Training-Free Indoor White Space Exploration

Dongxin Liu, Fan Wu, *Member, IEEE*, Linghe Kong, *Member, IEEE*,
Shaojie Tang, *Member, IEEE*, Yuan Luo, *Member, IEEE*, and Guihai Chen, *Senior Member, IEEE*

***Abstract*— Exploration of white spaces has been recognized as a promising way to improve the utilization of wireless spectrums. Especially, indoor white space exploration has been shown to be much more challenging than that in outdoor environment. Most of existing works on indoor white space exploration infer availabilities of TV channels based on the correlations among the channels and locations, which is learned from the training data measured beforehand. However, the process of training data collection normally requires considerable time and devices, as well as human power. In this paper, we perform a measurement of indoor white spaces and study their characteristics. Based on the in-depth understanding of the characteristics, we propose a Training-free Indoor white space exploration MEchanism (TIME). In TIME, we design an algorithm for channel state inference based on Bayesian compressive sensing, as well as an incremental method for deployment of spectrum detectors. Furthermore, we present an algorithm to determine a proper number of spectrum detectors in need. Extensive real-world experiments are conducted to evaluate TIME's performance. The evaluation results demonstrate that TIME achieves competitive performances against the state-of-the-art training-based mechanisms.**

***Index Terms*— White space, spectrum measurement, spectrum exploration.**

I. Introduction

THE fast development of wireless networks and mobile communications leads to explosive wireless traffic growth, and thus the shortage of wireless spectrums becomes more and more serious given the existing static spectrum allocation strategy. However, it is observed that most licensed spectrums are not fully utilized, but in contrast, the amount of unlicensed spectrums, which are free to use, is very limited. To deal with the growing spectrum demand, the concept of dynamic spectrum access (DSA) was proposed, which allows unlicensed wireless devices to access the locally vacant licensed spectrums [1]–[5].

In 2008, Federal Communication Commission (FCC) issued a historic rule that allows the unlicensed access of vacant (or unoccupied) TV spectrums. After that, accessing vacant TV spectrums became a popular implementation of DSA. People call vacant TV spectrums "TV white spaces" or simply "white spaces". Unlicensed devices are allowed to access the locally unoccupied TV spectrums, but should not interfere with any of the licensed transmissions (e.g., TV broadcast). Therefore, all user devices (especially the unlicensed ones) are required to explore whether a TV channel is occupied or not before using it.

There are mainly two kinds of white space exploration: outdoor and indoor white space exploration. Most of the prior works focus on the outdoor white space exploration [6]–[8], where two approaches are spectrum sensing and geo-location database. The spectrum sensing approach requires every unlicensed device directly detect the state of a TV channel before using it, which is expensive on devices and energy consumption. And it is difficult for off-the-shelf devices to detect the existence of TV signals at high accuracy. In contrast, the geo-location database approach, maintained by FCC, needs no hardware and is easy to implement. Most white space devices and standards are designed based on it. A user device gets to know the white space availability by querying an online database, which stores a white space availability map. The map indicates the availability of wireless spectrums at different outdoor locations. However, in the indoor scenario, we cannot directly apply the approach of geo-location database. This is because there are relatively more obstacles (e.g., walls) than that of the outdoor scenario. Directly applying geo-location database to the indoor scenario would miss considerable amount of white spaces [9]. To solve this problem, Ying *et al.* [9] carried out a large-scale white space measurement, and proposed the first indoor white space exploration system, namely WISER. Then, Liu *et al.* [10] studied the linear dependence among different channels and locations, and proposed a cost-efficient indoor white space exploration scheme FIWEX.

Existing indoor white space exploration approaches [9], [10] try to deploy spectrum detectors at a small number of indoor locations, and infer a complete availability map of indoor white space through methods of data reconstruction. Their data reconstruction modules and spectrum detector deployment methods rely on the knowledge of training data sets. This means that results of indoor white space measurement should be collected as training data before running the white space

Manuscript received May 2, 2016; revised August 5, 2016; accepted August 28, 2016. Date of publication September 5, 2016; date of current version October 13, 2016. This work was supported in part by the State Key Development Program for Basic Research of China (973 project) under Grant 2014CB340303, in part by China NSF under Grant 61672348, Grant 61672349, Grant 61422208, Grant 61472252, Grant 61473109, Grant 61303202, Grant 61272443, and Grant 61133006, in part by the Shanghai Science and Technology Fund under Grant 15220721300, in part by the CCF-Tencent Open Fund, and in part by the Scientific Research Foundation for the Returned Overseas Chinese Scholars. *(Corresponding author: Fan Wu.)*

D. Liu, F. Wu, L. Kong, Y. Luo, and G. Chen are with the Shanghai Key Laboratory of Scalable Computing and Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: liudx@sjtu.edu.cn; wu-fan@sjtu.edu.cn; linghe.kong@sjtu.edu.cn; luoyuan@cs.sjtu.edu.cn; gchen@cs.sjtu.edu.cn).

S. Tang is with the Department of Information Systems, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: tangshaojie@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2016.2606258

exploration system. The process of training data collection normally requires considerable time, devices, as well as human power, especially in a large and complicated indoor environment. Each time one wants to deploy an indoor white space exploration system in a building, she always needs to perform the indoor white space measurement for a period of time, since different buildings have different indoor white space characteristics. Furthermore, the measurement may be unable to be carried out, if there are not sufficient number of spectrum detectors. Therefore, a training-free mechanism for indoor white space exploration is highly attractive.

In this paper, we propose a Training-free Indoor white space exploration MEchanism, namely TIME. TIME does not need any training data, and thus reduces the complexity of system deployment. There are three major challenges when designing TIME: (A) How to reconstruct the complete indoor white space availability map based on the incomplete sensing data? (B) How to deploy the spectrum detectors? (C) How to determine the number of spectrum detectors? Existing works have solved these challenges based on the correlations among indoor white spaces studied from the training data. However, when there is no training data, these challenges cannot be directly solved, due to the lack of a priori knowledge. To solve these problems, we consider Bayesian compressive sensing as the data reconstruction algorithm, since it provides a full posterior density function for the reconstructed data, which is a useful tool for training-free deployment of spectrum detectors. Bayesian compressive sensing cannot be directly applied to the deployment process because of the unique sampling matrices of indoor white spaces. In our work, we design the sampling matrix for indoor white spaces in a novel way, and incrementally deploy the spectrum detectors. We determine the location of each spectrum detector one by one, such that the differential entropy [11] of the reconstruction result is minimized in every iteration. We also propose an innovate algorithm to determine the proper number of spectrum detectors based on the covariances of reconstruction.

