.

1) We identify the challenges of using crowdsourced mobile users for spectrum sensing and propose an approach for allocating tasks to mobile users without access to their individual locations.
2) We propose to use private spatial decomposition to represent mobile users' location data and design an approach to add differentially private noise that provides a good trade-off between privacy and utility.
3) Extensive experiments are carried out on real-world datasets to show the effectiveness of our proposed approach.

II. BACKGROUND

In this section, we give the introduction on differential privacy (DP), Private Spatial Decomposition (DSP) and geocast.

A. Differential Privacy

The privacy guarantee provided in our proposed framework is ϵ-differential privacy [12], [13]. Differential privacy can provide the protection of the dataset from attackers with arbitrary background. Specifically, differential privacy guarantees that when given a sanitized result, the adversary does not know whether the targeted individual is in the dataset or not. The differential privacy is defined as follows:

***Definition** 1 (Differential privacy):* A randomized mechanism M satisfies ϵ-differential privacy if for any two datasets D_1 and D_2 differing in only one tuple, and for all outputs $O \in \text{range}(M)$, the following inequity always holds:

$$\ln \frac{\Pr[M(D_1) = O]}{\Pr[M(D_2) = O]} \leq \epsilon, \tag{1}$$

where ϵ is a parameter to bound the ratio of probability distributions for two datasets. Specifically, ϵ indicates the amount of privacy protection in the mechanism, and it is usually regarded as the $privacy\ budget$. A private mechanism with a smaller ϵ gives better privacy protection.

In order to achieve the ϵ-differential privacy, a private mechanism should randomize the query result. There are many kinds of private mechanisms to achieve differential privacy, including laplacian mechanism, exponential mechanism and geometric mechanism.

In our setting, we need to consider a private mechanism with several stages of analyses M_i. The total privacy budget ϵ for the composition of M_i can be calculated by the following results [14]:

***Theorem** 1 (Sequential composition):* If $\{M_i\}$ are a set of analyses and each provide ϵ_i-differential privacy, then their sequential composition can provide $(\sum \epsilon_i)$-differential privacy.

***Theorem** 2 (Parallel composition):* If $\{M_i\}$ are a set of analyses and each provide ϵ_i-differential privacy, then their parallel composition can provide $\max \epsilon_i$-differential privacy.

B. Private Spatial Decomposition

The Private Spatial Decomposition (PSD) firstly introduced in [10] is an approach to construct a spatial dataset that guarantees differential privacy. A PSD provides a spatial index where each index node is referred to an associated spatial region, and the value for each index node is the noisy count of points located in this spatial region.

There are many types of data structures for the spatial index. The choice of data structure influences the data accuracy. In space-based decomposition such as grids and quad trees, the construction of spatial index is independent of the data point locations. Therefore, privacy budget is only consumed when noise is added to the count in each index node. The space-based decomposition is simple but it will cause the over-partitioning or under-partitioning when the data points are not uniformly distributed. In objective-based decomposition such as k-d trees [10], the spatial index is constructed based on the locations of data points. Specifically, partitioning the whole area into several index nodes requires the spatial distribution of the data points. Therefore, the privacy budget is consumed in both index node construction and noisy count. The objective-based decomposition can avoid the unbalance in partitioning but it is not robust as a slight change in PSD parameter will degrade the data accuracy greatly.

The Adaptive Grid (AG) approach is proposed in [15]. AG partitions the whole area into two-level grids. At the first level, the area is partitioned uniformly into several cells. The granularity of the second-level grids is based on the noisy count of the cells at the first-level grid. In this way, AG inherits the robustness and simplicity of space-based decomposition approach and can avoid the over-partitioning or under-partitioning problem.

C. Geocast

Geocast is a method of message dissemination in a allocated region [11]. There are two ways to implement the message dissemination: infrastructure-based mode and the infrastructure-less mode. In infrastructure-based mode, the message broadcaster send messages to each receiver located in the cast region. The communication cost is proportional to the number of notified receivers. In infrastructure-less mode, the messages can be relayed hop-by-hop among the receivers in the cast region, and an efficient way to relay hop-by-hop is to use geographic routing. In this case, the broadcaster do not need to know the exact locations of the receivers in the cast region and only need to send a few pieces of messages to some of them. Therefore, the communication cost of the latter mode is cheaper.

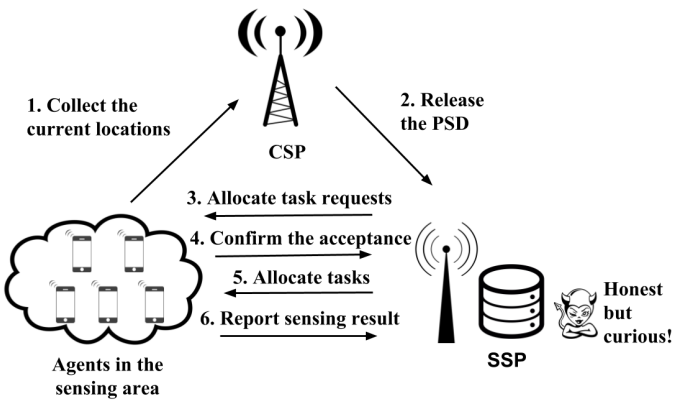


Fig. 1: Framework for task allocation in spectrum sensing.

