Geo-Indistinguishablility for Crowdsourced-Based Radio Environment Map Construction

Shahira Amin*, Liang Li†, Yuanxiong Guo‡, Miao Pan§, and Yanmin Gong*
* Department of Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, TX 78249
† School of Cyber Engineering, Xidian University, Xi'an 710071
‡ Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249
§ Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204
tsz035@my.utsa.edu, liliang_1127@outlook.com, yuanxiong.guo@utsa.edu, mpan2@uh.edu, yanmin.gong@utsa.edu

GLOBECOM 2020 - 2020 IEEE Global Communications Conference | 978-1-7281-8298-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/GLOBECOM42002.2020.9348142

***Abstract*—The aim of this paper is to preserve location privacy of crowdsourced-based spectrum sensing agents using geo-indistinguishability. We considered database-driven dynamic spectrum access, where a radio environment map provides spectrum availability information for dynamic spectrum access management. Moreover, we assumed crowdsourced-based spectrum sensing, where a pool of allocated mobile users, called crowdsourced-based spectrum sensing agents, sense the spectrum and report their actual location and the received signal strength to the spectrum manager that constructs a radio environment map. This discloses location information of crowdsourced-based spectrum sensing agents and violates their location privacy. Consequently, crowdsourced-based spectrum sensing agents could be discouraged to participate in spectrum sensing. In our paper, to solve the problem of location disclosure, we adopted planar Laplacian mechanism, where each crowdsourced-based spectrum sensing agent reports an obfuscated location instead of its actual location, which achieves geo-indistinguishability. Our simulation results were based on real-world CRAWDAD dataset. Our results showed that with a moderate privacy level, location privacy of crowdsourced-based spectrum sensing agents was preserved while the effect of introduced location noise on the accuracy of radio environment map was insignificant.**

***Index Terms*—dynamic spectrum access, received signal strength, spectrum availability, location privacy, prediction accuracy**

I. Introduction

Dynamic spectrum access (DSA) is a promising solution for the wireless spectrum scarcity problem. In DSA, a secondary user is allowed to access the spectrum when the primary user is absent, which leads to better spectrum utilization. In a database-driven DSA system, the spectrum manager (SM) manages the spectrum using a radio environment map (REM), proposed in [1], [2], which is a database that contains information about spectrum availability over an area of interest. Constructing an accurate REM is essential for providing reliable spectrum availability information for better spectrum utilization.

There are several techniques for constructing REM in the literature. Some techniques are based on the radio propagation model [3], which ignore local environment factors and lead to REM inaccuracy. Other techniques are based on allocated wireless sensors in the area of interest that sense the dynamic signal strength, which is used to estimate spectrum availability. However, only a limited number of sensors could be allocated due to their high cost [4], which leads to poor coverage. To address this problem, the SM could allocate a pool of mobile users (MUs) [5]–[7] to sense the received signal strength (RSS) using their mobile devices, referred to in the literature as crowdsourced-based [8] spectrum sensing (CSS) agents. Using the reported RSS and their corresponding locations, the SM constructs REM based on kriging, a popular interpolation method in geo-statisitics.

Since CSS agents submit their sensitive location information, this may lead to privacy violations. Information such as an individual's home or work location, sexual preferences, political views, and religious tendencies could be easily inferred from the individual's actual location when collected and analyzed on regular basis [9]. For this reason, CSS agents could be unwilling to participate in spectrum sensing tasks. Thus, private location information should be protected against attackers who maliciously use location information for robbing, stalking, blackmailing, and other privacy threats [9].

In literature, several techniques have been proposed to address location privacy preservation. Some researchers adopted anonymization to achieve k-anonymity [10], [11], which guarantees that an adversary cannot distinguish an individual whose information is in the released data among at least $k-1$ other individuals whose information is also in the released data. However, if the adversary has sufficient prior knowledge, this method can no longer guarantee privacy. Others focused on encryption-based methods, which could hide location information by encryption, but requires huge computational cost [12] leading to undesired communication delays. Another approach proposed in literature is perturbation-based methods, which guarantee differential privacy or geo-indistinguishability [13], a generalized definition of differential privacy used in geo-statistics.