The main contributions of this paper are summarized as follows.

- We perform an indoor white space measurement in our complex of offices and labs for a period of two weeks, and study the characteristics of indoor white spaces. The measurement results show the correlations and stability of indoor white spaces. By carefully exploiting these two characteristics, we come up with a mechanism for training-free indoor white space exploration.
- We propose TIME, which is a training-free indoor white space exploration mechanism. It can effectively reduce the deployment complexity of indoor white space exploration system. TIME deploys the spectrum detectors in an incremental way, such that the differential entropy is minimized in every iteration, and reconstructs the complete availability map of indoor white spaces based on the incomplete sensing results. We also propose an algorithm to determine a proper number of spectrum detectors. To the best of our knowledge, TIME is the first training-free indoor white space exploration mechanism.
- We evaluate the performance of TIME with extensive real-world experiments. The evaluation results demonstrate that our training-free mechanism, TIME, achieves competitive performance to the state-of-the-art training-based mechanisms.

The rest of the paper is organized as follows. Section II briefly introduces the Bayesian compressive sensing theory. In Section III, we present our indoor white space measurement experiment and results observed. Section IV shows the detailed design of TIME. Evaluation results are given in Section V. In Section VI, we discuss some practical problems for indoor white space exploration. Section VII discusses related work. We conclude the paper in Section VIII.

II. Preliminary

These years, indoor white space exploration receives more and more attentions, since there are more indoor white spaces than outdoor environment [9]. However, the geo-location database approach, which is commonly utilized in outdoor white space exploration, cannot be directly applied to the indoor scenario, because of the existence of the indoor obstacles (e.g., walls). Existing indoor white space exploration works [9], [10] deploy spectrum detectors at part of the indoor locations, and recover the white space information over all locations through training based data reconstruction method. In our work, we design a Bayesian compressive sensing based reconstruction module, and present an incremental method to deploy spectrum detectors.

Compressive sensing (CS) is a new type of sampling technique, which is widely utilized in different realms [12]–[15]. According to the compressive sensing theory [16], [17], a signal with the feature of sparsity can be reconstructed with a high accuracy. Bayesian compressive sensing (BCS) [18]–[20] is a more advanced compressive sensing data reconstruction method based on sparse Bayesian learning and the Relevance Vector Machine (RVM) [21]. Different from the traditional compressive sensing, Bayesian compressive sensing provides a full posterior density function instead of the single point estimation, and gives error bars of the reconstructed data, which can be used to guide the design of the additional measurements.

Suppose $\boldsymbol{x}$ is an unknown vector in $\mathbb{R}^n$, which has sparse representation under some basis Ψ. Then, $\boldsymbol{x}$ can be represented by

$$\boldsymbol{x} = \Psi\boldsymbol{\omega} = \sum_{i=1}^{n} \psi_i \omega_i, \tag{1}$$

where $\Psi = [\psi_1, \psi_2, \ldots, \psi_n]$ is an $n \times n$ dimension basis, ψ_i is the ith column of Ψ, and ω_i is the corresponding coefficient. Usually the representation basis Ψ is an orthogonal matrix, such as discrete wavelet transform (DWT) basis and discrete cosine transform (DCT) basis. If the coefficient vector $\boldsymbol{\omega}$ is sparse, which means that most components of $\boldsymbol{\omega}$ have negligible amplitudes, then $\boldsymbol{x}$ is sparse under basis Ψ, and can be accurately reconstructed by a small number of measurements

$$\boldsymbol{y} = \Phi\boldsymbol{x} = \Phi\Psi\boldsymbol{\omega}, \tag{2}$$

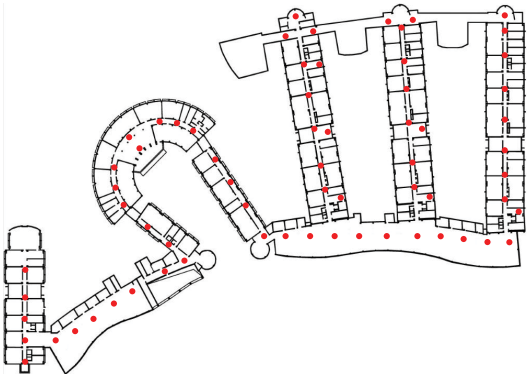


Fig. 1. Floor map with 66 measurement locations, each denoted by a red dot.

where $\Phi = [\phi_1, \phi_2, \ldots, \phi_m]^T$ is an $m \times n$ ($m \ll n$) sampling matrix, and every row ϕ_i^T represents a sample or a measurement. The problem of solving $\boldsymbol{\omega}$ based on $\boldsymbol{y}$, Φ and Ψ is ill-posed since $m \ll n$. However, if $\boldsymbol{\omega}$ is sparse, such an ill-posed problem can be solved via an ℓ_1-regularized formulation [16]

$$\hat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \{\|\boldsymbol{y} - \Phi\Psi\boldsymbol{\omega}\|_2^2 + \rho\|\boldsymbol{\omega}\|_1\}, \tag{3}$$

where the scalar ρ controls the relative importance of precise fit to the measurement and sparseness. In real-world applications, the measurements may be noisy. Bayesian compressive sensing considers a common zero mean Gaussian noise with unknown variance σ^2. The measurement result becomes:

$$\boldsymbol{y} = \Phi\Psi\boldsymbol{\omega} + \boldsymbol{n}, \tag{4}$$

where $\boldsymbol{n} \in \mathbb{R}^m$ is the Gaussian noise. Given Φ, Ψ, and $\boldsymbol{y}$, Bayesian compressive sensing provides the posterior density function for $\boldsymbol{\omega}$ with mean $\boldsymbol{\mu}$ and covariance Σ. The values of $\boldsymbol{\mu}$ and Σ are calculated by the Bayesian compressive sensing module.