III. System and adversarial model

A. System Model

In our proposed framework, there are three basic entities including spectrum service provider (SSP), cellular service provider (CSP) and mobile users (or sensing agents). The SSP needs to generate the database for DSA systems, and has a pool of sensing agents who sense the spectrum with advanced sensing capabilities. The CSP is a trusted party which has access to the true location information of sensing agents and help agents protect their location privacy. Here, it is reasonable to include CSP in our framework. Cellular companies like ATT have users' dynamical real locations, they are willing to provide such service if they get payed, and they are trusted because they sign contracts for location privacy protection. However, CSP could not replace SSP as there is no expertise or researcher on spectrum sensing in CSP.

Fig. 1 also shows how our framework works. Firstly, CSP collects location information of agents (step 1). Then, the CSP releases the private spatial decomposition of the collected location data to SSP (step 2). According to the spatial decomposition, SSP determines the regions for geo-cast and sends task requests to the sensing agents located in the cast regions (step 3). The agents who receive task requests from SSP respond to SSP if they accept the task requests (step 4). With the response from agents, SSP makes the agent selection among respondents and sends the task confirmation to the selected agents (step 5). The selected agents finish their assignments and send their sensing reports to SSP (step 6). By fusing the sensing reports from the agents, SSP can determine the availability of the requested spectrum and then inform the secondary users on the sensing result.

B. Spectrum Sensing Model

With add-on components to agents' smartphones, each sensing agent is able to detect the presence of primary user on a specific channel. According to [16], the signal strength detected by smartphone could be modeled as follows:

$$P_i = P_0(\frac{d_0}{d_i})e_i^X e_i^Y. \quad (2)$$

From the signal propagation model described above, there are three types of effects of the mobile communication ratio channel that limit the sensing performance, i.e., path loss, multipath induced fading and shadow fading [17], [18]. The effect of path loss is the major effect on the signal strength, but we do not need to take path loss into our consideration in our geo-cast and agents selection algorithm because we assume that the path loss is approximately the same for all sensors within the sensing region. Multipath fading is usually caused by atmospheric ducting, ionospheric reflection and refraction, and reflection from water bodies and terrestrial objects such as mountains and buildings. In practice, the multipath fading is usually ignored in sensing systems based on energy detection [19].

Shadow fading, also called large-scale fading is the result of obstacles blocking in the propagation path between the primary transmitter and the agents [18]. The shadow fading is spatially correlated, and we can model the correlation between two detectors as [18]:

$$R(d) = e^{-\alpha d}, \quad (3)$$

where α is an environmental factor and d is the distance between two detector. If the correlation of two detectors is small enough, we can say the two detector is uncorrelated. We define decorrelation distance as the minimum distance for uncorrelation.

After receiving sensing report from agents, SSP fuses the local sensing report by soft combining or hard combining for hypothesis test [20]. The decision error probability is related to the number of uncorrelated sensing report for fusion. Specifically, we set up a target for the decision error probability, the number of uncorrelated sensing reports should be equal to or more than the diversity order ($\overline{\text{div}}$) [21]. We need to select at least $\overline{\text{div}}$ uncorrelated agents for one sensing task to ensure the sensing performance.

C. Adversarial Model

We assume that the CSP is trusted and has access to location information of all sensing agents. We assume the SSP is "honest but curious", which means it would honestly follow the protocols to generate databases, but has the motivation to learn private location information of the sensing agents. We also assume that the communication channels among SSP, sensing agents and CSP are secure and reliable. However, we do not consider outsider attacks in this paper. Based on our assumption, there are two types of information leakage. The first type of information leakage happens during task allocation, when the adversary may infer the private location information of individual sensing agents from the aggregated statistics released by the CSP. This privacy leakage is critical since a large population of potential sensing agents may be involved. In our proposed privacy-preserving mechanism, we focus on how to prevent this type of information leakage and provide differential privacy guarantee for participating sensing agents. The second type of information leakage happens when the agents in the cast region accept the task requests, which discloses the real locations of agents. However, this type of information leakage is out of the scope of this paper

and may be mitigated through mechanisms such as location perturbation.

D. Design Goals

The goal of our proposed framework is to correctly detect the presence of primary user, which is directly affected by the correlation of the sensing reports. To achieve this goal, SSP sends task requests to the agents and selects some of the responding agents to perform the spectrum sensing. In order to protect the location information of agents, noises are added to the spatial index. The noisy counts will be overrated or underrated, which will affect the task request allocation to the agents and then affect the system overhead and assignment success rate. In our proposed framework, we focus on the following three performance metrics:

- **Overall Acceptance Rate**. Overall Acceptance Rate is the ratio of number of agents accepting task requests and the number of total task requests. Our major goal is to ensure that the overall acceptance rate is close to 100%.
- **System Overhead**. System overhead here includes the communication overhead and the computation overhead. Both of them are dependent on the average number of notified agents during the geo-cast process.
- **Overall Correlation of Reports**. Overall Correlation of Reports will directly affect the quality of the sensing result. We should guarantee the correlation between each pair of reports as small as possible.