In our work, we were motivated to use geo-indistinguishability to preserve location privacy of CSS agents due to the limitations observed in anonymization and encryption-based methods:

- We used geo-indistinguishability which guarantees location privacy *regardless of the prior knowledge available to the adversary*. To achieve geo-indistinguishability, each CSS agent adopted a planar Laplacian mecha-

978-1-7281-8298-8/20/$31.00 ©2020 IEEE

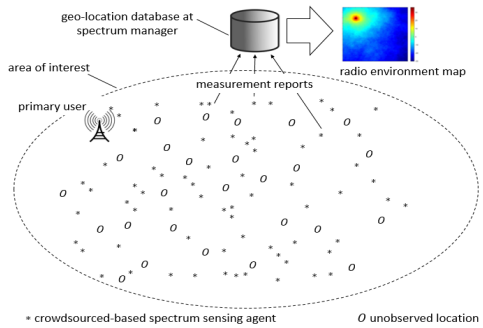


Fig. 1. System model

nism [13], which is *computationally inexpensive*, to obfuscate/randomize its location information locally.
- We evaluated how privacy impacts the radio environment map construction performance through extensive simulations using a real-world dataset.

The rest of our paper is organized as follows. In Section II, we define our problem. In Section III, we explain our proposed privacy preserving OK-based REM construction method. In Section IV, we present and discuss our simulation results. In Section V, we give our recommendations for future research. In Section VI, we conclude our work.

II. Problem Statement

In this section, we explain our system model, threat model, and location privacy preservation of CSS agents.

A. System Model

In our framework shown in Fig. 1, we considered two parties: the SM and MUs. The SM constructs REM over an area of interest denoted by $\mathcal{D}$ and utilizes the resulting REM to manage DSA. The MUs work as sensing agents referred to in the rest of our paper as CSS agents. Each CSS agent senses the RSS at its location using its mobile device and sends a measurement report to the SM. Denoting the actual location of CSS agent i by $\mathbf{x}_i$, and the RSS at $\mathbf{x}_i$ by $Z(\mathbf{x}_i)$, the CSS agent at $\mathbf{x}_i$ reports $R(\mathbf{x}_i) = \big(\hat{\mathbf{x}}_i, Z(\mathbf{x}_i)\big)$, where $\hat{\mathbf{x}}_i$ is an obfuscated location of CSS agent i. We considered N CSS agents located at $\{\mathbf{x}_i : i = 1, 2, \ldots, N\}$. The N CSS agents measure the RSS at their locations: $\{Z(\mathbf{x}_i), i = 1, 2, \ldots, N\}$ and send their measurement reports $\{R(\mathbf{x}_i) = (\hat{\mathbf{x}}_i, Z(\mathbf{x}_i)), i = 1, 2, \ldots, N\}$ to SM. Based on the received measurement reports from CSS agents, the SM predicts the RSS $\{Z(\mathbf{y}_j) : j = 1, 2, \ldots, M\}$ at unobserved locations: $\{\mathbf{y}_j : j = 1, 2, \ldots, M\}$, and then constructs REM over the area of interest $\mathcal{D}$ based on its measurement predictions. In our paper, we focused on one primary user. We assumed that the SM knows the location $\mathbf{x}_0$, licensed band, and transmission schedule of the primary user.

B. Threat Model

We assumed that CSS agents are trusted to perform spectrum sensing. CSS agents would not intentionally report false measurements to poison REM. On the other hand, we aimed to protect CSS agents' location information against inference attacks from an adversary with arbitrary prior knowledge. The adversary could be the honest-but-curious SM that follows the protocol trustfully, but tends to infer CSS agents' location information. The adversary could also be an eavesdropper that eavesdrop messages exchanged between CSS agents and SM.

C. Location Privacy Preservation of Crowdsourced-Based Spectrum Sensing Agents

In our work, we adopted geo-indistinguishability, proposed by Andre et al. in 2013 [13], to preserve location privacy of CSS agents. Geo-indistinguishability provides strong privacy guarantees against attackers with arbitrary prior knowledge. The definition of ϵ-geo-indistinguishability is as follows [13]:

***Definition** 1 (ϵ-Geo-Indistinguishability):* A mechanism $\mathcal{M}$ satisfies ϵ-geo-indistinguishability, iff for all output $\mathcal{O}$ and all locations $\mathbf{x}$, $\mathbf{x}^{'}$ where $\|\mathbf{x} - \mathbf{x}^{'}\| \leq r$, $r > 0$:

$$\ln \frac{Pr(\mathcal{M}(\mathbf{x}) = \mathcal{O})}{Pr(\mathcal{M}(\mathbf{x}^{'}) = \mathcal{O})} \leq \epsilon,$$

where ϵ indicates the level of privacy at one unit of distance. In this paper, we used planar Laplacian mechanism to achieve ϵ-geo-indistinguishability as shown in Algorithm 1 [13]. Using planar Laplacian mechanism, each CSS agent computes its obfuscated location $\hat{\mathbf{x}}_i$, and then reports $\hat{\mathbf{x}}_i$ to SM instead of reporting its actual location.