Bayesian compressive sensing also provides an effective way to incrementally design the sampling matrix Φ. Suppose the current sampling matrix is Φ, which is an $m \times n$ matrix, a new measurement means adding a new row ϕ_{m+1}^T to Φ. An optimal ϕ_{m+1}^T should be designed to reduce the uncertainty as much as possible, which means to minimize the differential entropy [11]. According to the Bayesian compressive sensing theory [18], minimizing the differential entropy equals to maximize

$$\phi_{m+1}^T \Sigma \phi_{m+1} = \phi_{m+1}^T \mathrm{Cov}(\boldsymbol{\omega})\phi_{m+1} = \mathrm{Var}(\boldsymbol{y}_{m+1}). \tag{5}$$

This means that ϕ_{m+1}^T should maximize the variance of the measurement $\boldsymbol{y}_{m+1}$, and the eigenvector of Σ with the largest eigenvalue is a good choice of ϕ_{m+1}.

III. Indoor White Space Measurement

In this section, we introduce our indoor white space measurement experiment. We perform the measurement on the 3rd floor of a complex of offices and labs for two weeks (Nov. 3, 2014 - Nov. 16, 2014) to explore the characteristics of indoor white spaces. The floor map is shown in Fig. 1. The observations provide important guidelines for designing our training-free indoor white space exploration mechanism.

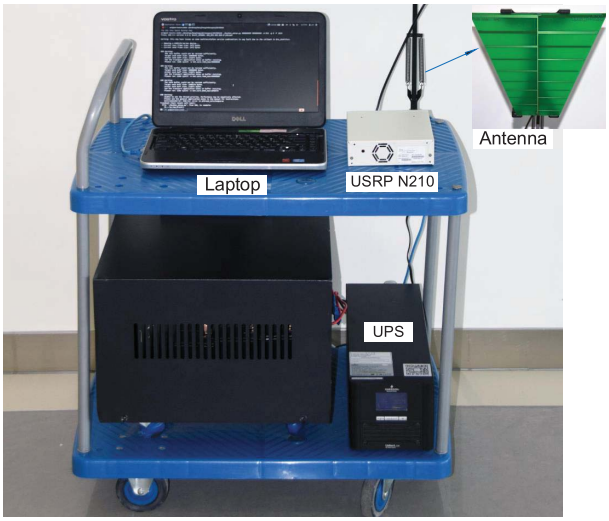


Fig. 2. Our measurement device with a laptop, an USRP N210, a log periodic PCB antenna, and an uninterrupted power supply (UPS).

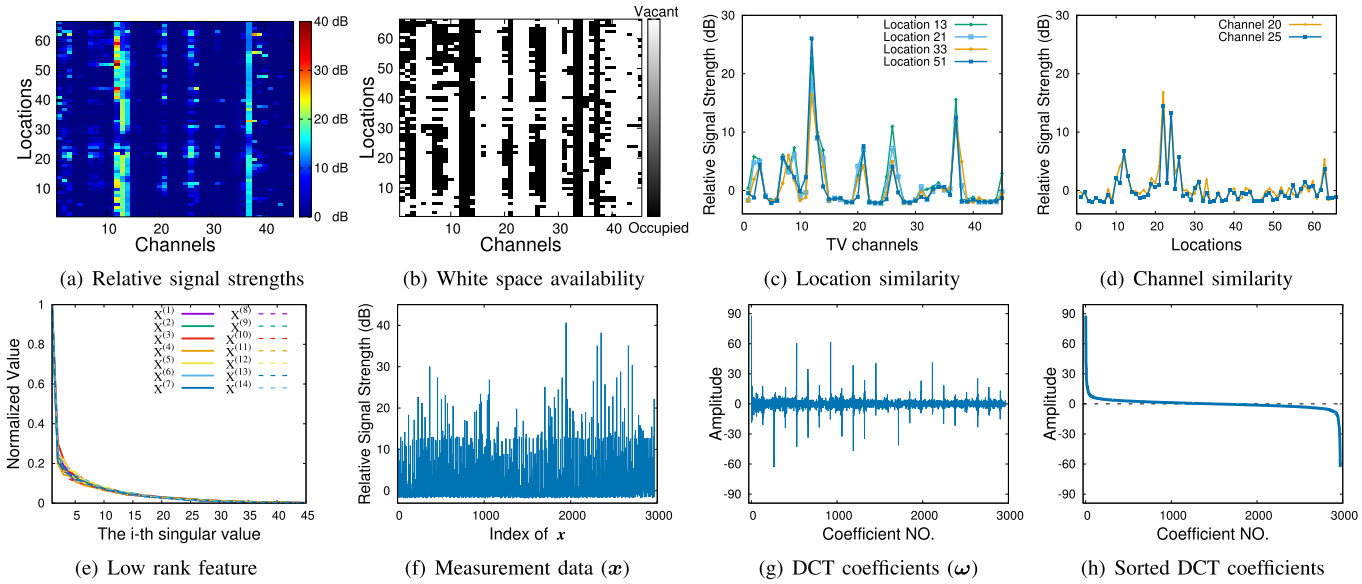
A. Equipment and Setup

Our measurement device (Fig. 2) consists of a laptop, an USRP N210, a log periodic PCB antenna (400-1000 MHz), and an uninterrupted power supply (UPS). The USRP N210 is equipped with SBX daughter-board with 5-10 dBm noise figure.

We measure the UHF TV channels with central frequencies within 470 MHz - 566 MHz and 606 MHz - 870 MHz. The bandwidth of a TV channel is 8 MHz. We measure a total of 45 TV channels. Several methods have been proposed to detect the presence of signal transmissions, such as energy detection, matched filtering, and waveform-based sensing. Here, we choose energy detection, since it is a commonly used spectrum sensing method with low computation and implementation complexity. Please refer to [22] for a survey of spectrum sensing algorithms for cognitive radio applications and [23], [24] for the knowledge of energy detection. We use a GUN Radio FFT program with a bin size 1024 and sample rate 4 MHz. The signal strength of a channel is calculated by averaging the values of all the corresponding bins. We compare the detected signal strengths with a threshold. If a channel's signal strength is greater than the threshold, we consider this channel as occupied, otherwise unoccupied. The same as [9], we use a threshold of -84.5 dBm/8 MHz. Although the threshold is higher than that suggested by FCC (i.e., -114 dBm/8 MHz) [25], we set the threshold in this way due to the limitation of our hardware. We have to note that the vacant channels detected using our prototype may be not safe to access in practice. However, our measurement observations and designed mechanism are not limited to any specific threshold. If the devices (e.g., ThinkRF WSA5000 [26]) are sensitive enough to support a threshold of -114 dBm/8 MHz, the vacant channels detected by our prototype would be safe to use.