IV. PRIVATE SPATIAL DECOMPOSITION CONSTRUCTION

It is of high cost for SSP to drop the task requests to the agents in the whole sensing area. Spatial decomposition [10] partitions the sensing area into several sub-regions and provides spatial indexes, which helps SSP efficiently choose some of regions for task request dropping rather than the whole area. In our system model, CSP has access to all the real locations of agents and constructs the spatial decomposition for SSP. In our spatial decomposition construction, we will use Adaptive Grid (AG) approach proposed in [15]. In this section, we will discuss on how to use AG in our construction.

A. Laplacian Mechanism Based AG

In the AG method, Laplacian Mechanism is widely used to add random noises to the counts. A two-level grid is used onto the area. In the first level, AG partitions the whole area into $m_1 \times m_1$ cells. m_1 is chosen as:

$$m_1 = \max\left(10, \left\lceil \frac{1}{4}\sqrt{\frac{N \times \epsilon}{10}} \right\rceil\right), \tag{4}$$

where N is the total number of agents in the whole sensing area and ϵ is the total privacy budget.

After the first-level grid calculation, we get the counts of agents for each cell in the first-level grid. Then we spend ϵ_1 in the mechanism, which adds noises to the counts to guarantee differential privacy for each cell. In this step, ϵ_1 is equal to $\alpha \times \epsilon$, where $\alpha \in (0, 1)$ decides how the total privacy budget ϵ is allocated into the two levels. Then AG further partition each level-1 cell into $m_2 \times m_2$ sub-cells. m_2 is chosen as:

$$m_2 = \left\lceil \sqrt{\frac{N' \times \epsilon_2}{\sqrt{(2)}}} \right\rceil, \tag{5}$$

where N' is the noisy count of one level-1 cell and ϵ_2 is remaining privacy budget by $(1-\alpha) \times \epsilon$. After creating the second level grid, the mechanism adds noises to the count for each level-2 sub-cell with the remaining privacy budget ϵ_2.

B. Geometric Mechanism Based AG

The widely used Laplacian Mechanism to add random noise has some limitations for our framework. Firstly, with Laplacian Mechanism, we cannot guarantee that all the noisy counts are integer number within the limited value range, e.g., $[0, 10]$, where will decrease the utility of data. Secondly, even if we round the noisy count results to the nearest integer and remap the infinite value range to the finite range for Laplacian Mechanism, the utility of data is still not optimal [22]. The task request allocation introduced in the next section requires the calculation of the expected diversity and overall acceptance rate, where the noisy counts in spatial index should have limited value range and be integer. Therefore, in order to maximize the utility, we choose the truncated α-geometric mechanism to add noises in AG, which is regarded as the universally utility-maximizing privacy mechanism for count query [22]. By truncated α-geometric mechanism, we reduce the expected loss by a factor of $\frac{1+\alpha}{2\sqrt{\alpha}}$ compared to the laplacian mechanism [22].

The truncated α-geometric mechanism add noises to the counts by the following distribution:

$$\begin{aligned}
&P[\delta < -c] = 0;\\
&P[\delta = -c] = \frac{\alpha^c}{1+\alpha};\\
&P[\delta = n] = \frac{1-\alpha}{1+\alpha}\alpha^n;\\
&P[\delta = N-c] = \frac{\alpha^{n-c}}{1+\alpha};\\
&P[\delta > N-c] = 0,
\end{aligned} \tag{6}$$

where δ is added noise, c is the true count result, N is the upper bound of the range, α is a constant within [0, 1] and chosen by $\alpha = \frac{1}{e^\epsilon}$, and $n \in (-c, N-c)$. Actually, the noisy result by truncated α-geometric mechanism can guarantee the ϵ-differential privacy.

In the construction of the private spatial, we still use two-level grid in the DSP construction. In the first level, we need to partition the whole region into $m_1 \times m_1$ parts. we can minimize the noisy error and the non-uniformity error by choosing m_1. As the noisy error is related to the standard deviation of the distribution in the truncated α-geometric mechanism, we need to consider the standard deviation when we choose the m_1. However, it is difficult to calculate the exact standard deviation for the distribution used in the mechanism as the standard deviation will change slightly with the different real counts and the upper bound N. Therefore, we use the standard deviation from the α-geometric mechanism, which is approximate to

the standard deviation for truncated mechanism, when we consider the granularity. And the standard deviation for the distribution in α-geometric mechanism is $\frac{\sqrt{2}}{\epsilon}$ the same as laplacian distribution. Then we choose m_1 as the laplacian mechanism by (4).

For the second level of grids, we further partition each 1-level region into $m_2 \times m_2$ sub-cell. As in the truncated geometric mechanism we can ensure that the noisy count is integer and non-negative, then m_2 is chosen by:

$$m_2 = \left\lceil \sqrt{\frac{N' \times \epsilon_2}{5}} \right\rceil, \tag{7}$$

where N' is the noisy count of one level-1 cell and ϵ_2 is remaining privacy budget by $(1-\alpha) \times \epsilon$.