Algorithm 1 Planar Laplacian Mechanism for Achieving ϵ-Geo-Indistinguishability

Input: real location $\mathbf{x}_i$, ϵ
1: Sample θ uniformly in $[0, 2\pi)$;
2: Sample p uniformly in $[0, 1)$;
3: $r \leftarrow -\frac{1}{\epsilon}(W_{-1}(\frac{p-1}{e}) + 1)$, where W_{-1} is the Lambert W function;
4: $\mathbf{u}_i \leftarrow \mathbf{x}_i + [r * \cos\theta, r * \sin\theta]$;
5: $\hat{\mathbf{x}}_i \leftarrow \mathbf{u}_i$;
Output: obfuscated location $\hat{\mathbf{x}}_i$

III. Privacy Preserving Ordinary Kriging-Based Radio Environment Map Construction

In this section, we explain the steps to constructing REM. When SM receives the measurement reports $\{R(\mathbf{x}_i)\}$ from CSS agents, SM constructs REM using OK. There are three main steps to OK-based REM construction: measurement detrending, semivariogram parameter estimation, and measurement prediction.

A. Measurement Detrending

Generally, kriging assumes the following model [14]: $Z(\mathbf{x}_i) = \mu(\mathbf{x}_i) + \delta(\mathbf{x}_i)$, where $\mu(\cdot)$ is the mean capturing the path loss, and $\delta(\cdot)$ is the residue capturing the shadowing effect. Ordinary kriging further assumes that $Z(\mathbf{x}_i)$ is intrinsically stationary, i.e. [14]: $\mathbb{E}\big[Z(\mathbf{x}_i)\big] = \mu(\mathbf{x}_i) = \mu$, $\mathbb{E}\big[\big(Z(\mathbf{x}_i) - Z(\mathbf{x}_j)\big)^2\big] = 2\gamma(h_{i,j})$, where μ is an unknown constant; $h_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|$ is the distance between two

locations, and $\gamma(\cdot)$ is the semivariogram function that models the correlation between two locations.

In radio environment, the mean capturing path loss $\mu(\mathbf{x}_i)$ is not constant, but is a function of location $\mathbf{x}_i$. Therefore, we first performed detrending in order to make $Z(\mathbf{x}_i)$ an intrinsically stationary process with a constant mean. At SM, using the knowledge of the obfuscated location $\hat{\mathbf{x}}_i$ and RSS $Z(\mathbf{x}_i)$, we estimated $\mu(\hat{\mathbf{x}}_i)$, and then subtracted $\mu(\hat{\mathbf{x}}_i)$ from $Z(\mathbf{x}_i)$. The path loss at location $\hat{\mathbf{x}}_i$ is estimated as [15]:

$$P(\hat{\mathbf{x}}_i) = \alpha 10 \log_{10}(d(\hat{\mathbf{x}}_i, \mathbf{x}_0)) + P_0, \tag{1}$$

where d_i is the distance between $\hat{\mathbf{x}}_i$ and the location of the primary user $\mathbf{x}_0$, and α and P_0 are path loss parameters calculated empirically. The detrended measurement at $\mathbf{x}_i$ is:

$$S(\mathbf{x}_i) = Z(\mathbf{x}_i) - P(\hat{\mathbf{x}}_i), \tag{2}$$

From (1), it is clear that we estimated the path loss at obfuscation location $\hat{\mathbf{x}}_i$ instead of actual location $\mathbf{x}_i$.