B. Measurement Results and Observations

In the indoor white space measurement, we choose 66 typical indoor locations (red dots in Fig. 1), and measure the



(a) Relative signal strengths (b) White space availability (c) Location similarity (d) Channel similarity

(e) Low rank feature (f) Measurement data ($\boldsymbol{x}$) (g) DCT coefficients ($\boldsymbol{\omega}$) (h) Sorted DCT coefficients

Fig. 3. Indoor white space measurement observations.

signal strengths of all the 45 TV channels at each location. Our measurement device is loaded on a movable cart. We measure TV channels' signal strengths at one location after another. We perform one round of measurement each day. The measurement lasts for a period of two weeks. It would be ideal to simultaneously measure the TV channels at all the locations. However, that would need 66 sets of spectrum detectors, which exceed the budget of our research funding. Ying *et al.* [9] have shown that TV channels are short-time stable in terms of hours. In our work, a round of measurement over all 66 locations needs no more than two hours, and thus the results of the asynchronous measurement can still provide some insights. Actually, there exist correlations among indoor white spaces at different times. In this paper, we just study the space and channel correlation. Utilizing time correlation to improve the system performance is one of our future work.

For the convenience of representation, we convert the absolute signal strengths to relative values by subtracting the threshold (-84.5 dBm) from them. After each round of measurement, we get a 66×45 matrix

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,44} & x_{1,45} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,44} & x_{2,45} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{66,1} & x_{66,2} & \cdots & x_{66,44} & x_{66,45} \end{pmatrix}, \tag{6}$$

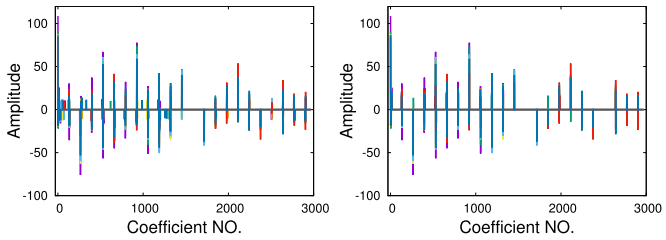
where $x_{i,j}$ refers to channel j's relative signal strength at location i. Since Bayesian compressive sensing can reconstruct vector data with a high accuracy, for the convenience of calculation, we transform matrix X to a vector in a row-wise order:

$$\boldsymbol{x} = (\underbrace{x_{1,1}, x_{1,2}, \ldots, x_{1,45}}_{\text{Location 1}}, \underbrace{x_{2,1}, \ldots, x_{2,45}}_{\text{Location 2}}, \ldots, x_{66,45})^T, \tag{7}$$

where $\boldsymbol{x}$ is a vector of 2970 dimensions ($2970 = 66 \times 45$). We perform one round of measurement every day, and get a corresponding relative signal strength matrix X, which contains the relative signal strengths of the 45 TV channels at 66 indoor locations. After the two weeks' measurement, we get a total of 14 measurement data sets. We use $X^{(i)}$ and $\boldsymbol{x}^{(i)}$ to denote the measurement data set in the ith day. From the indoor white space measurement, we have the following observations.

First, we draw measurement results of the first day (Nov. 3, 2014). Fig. 3(a) shows the 45 TV channels' relative signal strengths over all the 66 indoor locations. We observe that even for the same channel, its signal strengths may differ from each other at different locations. Fig. 3(b) is the corresponding availability map of indoor white spaces, where black blocks refer to occupied channels and white blocks refer to white spaces. The state of each channel is different at different locations, and this means that a channel may be occupied at some indoor locations while vacant at others. The differences on a TV channel's signal strengths are caused by the complicated indoor obstacles (e.g., walls), and make the commonly used outdoor white space exploration approach (i.e., geo-location database) invalid in the indoor environment. If we directly apply the outdoor exploration methods to the indoor environment, we may lose a number of potential white spaces and get a conservative result.

Second, the measurement results show that there exist correlations among the indoor white spaces, which means that the signal strengths of TV channels are not independent at different channels or locations. Intuitively, correlations of indoor white spaces represent the redundancies in X and $\boldsymbol{x}$. The redundancies in matrices indicate their property of low rank, while those in vectors infer their property of sparsity. We use the first day's measurement data as an example to illustrate the correlations among indoor white spaces. Fig. 3(c) demonstrates that the relative signal strengths of all 45 TV channels are similar at location 13, 21, 33, and 51, which means that the 13rd, 21st, 33rd, and 51st rows in $X^{(1)}$ have similar values. Similarly, Fig. 3(d) shows that the relative



(a) Absolute amplitudes > 10 (b) Top 50 coefficients

Fig. 4. Stability of correlations.

signal strengths of channel 20 and 25 are similar at all 66 locations. This means that the 20th and 25th columns in $X^{(1)}$ are similar. These two figures demonstrate that the rows and columns of $X^{(1)}$ are not independent. Besides, for a matrix X, if the vector that contains all its singular values is sparse, then the matrix is low rank. In Fig. 3(e), we draw the distribution of the singular values from the 14 relative signal strength matrices. The X-axis represents the ith singular values, and the Y-axis stands for the normalized singular values. Fig. 3(e) suggests that most of the energy is contributed by the top few singular values, which reveals the approximate low rank structure of $X^{(1)}, X^{(2)}, ..., X^{(14)}$.