V. TASK REQUEST ALLOCATION

In this section, we focus on how SSP selects the cast regions from the spatial decomposition. Firstly, we introduce the definition of expected diversity and acceptance rate. Then we describe how to construct geocast regions based on the private spatial decompositions.

A. Expected Diversity and Expected Average Correlation

In our spectrum sensing model, it is necessary for SSP to select uncorrelated agents so that we can reduce the shadow fading in the spectrum sensing. The targeted number of uncorrelated agents is called diversity order. During the geocast region construction, SSP need to select the regions for task request dropping, whose number of uncorrelated agents is at least $\overline{\text{div}}$ and the average correlation between them is small enough. However, SSP doesn't know the real location of each agent in a sub-cell from the spatial decomposition, so SSP doesn't know the number of uncorrelated agents and also their correlations in each sub-cell. Therefore, given a spatial index, SSP should consider the expected diversity and the expected average correlation for each sub-cell when SSP constructs the geo-cast region.

Expected diversity is defined as the expected number of uncorrelated agents in a given region with a certain count number. As the real location of each agent is unknown in a sub-cell from the spatial decomposition, we assume that each agent is uniformly, identically and independently distributed(i.i.d.) in their located region. If we choose two agents from a given region, then the expected correlation between this two uniform distributed agents could be calculated by $\int\int R(|x_1 - x_2|) f(x_1) f(x_2) dx_1 dx_2$, where x_1 and x_2 denote the 1^{th} agent and 2^{th} agent, $R(\cdot)$ is the correlation function by (3) and $f(\cdot)$ is the location distribution. When we select more agents in the region, then the expected average correlation among them will increases as they are closer in the region. Therefore, the expected average correlation (EAC) is related to the number of agents we choose in the region, and it can be modeled as:

$$\text{EAC}(k) = \int ... \int \dot{R}(k) f^k(x_i) dx_1 ... dx_k, \tag{8}$$

where $\dot{R}(k) = \frac{\sum_{i=0}^{k-1} \sum_{j=i+1}^{k} R(|x_i - x_j|)}{C_k^2}$, $f(\cdot)$ is the distribution function, and x_i denotes i^{th} agent.

After getting the expected correlation from the equation above, we can define the expected diversity as follow:

***Definition** 2 (Expected diversity):* In a given region c_i with n_i number of agents, the expected diversity of this region is k if EAC(k) is less than or equal to $R(d_0)$ and EAC$(k+1)$ is larger than $R(d_0)$ for $k < n_i$, or EAC(k) is less than or equal to $R(d_0)$ for $k = n_i$, where d_0 is the decorrelation distance and $R(\cdot)$ is the correlation function.

***Proposition** 1 (Composition):* If two expected diversities of two regions c_1, c_2 are d_1 and d_2, then the expected diversity of $\{\ c_1 \cap c_2\ \}$ is equal to d_1 + d_2.

Actually, when the number of agents we choose in the region becomes large, it is much more complicated to calculate the expected average correlation EAC(count) by (8), then it is also complicated to get the expected diversity. However, as the expected average correlation is propositional to $\frac{n}{l}$, where n is the count number and l is the size of the region. We can use the ratio $\frac{n}{l}$ to indicate how large the expected average correlation for a region is. Besides, in [17], a simulation on the number of uncorrelated sensors is given, which suggests that the number of uncorrelated sensors is a random variable with a Rayleigh distribution with the parameter $\frac{R}{d_0}$, where R is the radius of the given region and d_0 is the decorrelation distance. So according to the feature of Rayleigh distribution, the expected number of uncorrelated sensors is approximately equal to $1.253\frac{R}{d_0}$. Specifically, for a given region with a certain number of agents inside, the expected diversity can be calculated by:

$$E[\text{div}] = \min(1.253\frac{R}{d_0}, \text{count}). \tag{9}$$

In order to have a good sensing performance, we need to guarantee the diversity order in the cast region and minimize the average correlation. Therefore, we need to choose the region for geo-cast, where the expected diversity is high enough and the expected average correlation is as low as possible.

B. Overall Acceptance Rate

When the SSP constructs the geocast region, it needs to ensure that the probability of at least N agents accepting the task requests is high enough for a successful task. We assume that the individual acceptance rate (IAR) of each agent is the same, which is denoted by p.

In a given region, the overall acceptance rate of at least k agents accepting among m agents could be modeled as:

$$\text{OAR}(k, m) = 1 - \sum_{k-1}^{i=0} C_m^i (p)^i (1-p)^{m-i}. \tag{10}$$

When we construct the geocast region, the overall acceptance rate for the cast region should be equal to or larger than the threshold $\overline{\text{OAR}}$, which guarantees that enough number of agents located in the cast region would accept the task requests.

C. Geocast Region Construction

When SSP receives the spatial decomposition from CSP, SSP needs to select regions from the spatial decomposition, where some sub-tasks will be dropped. During the cast region

construction, we need to ensure the quality of our geocast and keep the overhead as small as possible. In details, we need to guarantee that the expected diversity of the cast region by (9) and the overall acceptance rate by (10) are both high enough. Besides, the number of notified agents located in the cast region and the expected average correlation for the cast region should be minimal.