B. Semivariogram Parameter Estimation

The exponential semivariogram is commonly used to model radio systems involving shadowing [16] and could be expressed as [4]:

$$\gamma(h) = \beta_1(1 - \exp(-\frac{h}{\beta_2})), \tag{3}$$

where β_1 represents the variance of the RSS measurements and is called $sill\ variance$, and β_2 is known as the $range$ indicating the exponential decay in the semivariogram. In our paper, we assumed that both β_1 and β_2 are unknown. To obtain β_1 and β_2, we trained the exponential semivariogram model in (3) using the detrended measurements in (2). First, we calculated a set of empirical semivariogram values $\{\hat{\gamma}(\hat{h}_{lag})\}$ using the detrended measurements as follows [4]:

$$\hat{\gamma}(\hat{h}_{lag}) = \frac{1}{2|H(\hat{h}_{lag})|} \sum_{(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j) \in H(\hat{h}_{lag})} (S(\mathbf{x}_i) - S(\mathbf{x}_j))^2, \tag{4}$$

where $\hat{h}_{lag}$ is a set of pre-calculated separation distances called lag distances, and $H(\hat{h}_{lag}) = \{(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j) | \hat{h}_{lag} - \sigma \leq \|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\| \leq \hat{h}^{lag} + \sigma\}$, where σ is a small tolerance.

Second, using the empiricial semivariogram values $\{(\hat{h}_i^{lag}, \hat{\gamma}(\hat{h}_i^{lag}))\}$ in (4), we trained the exponential semivariogram model in (3) by *minimizing the mean squared error* between the model and empirical semivariogram values [4]. From (4), it is clear that, instead of using actual locations $\{\mathbf{x}_i\}$, we used obfuscated locations $\{\hat{\mathbf{x}}_i\}$ to calculate both the detrended measurements and lag distances $\{\hat{h}_i^{lag}\}$, which are used in calculating the empirical semivariogram values $\hat{\gamma}(\hat{h}_i^{lag})$.

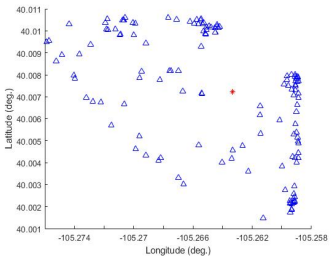


Fig. 2. Measurements and base station (primary user) locations in group D (CRAWDAD dataset)

C. Measurement Prediction

Based on the estimated semivariogram model $\gamma(h; \beta_1, \beta_2)$, we predicted the RSS at an unobserved location $\mathbf{y}_j$ as [4]:

$$\begin{aligned}\hat{Z}(\mathbf{y}_j) =& P(\mathbf{y}_j) + \hat{S}(\mathbf{y}_j) \\ =& \alpha 10 \log_{10}(d(\mathbf{y}_j, \mathbf{x}_0)) + P_0 + \sum_{i=1}^{N} w_i \cdot S(\mathbf{x}_i),\end{aligned} \tag{5}$$

where $\sum_{i=1}^{N} w_i = 1$ are normalized weights. The optimal set of weights $\{w_i\}_{i=1}^{N}$ minimizes the prediction variance $Var\big[\varepsilon(\mathbf{y}_j)\big] = Var\big[\hat{S}(\mathbf{y}_j) - S(\mathbf{y}_j)\big]$, which could be written as follows [4]:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \rho \end{bmatrix} = \begin{bmatrix} \gamma(\hat{h}_{1,1}) & \dots & \gamma(\hat{h}_{1,N}) & 1 \\ \gamma(\hat{h}_{2,1}) & \dots & \gamma(\hat{h}_{2,N}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\hat{h}_{N,1}) & \dots & \gamma(\hat{h}_{N,N}) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(\hat{h}_{1,j}) \\ \gamma(\hat{h}_{2,j}) \\ \vdots \\ \gamma(\hat{h}_{N,j}) \\ 1 \end{bmatrix}, \tag{6}$$

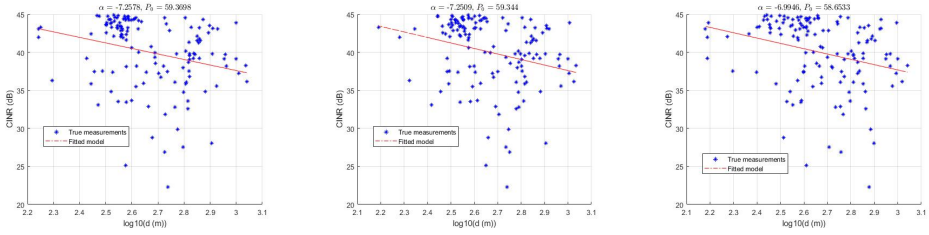
where ρ is the Lagrange multiplier. From (6), it could be seen that the set of weights $\{w_i\}$ depends on the trained semivariogram model and distance matrix $\{\hat{h}_{i,j}\}$ which were calculated using obfuscated locations.