We also explore the sparsity of vector $\boldsymbol{x}$. Fig. 3(f) plots the value of $\boldsymbol{x}^{(1)}$. The X-axis represents the index of $\boldsymbol{x}^{(1)}$, and the Y-axis stands for the values. Fig. 3(g) shows the corresponding coefficients $\boldsymbol{\omega}^{(1)}$ after the discrete cosine transform (DCT)

$$\boldsymbol{\omega}^{(1)} = \Psi^{-1}\boldsymbol{x}^{(1)}, \tag{8}$$

where Ψ refers to the DCT basis. The sparsity of $\boldsymbol{x}^{(1)}$ on the basis Ψ means that the vector $\boldsymbol{\omega}^{(1)}$ should be sparse or approximately sparse. In order to intuitively illustrate the sparsity of $\boldsymbol{\omega}^{(1)}$, we sort the elements of $\boldsymbol{\omega}^{(1)}$ in a descending order, and show the results in Fig. 3(h). We observe that most of the coefficients have negligible amplitudes. Actually, only 81 of the 2970 have relatively large absolute coefficients (> 10) in $\boldsymbol{\omega}^{(1)}$. The average number of coefficients with absolute value larger than 10 is 86.4 for all of the 14 rounds of measurement results. This means that $\boldsymbol{\omega}$ is sparse and the DCT basis is a sparse basis for the measurement data $\boldsymbol{x}$.

At last, we study the stability of the indoor white spaces' correlations. Here, we focus on vector $\boldsymbol{x}$. We perform DCT on 7 rounds of measurement results $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, ..., \boldsymbol{x}^{(7)}$, and calculate their DCT coefficients $\boldsymbol{\omega}^{(1)}, \boldsymbol{\omega}^{(2)}, ..., \boldsymbol{\omega}^{(7)}$. Fig. 4(a) illustrates all the DCT coefficients from these 7 rounds of measurement results, whose absolute amplitudes are larger than 10. Fig. 4(b) shows the corresponding top 50 coefficients with the largest absolute amplitudes. It is observed that the value of these large coefficients in $\boldsymbol{\omega}^{(1)}, \boldsymbol{\omega}^{(2)}, ..., \boldsymbol{\omega}^{(7)}$ may be different in different measurement data sets, but the locations of them are relatively consistent.

C. Summary

The main observations of the indoor white space measurement experiment are as follows.

- The signal strengths of TV channels differ from each other at different indoor locations.
- There exist correlations among indoor white spaces.

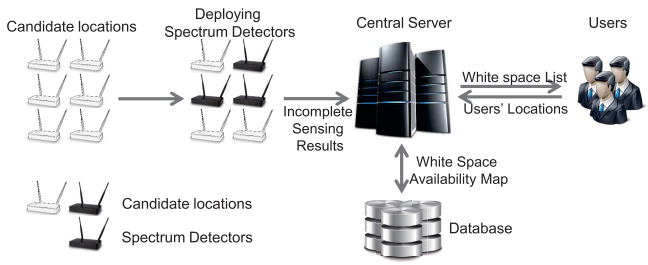


Fig. 5. The system model: a part of the candidate locations are selected to deploy spectrum detectors. The partial sensing results are submitted to the central server where the complete indoor white space availability map is calculated based on the partial sensing results. Users could get the status of TV channels by submitting their indoor locations.

- The correlations among indoor white spaces are stable over time.

These observations provide important guidelines for designing our training-free indoor white space exploration mechanism.

IV. System Design

In this section, we provide the details of our system design. We first introduce the system model, and then present the data reconstruction algorithm. Next, we propose the incremental deployment algorithm of spectrum detectors. At last, we present the method determining proper number of spectrum detectors.

A. System Model

The system architecture of TIME is shown in Fig. 5. At first, we select N typical indoor locations that cover the main indoor areas, and call them *candidate locations*. Just like the existing works [9], [10], we assume that the candidate locations are easily selected to cover each room and corridor of the target building. Given N candidate locations, we aim to deploy spectrum detectors at M of them. We denote the set of candidate locations using

$$\mathcal{V} = \{1, 2, \cdots, N\}, \tag{9}$$

while the set of locations with spectrum detectors deployed is denoted by

$$\mathcal{S} = \{l_1, l_2, \cdots, l_M\}, \tag{10}$$

where $1 \leq l_1 < l_2 < \cdots < l_M \leq N$. Each spectrum detector i is deployed at candidate location l_i. It is clear that $\mathcal{S}$ is a subset of $\mathcal{V}$ ($\mathcal{S} \subseteq \mathcal{V}$). Ideally, if we deploy a spectrum detector at each of the N candidate locations ($M = N, \mathcal{S} = \mathcal{V}$), we can perform a complete sensing, and get an accurate white space availability map at all candidate locations. However, this kind of approach costs too much due to the need of a large number of spectrum detectors. Instead, we only deploy spectrum detectors at a part of the candidate locations ($M < N, \mathcal{S} \subsetneq \mathcal{V}$), and perform an incomplete sensing.

We assume that the relative signal strengths of TV channels at candidate location $k \in \mathcal{V}$ store in

$$\boldsymbol{x}_k = (x_{k,1}, x_{k,2}, \cdots, x_{k,C})^T \in \mathbb{R}^C, \tag{11}$$

where C refers to the number of TV channels, and $x_{k,j}$ is the relative signal strength of channel j at location k.

The complete indoor white space information, which stores in $\boldsymbol{x}$, can be denoted by

$$\boldsymbol{x} = (\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_N^T)^T \in \mathbb{R}^{NC}. \tag{12}$$

Each spectrum detector i, which is deployed at candidate location l_i, submit $\boldsymbol{x}_{l_i}$ to the central server at regular time intervals. The central server then collects the incomplete sensing results $\boldsymbol{x}_{l_1}, \boldsymbol{x}_{l_2}, \cdots, \boldsymbol{x}_{l_M}$ in a vector

$$\boldsymbol{y} = (\boldsymbol{x}_{l_1}^T, \boldsymbol{x}_{l_2}^T, \cdots, \boldsymbol{x}_{l_M}^T)^T \in \mathbb{R}^{MC}, \tag{13}$$

and performs the data reconstruction to estimates the complete white space information $\boldsymbol{x}$. After that, the central server calculates the indoor white space availability map based on the reconstructed $\boldsymbol{x}$, and stores the map into its database. When users want to access indoor white spaces, they submit their indoor locations to the central server. Given a user's indoor location, the central server finds the candidate location, which is the nearest to the user, and returns a list of the white spaces according to the candidate location. In the returned list of white spaces, the central server excludes the channels occupied by licensed users and existing unlicensed white space users within the estimated interfering distance.