In order to achieve the above goal, we design the following algorithm. The algorithm takes the private spatial decomposition as the input, and outputs the geocast regions where the task requests will be dropped. Firstly, the cast region is initialized as an empty set. Then the expected diversity of each sub-cell and ratio $\frac{n}{l}$ for the sub-cell are calculated. We order all the sub-cells by the ratio $\frac{n}{l}$, then pick the sub-cell with smallest ratio into the cast region. The overall acceptance rate for the cast region is updated. We continue to add the next sub-cell from the ordered set into the geo-cast region; if the overall acceptance rate is equal to or larger than the threshold, then we stop our construction and output the cast region. Algorithm 1 gives the details of our proposed greedy algorithm for the geocast region construction.

Algorithm 1 Greedy Algorithm to construct geocast region

Require: PSD, $\overline{\text{OAR}}$, $\overline{\text{div}}$
Ensure: Geocast region Ω
1: Initialize $\Omega = \emptyset$, $\text{div} = 0$;
2: Let $\text{div}(\cdot)$ denote the expected diversity of the region;
3: Let $\text{ratio}(\cdot)$ denote the ratio of the number of agents in a region and the size of the region;
4: Let $\text{OAR}(\cdot)$ denote the overall acceptance rate OAR_k of a region;
5: Let $\mathcal{Q}$ denote the set of subcell in the sensing area;
6: **repeat**
7: **if** $\mathcal{Q} = \emptyset$ **then**
8: **return** Ω
9: **else**
10: $c^* \leftarrow \text{argmin}_{c \in \mathcal{Q}} \text{ratio}(c)$;
11: $\mathcal{Q} \leftarrow \mathcal{Q} \setminus \{c^*\}$;
12: $\Omega \leftarrow \Omega \cup \{c^*\}$;
13: $\text{OAR} \leftarrow \text{OAR}(\Omega)$;
14: $\text{div} \leftarrow \text{div} + \text{div}(c^*)$;
15: **end if**
16: **until** $\text{OAR} \geq \overline{\text{OAR}}$ and $\text{div} \geq \overline{\text{div}}$
17: **return** Ω;

VI. AGENT SELECTION

After SSP sends task requests to the agents located in the cast region, agents will send their acceptance confirmation and their current real locations back to SSP if they accept the task. From the pool of agents who decide to accept the task, SSP need to select N agents to perform the sensing task. There are several methods on the sensor selection to achieve an optimal result: correlation based method, iterative partitioning based method and radius information based method [17]. Here, we choose the correlation based algorithm to select the agents as the correlations between selected agents directly affect the final sensing result.

The correlation based algorithm uses the correlations by (3) and select N agents from M respondents, among which the sum of correlations is the minimal. Specifically, the objective for this algorithm is as follow:

$$\begin{aligned} &\min \sum_{i=1}^{M} \sum_{j=1}^{M} a_i a_j R(d_{ij}) \\ &\text{subject to} \sum_{i=1}^{M} a_i = N \\ &\quad a_i a_j R(d_{ij}) \leq R(d_0) \\ &\quad a_i \in \{0,1\}, i = 1, 2, ..., M, \end{aligned} \tag{11}$$

where N is the targeted number of selected agents for sensing task, M is the total number of respondents, a_i denotes whether the i^{th} agent is selected, $R(\cdot)$ is the correlation function by (3), d_{ij} is the distance between the i^{th} agent and j^{th} agent, and d_0 is the decorrelation distance.

With the response from the agents who accept the task, SSP will select agents for sensing. If the number M of the respondents is less than N, then the sensing task fails. If the number M is exactly the same as N, then SSP need to check whether each pair of agents are uncorrelated and choose all the respondents if it is true. Otherwise, the SSP fails to complete the sensing task. Here, we need to discuss how to select the agents when M is larger than N. Firstly, SSP calculated the correlations among the M agents by (3) with their real locations. Then SSP put all the responded agents in the selection set, which means all the agents are selected. SSP removes the agents one by one from the selection set according to their correlations with the remaining agents until there are N agents in the selection set. SSP checks the correlation among the N agents to confirm that all pairs of agents are uncorrelated. Finally, SSP sends final confirmation messages to the selected agents for task sensing. Algorithm 2 gives the details for agent selection.

VII. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed framework and investigate the trade-offs between privacy and detection performance provided by our approach.

A. Experiment Setup

We use Gowalla, a real-world dataset to simulate the spatial distribution of the mobile agents. The Gowalla dataset contains $6,442,890$ check-ins with each check-in record including user id, check-in time, latitude, longitude and location id. In our experiments, we use the latest check-in of each user as his/her current location. Specifically, we choose the check-ins within the region $[32.58, 32.91; -97.42, -97.09]$, which contains 544 check-ins. The spatial distribution of the 544 points is shown in Fig. 2. We consider the size of the sensing area in our simulation as $500 \times 500m$ and use the spatial distribution of these 544 check-ins to simulate the distribution of agents for spectrum sensing in the urban area.