IV. Simulation Results

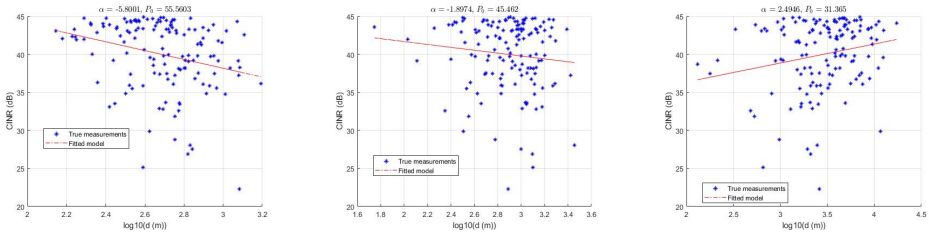
In this section, we present and discuss our simulations. Our results showed that by applying ϵ-geo-indistinguishability, the privacy of CSS agents was protected, and at the same time, a fairly accurate REM was constructed for a specific range of privacy budget ϵ.

A. Simulation Setup

1) Dataset:: In our simulations, we used CRAWDAD dataset [17], which was collected at the University of Colorado Boulder and contains measurements of the WiMAX network serving the University of Colorado campus. The WiMAX network consists of five base stations operating on different channels at 2.5 GHz, with an educational license granted to the University. The measurements include carrier to interference plus noise ratio (CINR), relative constellation error (RCE), error vector magnitude (EVM), and subcarrier spectrum flatness. The measurements were taken

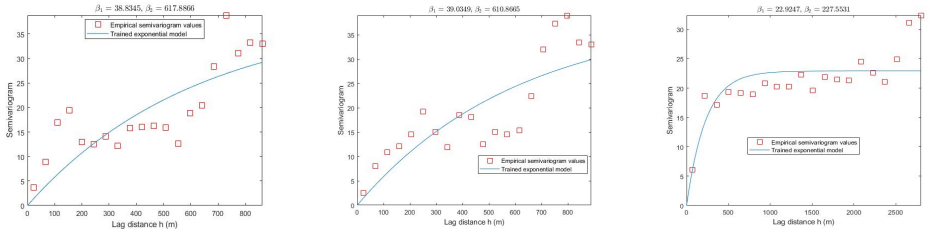


(a) no location obfuscation (b) $\epsilon^* = 1.0$ (c) $\epsilon^* = 0.5$

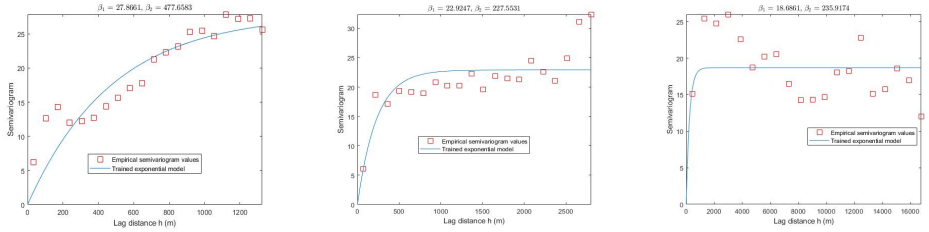


(d) $\epsilon^* = 0.2$ (e) $\epsilon^* = 0.05$ (f) $\epsilon^* = 0.01$

Fig. 3. Measurement detrending



(a) no location obfuscation (b) $\epsilon^* = 1.0$ (c) $\epsilon^* = 0.5$



(d) $\epsilon^* = 0.2$ (e) $\epsilon^* = 0.05$ (f) $\epsilon^* = 0.01$

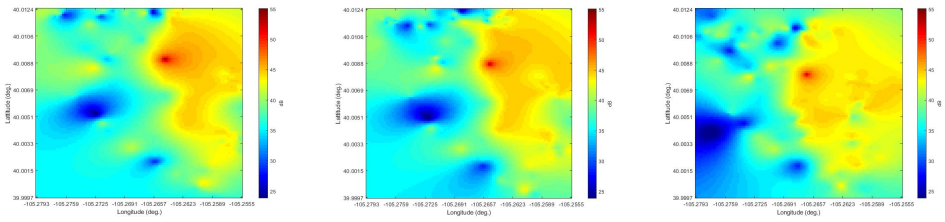
Fig. 4. Semivariogram parameter estimation