Same as [9] and [10], we assume that the white space availability at a given location is the same as that of its nearest candidate one. We leave the case, in which this assumption does not hold, to our future work.

The remaining part of this Section is organized as follows. In Section IV-B, we propose the Bayesian compressive sensing based reconstruction algorithm. Then, we present the incremental deployment method of spectrum detectors in Section IV-C. At last, Section IV-D shows how to determine the number of spectrum detectors.

B. Availability Map Reconstruction

As we mentioned before, in order to perform efficient indoor white space exploration, we can only deploy spectrum detectors at part of the candidate locations, and get an incomplete sensing result $\boldsymbol{y}$. Here, we consider that the spectrum detectors have already been deployed at a part of the candidate locations. We will present the algorithm for deployment of spectrum detectors in Section IV-C. The central server gets the incomplete sensing result $\boldsymbol{y} \in \mathbb{R}^{MC}$, and tries to estimate the complete data $\boldsymbol{x} \in \mathbb{R}^{NC}$. We have

$$\boldsymbol{y} = \Phi \boldsymbol{x}, \tag{14}$$

where Φ is a sampling matrix representing the deployment of spectrum detectors. Φ is an $MC \times NC$ matrix:

$$\Phi = \begin{pmatrix} \Phi_{l_1,1} & \Phi_{l_1,2} & \cdots & \Phi_{l_1,N-1} & \Phi_{l_1,N} \\ \Phi_{l_2,1} & \Phi_{l_2,2} & \cdots & \Phi_{l_2,N-1} & \Phi_{l_2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Phi_{l_M,1} & \Phi_{l_M,2} & \cdots & \Phi_{l_M,N-1} & \Phi_{l_M,N} \end{pmatrix}, \tag{15}$$

where $\Phi_{i,j}$ is a $C \times C$ matrix, and

$$\Phi_{i,j} = \begin{cases} I_C & \text{if } i = j, \\ \mathbf{0} & \text{otherwise}, \end{cases} \tag{16}$$

with I_C referring to the $C \times C$ identity matrix.

Our measurements in Section III have shown that different TV channels or different indoor locations are dependent, and there exists redundancy in vector $\boldsymbol{x}$, which means that $\boldsymbol{x}$ has a sparse representation on an orthogonal basis. In this work, we choose DCT basis, as it has been widely adopted in signal processing and image compression. We have also verified that DCT works well on the indoor white space data $\boldsymbol{x}$ in Section III-B. Let $\boldsymbol{\omega}$ be the coefficients of $\boldsymbol{x}$ under DCT basis. We have

$$\boldsymbol{y} = \Phi \boldsymbol{x} = \Phi \Psi \boldsymbol{\omega}, \tag{17}$$

where Ψ refers to the DCT basis and $\Psi \Psi^T = I$. To deal with the measurement noise, we add a zero-mean Gaussian noise $\boldsymbol{n}$, which is commonly used, to the measurement results

$$\boldsymbol{y} = \Phi \Psi \boldsymbol{\omega} + \boldsymbol{n}, \tag{18}$$

where $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2)$.

After obtaining $\boldsymbol{y}$, we are ready to estimate $\boldsymbol{x}$. Although there is only one non-zero element in every row of the sampling matrix Φ according to its definition, we can still use compressive sensing to do the reconstruction [13]. This is because $\boldsymbol{x}$ is sparse under Ψ, and what we actually reconstruct is $\boldsymbol{w}$, where each component of $\boldsymbol{y}$ is a combination of $\boldsymbol{w}$. According to the Bayesian compressive sensing theory [18], given $\boldsymbol{y}$, we can get the posterior for $\boldsymbol{\omega}$ as a multivariate Gaussian distribution with mean and covariance:

$$\boldsymbol{\mu} = \frac{1}{\sigma^2} \Sigma \Psi^T \Phi^T \boldsymbol{y}, \tag{19}$$

$$\Sigma = \left(\frac{1}{\sigma^2} \Psi^T \Phi^T \Phi \Psi + A \right)^{-1}, \tag{20}$$

where A refers to the parameters in the prior of $\boldsymbol{\omega}$. Here, we use the hierarchical sparseness prior in a similar way as [18] and [21]. The value of A and σ^2 can be estimated based on the RVM model. Due to the space limitation, we omit the calculation process. Please refer to [21] for more details. In our work, we use a fast RVM algorithm [27] to improve the running speed. Since $\boldsymbol{x} = \Psi \boldsymbol{\mu}$, the posterior density function of $\boldsymbol{x}$ is also a multivariate Gaussian distribution with mean and covariance:

$$\mathrm{E}(\boldsymbol{x}) = \Psi \boldsymbol{\mu}, \tag{21}$$

$$\mathrm{Cov}(\boldsymbol{x}) = \Psi \Sigma \Psi^T. \tag{22}$$

The central server receives the incomplete sensing result $\boldsymbol{y}$ at regular time intervals, and then performs the Bayesian compressive sensing based data reconstruction to obtain the posterior density function of the complete data $\boldsymbol{x}$. We use $\mathrm{E}(\boldsymbol{x})$ as the estimation of $\boldsymbol{x}$, and the diagonal elements of $\mathrm{Cov}(\boldsymbol{x})$ as error bars to describe the estimation accuracy. Given $\hat{\boldsymbol{x}} = \mathrm{E}(\boldsymbol{x})$, we can easily change it to the matrix form $\hat{X}$, which contains the reconstructed relative signal strengths, based on a reverse process of equation (7). Intuitively, the white space availability map should be calculated by comparing the elements of $\hat{X}$ with 0. But FCC requires that unlicensed devices should not interfere with the licensed signal transmissions. This means that the white space exploration mechanism should avoid misidentifying an occupied TV channel as

vacant (false alarms). Hence, we compare the reconstructed relative signal strength $\hat{X}(i, j)$ with the protection range PR (which is less than 0), instead of 0, to reduce false alarms. The availability map is defined as

$$\mathrm{MAP}(i, j) = \begin{cases} 1 & \text{if } \hat{X}(i, j) < PR, \\ 0 & \text{if } \hat{X}(i, j) \geq PR, \end{cases} \tag{23}$$

where $\mathrm{MAP}(i, j) = 1$ if channel j is vacant at location i, and $\mathrm{MAP}(i, j) = 0$ otherwise.