In our simulation, we use MATLAB to simulate the PSD construction, geocast region construction, task request allocation and agent selection. With the pre-

Algorithm 2 Agent Selection Algorithm

Require: Agent locations, respondent set $\mathcal{Q}$
Ensure: Selection set Ω

```
Initialize Ω = Q, a_i = 1, i = 1, 2, ....., M ;
repeat
    if Σ_{i=1}^{M} a_i < N then
        return ∅
    else
        agent*, i ← argmax_{agent_i ∈ Ω} Σ_{j=1}^{M} a_j R(d_ij);
        Ω ← Ω \ agent*;
        a_i = 0;
    end if
until Σ_{i=1}^{M} a_i = N
for i=1,...,M do
    for j=1,...,M do
        if a_i a_j R(d_ij) > R(d_0) then
            return ∅
        end if
    end for
end for
return Ω
```

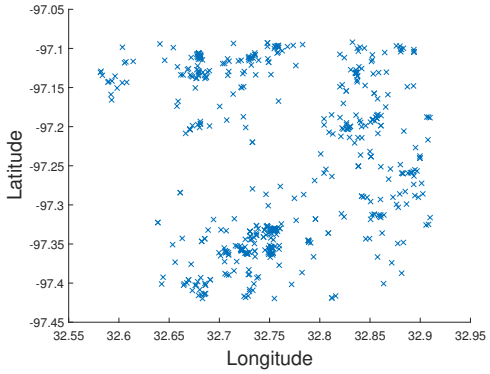
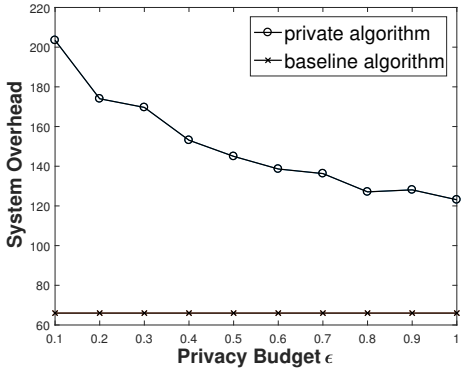
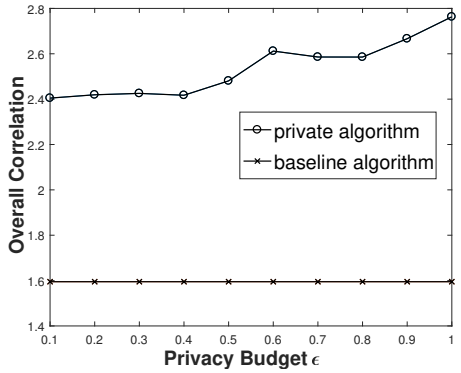


Fig. 2: The current spatial distribution of agents in the sensing area.

processed data, we construct the private spatial decomposition based on the geometric mechanism with $\epsilon \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ respectively. The choice of $\frac{\epsilon_1}{\epsilon}$ is less critical, and we set it as 0.5 in the construction. Then with the cell matrices from the PSD construction, we can construct geocast regions. In the geocast construction simulation, we set the decay factor α in the correlation function by (3) to be $0.1204/m$, which is the environmental parameter for urban area. We also assume that if the correlation calculated by (3) is less than 0.2 then the two agents are uncorrelated. So the decorrelation distance d_0 is equal to $\frac{\log 0.2}{\alpha} \approx 13m$. The individual acceptance rate is set as 0.5. As we need to guarantee that the sensing task will be successfully completed with a high probability, we set the threshold of the overall acceptance rate in (10) as 90%. According to simulation results given in [17], the lower bound for the number of uncorrelated agents is the expected number of uncorrelated agents, which is proportional to $\frac{R}{d_0}$. Here R is



(a)



(b)

Fig. 3: (a) Effect of varying ϵ on the system overhead with Overall Acceptance Rate $= 0.9$ and Individual Acceptance Rate $= 0.5$; (b) effect of varying ϵ on the overall correlation with Overall Acceptance Rate $= 0.9$ and Individual Acceptance Rate $= 0.5$.

the radius of the circle sensing region. Since the spatial index is rectangular in our simulation, we use an equivalent radius $R^* = \frac{\text{size}}{\sqrt{\pi}}$ instead to calculate the lower bound, which indicates the targeted diversity. The target diversity in our simulation is 27 following this calculation. Finally, we simulate the agent selection with the cast region, and output the final selection set and the experiment results for evaluation.

B. Evaluation Results

In this subsection, we evaluate the performance of our proposed approach. Specifically, we focus on system overhead and overall correlation, and give the analysis by comparing the simulation results.

1) Evaluation on system overhead and overall correlation for achieving privacy-preserving: In our proposed framework, the system overhead is related to the number of notified agents during the task allocation. In the simulation, we use the number of notified agents as the indicator of the system overhead, and evaluate how the privacy-preserving algorithm affects the system overhead. We set the overall acceptance rate (OAR) threshold as 0.9 and the individual acceptance rate (IAR) as 0.5. Fig. 3a compares the system overhead of our

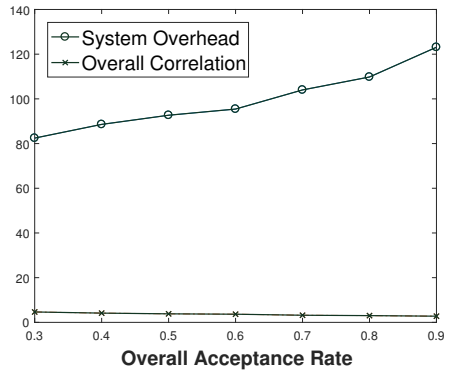


Fig. 4: Effect of varying overall acceptance rate threshold on the system overhead and overall correlation with $\epsilon = 1$ and Individual Acceptance Rate $= 0.5$.