using a portable spectrum analyzer over four distinct campaigns/groups (A,B,C, and D). In our simulations, we chose the CINR measurements in group D, focusing on channel 308 being studied in our experiments. Group D consists of 138 measurements taken at optimized locations. The base station (i.e. primary user) operating on channel 308 (named "Eng_North_GENI") is located at longitude $-105.26333°$ and latitude $40.00722°$ and has a BSID 3674210305. Fig. 2 shows the locations of the measurements (blue triangles) and base station (red asterisk).

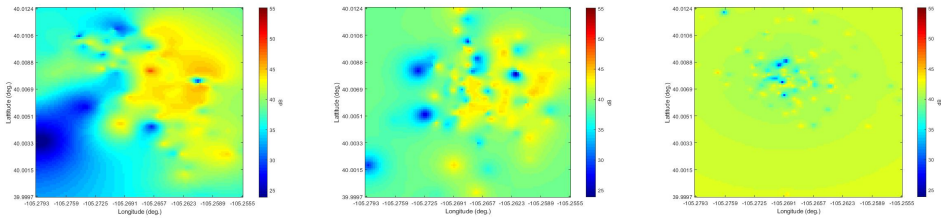
2) Privacy specifications:: To set the parameter ϵ, we assumed that all CSS agents choose a privacy level ϵ^* and a privacy radius r^*, where $\epsilon = \frac{\epsilon^*}{r^*}$. We performed our simulation based on $\epsilon^* = 0.01, 0.05, 0.2, 0.5, 1$ and $r^* = 20$ meters, which are commonly used values for location privacy problems in the literature [18]–[20].

B. Performance Evaluation

In Fig. 3, Fig. 4, and Fig. 5, we observed the effect of location obfuscation on the three steps to construct REM. Fig. 3 illustrates measurement detrending, where the path loss model in (1) (red line) is fitted to the CINR measurements (blue asterisk) to obtain the model parameters α and P_0. In Fig. 3(a), the fitted model parameters are $\alpha = -7.2578,\ P_0 = 59.3698$ when CSS agents do not obfuscate their locations. Fig. 3(b), Fig. 3(c), Fig. 3(d), Fig. 3(e), and Fig. 3(f) show the fitted model when CSS agents select their privacy levels $\epsilon^* = 1, 0.5, 0.2, 0.05, 0.01$, respectively. From Fig. 3(b), Fig. 3(c), Fig. 3(d), Fig. 3(e), and Fig. 3(f) it is clear that as privacy



(a) no location obfuscation (b) $\epsilon^* = 1.0$ (c) $\epsilon^* = 0.5$



(d) $\epsilon^* = 0.2$ (e) $\epsilon^* = 0.05$ (f) $\epsilon^* = 0.01$

Fig. 5. Measurement prediction (radio environment map)

level ϵ^* increases, the fitted model parameters α and P_0 are insignificantly affected by location obfuscation. In Fig. 3(b), the fitted model parameters ($\alpha = -7.2509,\ P_0 = 59.344$) are the closest to Fig. 3(a) followed by Fig. 3(c) ($\alpha = -6.9946,\ P_0 = 58.6533$), Fig. 3(d) ($\alpha = -5.8001,\ P_0 = 55.5603$), Fig. 3(e) ($\alpha = -1.8974,\ P_0 = 45.462$), then Fig. 3(f) ($\alpha = 2.4946,\ P_0 = 31.365$).

To detrend the CINR measurements, first, we converted the raw CINR measurements, which are given in linear scale, to dB. Second, we calculated the log of the distance between the base station and each measurement location. Third, using the calculated distance, we fitted the parameters α and P_0 in (1) to the CINR measurements as shown in Fig. 3. Since the difference between CINR and path loss is a constant term, which depends on the transmit power, antenna gain, and noise floor, converting CINR to path loss before fitting is not necessary. The path loss model parameters α and P_0 are automatically adjusted to account for the constant difference between CINR and path loss. Finally, to obtain the deterended measurements, we subtracted the path loss calculated using (1) from the CINR measurements.