C. Deployment of Spectrum Detectors

Deployment of Spectrum detectors is an important part of the indoor white space exploration, where different deployments may yield different system performances. However, existing works only focus on the training based deployment method, which means that the channels' information at different locations should be collected in advance as training data. Then, the indoor spectrum detectors are deployed based on the training data. The ideal way to collect the training data is to deploy spectrum detectors at all candidate locations, and to perform simultaneous signal strength collections. However, it is expensive and energy consuming. Even if we adopt the asynchronous way as described in Section III, collecting training data is also a hard job, especially when we need to deal with a large and complicated indoor environment.

In our work, we propose an incremental spectrum detectors deployment method, without a training process. Assuming that there are M spectrum detectors, we first randomly choose r candidate locations, and deploy a spectrum detector at each of them. Actually, we can also deploy the first r spectrum detectors according to the "importance" (e.g., spectrum demand) of different candidate locations. Then, we deploy the remaining spectrum detectors in an incremental way. The detailed deployment algorithm is shown in Algorithm 1. In Algorithm 1, $\mathcal{V}$ refers to the set of all candidate locations, $\mathcal{S}$ refers to the candidate locations with spectrum detectors, and $\mathcal{V} \setminus \mathcal{S}$ refers to the candidate locations without spectrum detector. Φ_d is the current sampling matrix with deployed spectrum detectors, and is initialized as $\varnothing$. We update its value when a new spectrum detector is deployed. If we deploy a spectrum detector at location $\hat{l}$, we can explore all the C channels at location $\hat{l}$, and thus add C rows to Φ_d. The random deployment process is described in lines 2-6. After that, we get the current sampling matrix Φ_d and the corresponding measurement result $\boldsymbol{y}_d = \Phi_d \boldsymbol{x} = \Phi_d \Psi \boldsymbol{\omega}$.

Then, we incrementally choose the locations of spectrum detectors one by one. In each iteration, we perform the Bayesian compressive sensing data reconstruction (function BCS in line 10) given Φ_d and $\boldsymbol{y}_d$. Here, the noise variance σ^2 is an input to BCS, which re-estimates it and calculates $\boldsymbol{\mu}_d$ and Σ_d. Initially, σ^2 is set to 1% of $\boldsymbol{y}_d$'s variance. According to the Bayesian compressive theory (equations (19)-(22)), the mean and covariance of $\boldsymbol{x}$ are

$$\mathrm{E}(\boldsymbol{x}) = \frac{1}{\sigma^2} \Psi \Sigma_d \Psi^T \Phi_d^T \boldsymbol{y}_d, \tag{24}$$

$$\mathrm{Cov}(\boldsymbol{x}) = \Psi \Sigma_d \Psi^T, \tag{25}$$

Algorithm 1 Spectrum Detectors Deployment

Input : $\mathcal{V}$: Set of all candidate locations,
Ψ: DCT basis,
r: Number of spectrum detectors randomly deployed,
M: Total number of spectrum detectors,
Output: $\mathcal{S}$: Set of candidate locations with spectrum detectors.

1 $\mathcal{S} \leftarrow \varnothing$; $\Phi_d \leftarrow \varnothing$;
2 **for** $i = 1$ ***to*** r **do**
3 *Randomly choose a location* $\hat{l}$ *from* $\mathcal{V} \setminus \mathcal{S}$;
4 *Deploy a spectrum detector at location* $\hat{l}$;
5 $\mathcal{S} \leftarrow \mathcal{S} \cup \{\hat{l}\}$;
6 $\Phi_d \leftarrow [\Phi_d^T, \Phi_{\hat{l}}^T]^T$;
7 **for** $i = r + 1$ ***to*** M **do**
8 $\boldsymbol{y_d} \leftarrow$ *Measurement result under* Φ_d;
9 $\sigma_{ini}^2 \leftarrow var(\boldsymbol{y}_d)/100$;
10 $[\boldsymbol{\mu}_d, \Sigma_d, \sigma^2] \leftarrow BCS(\boldsymbol{y}_d, \Psi, \Phi_d, \sigma_{ini}^2)$;
11 $\hat{l} \leftarrow NextDetector(\mathcal{V}, \mathcal{S}, \Psi, \Sigma_d)$;
12 *Deploy a spectrum detector at location* $\hat{l}$;
13 $\mathcal{S} \leftarrow \mathcal{S} \cup \{\hat{l}\}$;
14 $\Phi_d \leftarrow [\Phi_d^T, \Phi_{\hat{l}}^T]^T$;
15 **return** $\mathcal{S}$

where $\Sigma_d = (\frac{1}{\sigma^2} \Psi^T \Phi_d^T \Phi_d \Psi + A)^{-1}$. After the data reconstruction, we try to find the currently best location to deploy next spectrum detector. Since the estimated posterior on $\boldsymbol{x}$ is a multivariate Gaussian distribution with means $\mathrm{E}(\boldsymbol{x})$ and covariances $\mathrm{Cov}(\boldsymbol{x})$, the differential entropy [11] of $\boldsymbol{x}$ is

$$\begin{aligned} h(\boldsymbol{x}) &= -\int p(\boldsymbol{x}) \log p(\boldsymbol{x}) d\boldsymbol{x} \\ &= \frac{1}{2} \log |\mathrm{Cov}(\boldsymbol{x})| + c_1 \\ &= \frac{1}{2} \log |\Psi \Sigma_d \Psi^T| + c_1 \\ &= \frac{1}{2} \log |\Sigma_d| + c_2 \\ &= -\frac{1}{2} \log \left| \frac{1}{\sigma^2} \Psi^T \Phi_d^T \Phi_d \Psi + A \right| + c_2, \end{aligned} \tag{26}$$