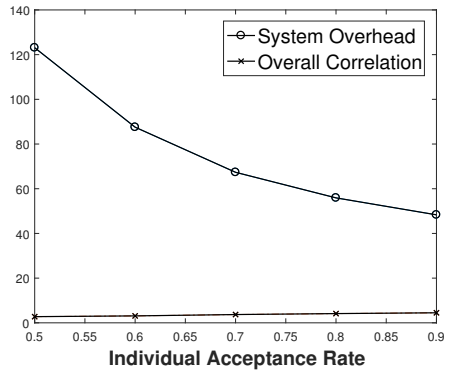


Fig. 5: Effect of varying individual acceptance rate on the system overhead and the overall correlation with $\epsilon = 1$ and Overall Acceptance Rate $= 0.9$.

private algorithm with the overhead of the baseline algorithm. We can observe that with higher privacy budget ϵ, the system overhead becomes smaller, and when we consider a non-private algorithm, the overhead is the lowest. Actually, when we make the algorithm more private, we need to add more noise to the spatial index, causing it more uncertain. Therefore, the actual number of notified agents will be larger than the wanted number. Besides, we can see that when ϵ is larger than 0.3, the system overhead for the privacy-preserving algorithm is close to the baseline algorithm. So we can choose an ϵ larger than 0.3 in our framework, then the system overhead incurred by the privacy preserving mechanism will be very small.

As the correlation directly affects the sensing performance, we need to analyze if the private mechanism will degrade the sensing performance. In our simulation, we still compare the overall correlation for the private algorithm with the correlation for the baseline algorithm by setting OAR $= 0.9$ and IAR $= 0.5$. Fig. 3 shows that the overall correlation will change significantly with the varying privacy budget, and the overall correlation for the private mechanism is close to the overall correlation for the baseline algorithm. We conclude that our private mechanism can provide much utility and also guarantee the privacy protection.

2) Trade-off between overall acceptance rate and overhead: By varying the overall acceptance threshold from 0.3 to 0.9, we evaluate how the overall acceptance rate threshold affects the system overhead and the overall correlation. We use the privacy budget $\epsilon = 1$ and IAR $= 0.5$. Fig. 4 shows that a high overall acceptance rate threshold causes a high system overhead as a large number of agents are considered to guarantee the high OAR. However, a higher overall acceptance rate will give a little better sensing result.

3) Evaluation on the effect of varying individual acceptance rate: We still consider the effect of varying IAR from 0.5 to 0.9 on the system overhead and the overall correlation by setting $\epsilon = 1$ and OAR $= 0.9$. From Fig. 5, we conclude that the a lower individual acceptance rate will result in a higher system overhead but a lower overall correlation. When the individual acceptance rate is lower, more agents will be considered in the cast region to guarantee the overall acceptance rate.

VIII. Related Work

Several solutions to location privacy in mobile services have been proposed in literature. A common approach is location obfuscation. Location obfuscation is used a lot for location-based services, where mobile users protect their true locations by slightly altering, substituting, or generalizing their location and querying service providers with the obfuscated locations [23]–[27]. However, adding noise directly on location information usually greatly degrade the utility of location data. For the spectrum sensing setting, this means that mobile agents may be distant from the reported location, and would need to either travel to the reported location to complete the task, causing a high travel cost, or report at their current location, decreasing the utility of their sensing reports due to the discrepancy of the true and reported location. Some papers consider the scenario of querying private locations in an outsourced database, where the data owner and the querying entity have a trusted relationship and the outsourced service provider is untrusted [28], [29]. This is different from our setting because the SSP and the mobile agents have none pre-established trust relationship. Another common approach is using the statistics of location data, such as the approach adopted in [30]. Statistics of location data could both protect individual location and provide relative accurate information, and thus is a promising approach. However, it is hard to directly apply statistics for spectrum sensing tasks, because the trustworthiness and diversity of mobile agents also need to be included in the picture. To and Ghinita [31], [32] propose to use private spatial decomposition for privacy issues in mobile crowdsourcing, but the challenges of worker trustworthiness and diversity are still left unsolved.

Privacy in collaborative spectrum sensing has been studied in a few prior works [2], [33]. In these prior works, the privacy challenge in collaborative spectrum sensing mainly arises from the aggregated sensing reports. On the other hand, in our crowdsourced spectrum sensing setting, location privacy becomes a challenge as early as when sensing requests are allocated to mobile sensing agents, because mobile sensing

agents and the SSP no longer share a trust relationship. In [34], privacy issues of selecting crowdsouced mobile agents is considered, but mobile agents are selected through a bidding process, which is different from the approach we considered in this paper. In [27], the authors propose DPSense, which protect location privacy of mobile agents through location perturbation. As a result, mobile agents need to travel to the desired locations in order to perform a spectrum sensing task. For our approach, we try to minimize the effort of mobile mobiles by exploiting the spatial diversity during task allocation stage.