Fig. 4 shows semivariogram parameter estimation, where the exponential semivariogram model in (3) (blue line) is trained using the empirical semivariogram values in (4) (red squares) to estimate the model parameters β_1 and β_2. From Fig. 4(a), when CSS agents do not obfuscate their locations, the estimated model parameters are $\beta_1 = 38.8345,\ \beta_2 = 617.8866$. As shown in Fig. 4(b), when CSS agents select a privacy level $\epsilon^* = 1$, the estimated model parameters ($\beta_1 = 39.0349,\ \beta_2 = 610.8665$) are the closest to Fig. 4(a). This is followed by Fig. 4(c) ($\beta_1 = 32.4344,\ \beta_2 = 564.9401$), when CSS agents select their privacy level $\epsilon^* = 0.5$. Then, Fig. 4(d) ($\beta_1 = 27.8661,\ \beta_2 = 477.6583$), when CSS agents select their privacy level $\epsilon^* = 0.2$. Followed by Fig. 4(e) ($\beta_1 = 22.9247,\ \beta_2 = 227.5531$), when CSS agents select their privacy level $\epsilon^* = 0.05$. Then, Fig. 4(f) ($\beta_1 = 18.6861,\ \beta_2 = 235.9174$), when CSS agents select their privacy level $\epsilon^* = 0.01$. This means that a higher privacy level ϵ^* leads to more accurate exponential semivariogram parameter estimation.

Fig. 5 shows measurement prediction. Fig. 5(a) shows the REM constructed when CSS agents do not obfuscate their locations. We consider the REM in Fig. 5(a) as the ideal REM when compared to the privacy preserving REMs constructed when CSS agents choose different privacy levels. It could be seen in Fig. 5(a), Fig. 5(b), Fig. 5(c), and Fig. 5(d) that the base station (red spot) could be located, unlike in Fig. 5(e) and Fig. 5(f), where the base station could not be located. Fig. 5(b), Fig. 5(c), Fig. 5(d), Fig. 5(e), and Fig. 5(f) show the privacy preserving REMs when CSS agents select their privacy levels $\epsilon^* = 1, 0.5, 0.2, 0.05, 0.01$, respectively. The REM in Fig. 5(b) is more comparable to the ideal REM in Fig. 5(a) followed by Fig. 5(c), Fig. 5(d), Fig. 5(e), and Fig. 5(f). This implies that a larger privacy level $\epsilon^* >= 0.2$ slightly affects the accuracy of REM.

To measure the accuracy of REM, we calculated the MAE of measurement prediction. Using Monte Carlo simulation, we calculated the average of the MAE over 100 independent realizations. In each realization, we used k-fold cross validation. We set $k = \left\lfloor \frac{138}{10} \right\rfloor = 13$, where the 138 measurements are split into 13 smaller sets (folds) of measurements. For each fold of the 13 folds:

1) The exponential semivariogram model (3) is trained using $k - 1 = 12$ folds.
2) The resulting semivariogram model is tested using the remaining fold by calculating the MAE as follows

$$MAE = \frac{\sum_{i=1}^{D_t} |\hat{Z}(x_i) - Z(x_i)|}{D_t}, \tag{7}$$

where $\hat{Z}(x_i)$ is the estimated CINR of the CSS agent

TABLE I
MEAN ABSOLUTE ERROR

	Mean Absolute Error (MAE) dB
no location obfuscation	2.41
$\epsilon^* = 1$	2.53
$\epsilon^* = 0.5$	2.66
$\epsilon^* = 0.2$	2.93
$\epsilon^* = 0.05$	3.66
$\epsilon^* = 0.01$	4.19

located at x_i using the trained exponential semivariogram model; $Z(x_i)$ is the actual CINR of the CSS agent located at x_i, and D_t is the number of test measurements, set to $D_t = 10$ measurements in our experiment.

The MAE reported by the 13 folds is finally averaged. Table I compares the MAE of REMs constructed with different CSS agents' privacy levels. The ideal case, where the CSS agents do not obfuscate their locations, resulted in the lowest MAE. The MAE in the ideal case is equal to 2.41 dB, where errors occur due to ordinary kriging-based interpolation. As shown in Figures 5(b), 5(c), 5(d), 5(e), 5(f), the smaller the privacy level ϵ^*, the less accurate the REM, and hence, the lower the MAE. The MAE is equal to $2.53, 2.66, 2.93, 3.66, 4.19$ dB with $\epsilon^* = 1.0, 0.5, 0.2, 0.05, 0.01$, respectively.