IX. Conclusion

Crowdsourced spectrum sensing has great potential to improve the detection accuracy of primary users' presence, which is important for the implementation of dynamic spectrum access. In this paper, we have investigated the conflicting concerns of privacy and detection performance in crowdsourced spectrum sensing, and have proposed an approach that allows spectrum service provider to allocate tasks to a diverse pool of mobile agents without compromising their location privacy. Detailed evaluations on real-world database showed that the proposed approach achieves a good balance between privacy and utility by providing the differential privacy guarantee at low privacy budget and good detection performance of primary users' presence.

References

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE personal communications*, vol. 6, no. 4, pp. 13–18, 1999.

[2] Z. Gao, H. Zhu, S. Li, S. Du, and X. Li, "Security and privacy of collaborative spectrum sensing in cognitive radio networks," *IEEE Wireless Communications*, vol. 19, no. 6, 2012.

[3] S. J. Shellhammer, R. Tandra, J. Tomcik *et al.*, "Performance of power detector sensors of dtv signals in ieee 802.22 wrans," in *Proceedings of the first international workshop on Technology and policy for accessing spectrum*. ACM, 2006, p. 4.

[4] J. Howe, "The rise of crowdsourcing," *Wired magazine*, vol. 14, no. 6, pp. 1–4, 2006.

[5] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*. IEEE, 2005, pp. 131–136.

[6] O. Fatemieh, R. Chandra, and C. A. Gunter, "Secure collaborative sensing for crowd sourcing spectrum data in white space networks," in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*. IEEE, 2010, pp. 1–12.

[7] C. Song and Q. Zhang, "Achieving cooperative spectrum sensing in wireless cognitive radio networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 13, no. 2, pp. 14–25, 2009.

[8] B. Wang, K. R. Liu, and T. C. Clancy, "Evolutionary cooperative spectrum sensing game: how to collaborate?" *IEEE transactions on communications*, vol. 58, no. 3, 2010.

[9] I. Krontiris, F. C. Freiling, and T. Dimitriou, "Location privacy in urban sensing networks: research challenges and directions [security and privacy in emerging wireless networks]," *IEEE Wireless Communications*, vol. 17, no. 5, 2010.

[10] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu, "Differentially private spatial decompositions," in *Data engineering (ICDE), 2012 IEEE 28th international conference on*. IEEE, 2012, pp. 20–31.

[11] J. C. Navas and T. Imielinski, "Geocastâgeographic addressing and routing," in *Proceedings of the 3rd annual ACM/IEEE international conference on Mobile computing and networking*. ACM, 1997, pp. 66–76.

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

[13] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2006, pp. 486–503.

[14] F. McSherry and I. Mironov, "Differentially private recommender systems: building privacy into the net," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 627–636.

[15] W. Qardaji, W. Yang, and N. Li, "Differentially private grids for geospatial data," in *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*. IEEE, 2013, pp. 757–768.

[16] M. Schwartz, *Mobile wireless communications*. Cambridge University Press, 2004.

[17] Y. Selén, H. Tullberg, and J. Kronander, "Sensor selection for cooperative spectrum sensing," in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*. IEEE, 2008, pp. 1–11.

[18] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics letters*, vol. 27, no. 23, pp. 2145–2146, 1991.

[19] E. G. Larsson and M. Skoglund, "Cognitive radio in a frequency-planned environment: some basic limits," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4800–4806, 2008.

[20] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical communication*, vol. 4, no. 1, pp. 40–62, 2011.

[21] D. Duan, L. Yang, and J. C. Principe, "Cooperative diversity of spectrum sensing for cognitive radio systems," *IEEE transactions on signal processing*, vol. 58, no. 6, pp. 3218–3227, 2010.

[22] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," *SIAM Journal on Computing*, vol. 41, no. 6, pp. 1673–1693, 2012.

[23] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st international conference on Mobile systems, applications and services*. ACM, 2003, pp. 31–42.

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

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

[26] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 901–914.

[27] X. Jin, R. Zhang, Y. Chen, T. Li, and Y. Zhang, "Dpsense: Differentially private crowdsourced spectrum sensing," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 296–307.

[28] M. L. Yiu, G. Ghinita, C. S. Jensen, and P. Kalnis, "Enabling search services on outsourced private spatial data," *The VLDB Journal*, vol. 19, no. 3, pp. 363–384, 2010.

[29] B. Yao, F. Li, and X. Xiao, "Secure nearest neighbor revisited," in *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*. IEEE, 2013, pp. 733–744.

[30] J. W. Brown, O. Ohrimenko, and R. Tamassia, "Haze: Privacy-preserving real-time traffic statistics," in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2013, pp. 530–533.

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

[32] H. To, G. Ghinita, L. Fan, and C. Shahabi, "Differentially private location protection for worker datasets in spatial crowdsourcing," *IEEE Transactions on Mobile Computing*, vol. 16, no. 4, pp. 934–949, 2017.

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

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