V. OUR RECOMMENDATIONS FOR FUTURE RESEARCH

Our simulations showed that at a low privacy level $\epsilon = 0.01, 0.05$, a large amount of location noise was introduced, and hence, a highly inaccurate REM was obtained. To address this problem, we recommend in future work that the negative impact of noisy locations on REM be minimized at either SM side or CSS agents' side. In the former approach, assuming the SM knows the probability distribution of location noise, SM could apply uncertainty-aware interpolation to adjust measurement detrending, empirical semivariogram calculation, and distance matrix calculation. In the latter approach, CSS agents could use location obfuscation mechanisms other than planar Laplacian mechanism, which could be designed to minimize the impact of noisy locations on REM accuracy, and at the same time, achieve ϵ-geo-indistinguishability.

VI. CONCLUSION

We tackled the problem of location disclosure of crowdsourced-based spectrum sensing agents. Each crowdsourced-based spectrum sensing agent adopted planar Laplacian mechanism to obfuscate its location, which guarantees geo-indistinguishability. We evaluated the impact of location privacy on radio environment map construction performance using extensive simulations, which were based on CRAWDAD real-world dataset. Our results showed that with a moderate privacy level, both location privacy of crowdsourced-based spectrum sensing agents and a reasonably accurate radio environment map were achieved.

ACKNOWLEDGEMENT

The work of S. Amin and Y. Gong was supported in part by the U.S. National Science Foundation under grants US CNS-2029685 and CNS-1850523. The work of Y. Guo was supported in part by the U.S. National Science Foundation under grant US CNS-2029685. The work of M. Pan was supported in part by the U.S. National Science Foundation under grants US CNS-1646607, CNS-1801925, and CNS-2029569.

REFERENCES

[1] Y. Zhao, J. Gaeddert, K. K. Bae, and J. H. Reed, "Radio environment map enabled situation-aware cognitive radio learning algorithms," in *Software Defined Radio Forum (SDRF) technical conference*, 2006.

[2] Y. Zhao, L. Morales, J. Gaeddert, K. K. Bae, J.-S. Um, and J. H. Reed, "Applying radio environment maps to cognitive wireless regional area networks," in *2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*. IEEE, 2007, pp. 115–118.

[3] D. Gurney, G. Buchwald, L. Ecklund, S. L. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the tv white space," in *2008 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*. IEEE, 2008, pp. 1–9.

[4] Y. Hu and R. Zhang, "Secure crowdsourced radio environment map construction," in *2017 IEEE 25th International Conference on Network Protocols (ICNP)*. IEEE, 2017, pp. 1–10.

[5] Z. Huang and Y. Gong, "Differential location privacy for crowdsourced spectrum sensing," in *2017 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2017, pp. 1–9.

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

[7] 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*, 2016, pp. 296–307.

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

[9] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "Location privacy via geo-indistinguishability," *ACM Siglog News*, vol. 2, no. 3, pp. 46–69, 2015.

[10] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, 2007.

[11] R. Shokri, C. Troncoso, C. Diaz, J. Freudiger, and J.-P. Hubaux, "Unraveling an old cloak: k-anonymity for location privacy," in *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*, 2010, pp. 115–118.

[12] X. He, M. M. Islam, and H. Dai, "Privacy-preserving crowdsensing based radio environment mapping."

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

[14] N. Cressie, *Statistics for spatial data*. John Wiley & Sons, 2015.

[15] R. Augusto and C. Panazio, "On geostatistical methods for radio environment maps generation under location uncertainty," *Journal of Communication and Information Systems*, vol. 33, no. 1, 2018.

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

[17] M. Ton and C. Phillips, "CRAWDAD dataset cu/wimax (v. 2012-06-01)," Downloaded from https://crawdad.org/cu/wimax/20120601, Jun. 2012.

[18] S. Oya, C. Troncoso, and F. Pérez-González, "Is geo-indistinguishability what you are looking for?" in *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*, 2017, pp. 137–140.

[19] K. Chatzikokolakis, E. Elsalamouny, and C. Palamidessi, "Efficient utility improvement for location privacy," *Proceedings on Privacy Enhancing Technologies*, vol. 2017, no. 4, pp. 308–328, 2017.

[20] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "A predictive differentially-private mechanism for mobility traces," in *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 2014, pp. 21–41.