where c_2 is independent of the sampling matrix Φ_d. Minimizing the differential entropy is equal to maximizing the information we get [18]. So, when deploying the next spectrum detector, we choose a location that minimizes the differential entropy in equation (26). Actually, the differential entropy based method has been previously studied in the machine learning community under the name of experimental design or active learning [28], [29]. Deploying one more spectrum detector means adding C rows into Φ_d. Hence, we cannot directly apply the result presented by [18] (equation (5)). If we deploy the next spectrum detector at location $l \in \mathcal{V} \setminus \mathcal{S}$, we add an extra $C \times (NC)$ matrix Φ_l to Φ_d,

$$\Phi_l = (\Phi_{l,1}, \Phi_{l,2}, \cdots, \Phi_{l,N}), \tag{27}$$

where $\Phi_{i,j}$ is defined in equation (16). The new sampling matrix after deploying a spectrum detector at location l is

$$\Phi_{new} = \begin{bmatrix} \Phi_d \\ \Phi_l \end{bmatrix}. \tag{28}$$

We define $h_{new}(\boldsymbol{x})$ as the differential entropy under the new sampling matrix Φ_{new}, and have

$$\begin{aligned} & h_{new}(\boldsymbol{x}) - h(\boldsymbol{x}) \\ &= -\frac{1}{2}\log\left|\frac{1}{\sigma^2}\Psi^T\Phi_{new}^T\Phi_{new}\Psi + A\right| \\ &\quad +\frac{1}{2}\log\left|\frac{1}{\sigma^2}\Psi^T\Phi_d^T\Phi_d\Psi + A\right| \\ &= -\frac{1}{2}\log\left|\frac{1}{\sigma^2}\Psi^T\Phi_d^T\Phi_d\Psi + A + \frac{1}{\sigma^2}\Psi^T\Phi_l^T\Phi_l\Psi\right| \\ &\quad +\frac{1}{2}\log\left|\frac{1}{\sigma^2}\Psi^T\Phi_d^T\Phi_d\Psi + A\right| \\ &= -\frac{1}{2}\log\left|\Sigma_d^{-1} + \frac{1}{\sigma^2}\Psi^T\Phi_l^T\Phi_l\Psi\right| + \frac{1}{2}\log\left|\Sigma_d^{-1}\right| \\ &= -\frac{1}{2}\log\left|I_{NC} + \frac{1}{\sigma^2}\Psi^T\Phi_l^T\Phi_l\Psi\Sigma_d\right|, \end{aligned} \tag{29}$$

where I_{NC} is an $(NC)\times(NC)$ identity matrix. Since it is time consuming to calculate the determinant of the $(NC)\times(NC)$ matrix $I_{NC} + \frac{1}{\sigma^2}\Psi^T\Phi_l^T\Phi_l\Psi\Sigma_d$, we transform equation (29) to:

$$\begin{aligned} & h_{new}(\boldsymbol{x}) - h(\boldsymbol{x}) \\ &= -\frac{1}{2}\log\left|I_{NC} + \frac{1}{\sigma^2}\Psi^T\Phi_l^T\Phi_l\Psi\Sigma_d\right| \\ &= -\frac{1}{2}\log\left(\frac{|\Phi_l\Psi|\cdot|I_{NC} + \frac{1}{\sigma^2}\Psi^T\Phi_l^T\Phi_l\Psi\Sigma_d|\cdot|\Psi^T\Phi_l^T|}{|\Phi_l\Psi|\cdot|\Psi^T\Phi_l^T|}\right) \\ &= -\frac{1}{2}\log\left(\frac{|\Phi_l\Psi\Psi^T\Phi_l^T + \frac{1}{\sigma^2}\Phi_l\Psi\Psi^T\Phi_l^T\Phi_l\Psi\Sigma_d\Psi^T\Phi_l^T|}{|\Phi_l\Psi\Psi^T\Phi_l^T|}\right) \\ &= -\frac{1}{2}\log\left(\frac{|\Phi_l\Phi_l^T + \frac{1}{\sigma^2}\Phi_l\Phi_l^T\Phi_l\Psi\Sigma_d\Psi^T\Phi_l^T|}{|\Phi_l\Phi_l^T|}\right) \\ &= -\frac{1}{2}\log\left(\frac{|\Phi_l\Phi_l^T|\cdot|I_C + \frac{1}{\sigma^2}\Phi_l\Psi\Sigma_d\Psi^T\Phi_l^T|}{|\Phi_l\Phi_l^T|}\right) \\ &= -\frac{1}{2}\log\left|I_C + \frac{1}{\sigma^2}\Phi_l\Psi\Sigma_d\Psi^T\Phi_l^T\right|, \end{aligned} \tag{30}$$

where I_C is a $C\times C$ identity matrix. We now just need to calculate the determinant of $I_C + \frac{1}{\sigma^2}\Phi_l\Psi\Sigma_d\Psi^T\Phi_l^T$, which is a $C\times C$ matrix. Minimizing $h_{new}(\boldsymbol{x})$ is equal to maximizing $|I_C + \frac{1}{\sigma^2}\Phi_l\Psi\Sigma_d\Psi^T\Phi_l^T|$, where σ^2 and Σ_d can be calculated based on the current measurement results. For each of the remaining candidate location l without spectrum detector ($l \in \mathcal{V}\setminus\mathcal{S}$), we generate its corresponding Φ_l, and deploy the next spectrum detector at location $\hat{l}$, which maximizes $|I_C + \frac{1}{\sigma^2}\Phi_{\hat{l}}\Psi\Sigma_d\Psi^T\Phi_{\hat{l}}^T|$. This process is shown in Algorithm 2. In this way, we choose the current most informative location by minimizing the differential entropy. We repeat the above process (lines 8-14 in Algorithm 1) until all the spectrum detectors are deployed.

Algorithm 2 $NextDetector(\mathcal{V}, \mathcal{S}, \Psi, \Sigma_d)$

Input : $\mathcal{V}$: Set of all candidate locations,
$\mathcal{S}$: Set of candidate locations with spectrum detectors,
Ψ: DCT basis,
Σ_d: Covariance matrix.
Output: $\hat{l}$: Location of next spectrum detector.

$\hat{E} \leftarrow -\infty$;
foreach $l \in \mathcal{V}\setminus\mathcal{S}$ **do**
  $E \leftarrow |I_C + \frac{1}{\sigma^2}\Phi_l\Psi\Sigma_d\Psi^T\Phi_l^T|$;
  if $E > \hat{E}$ **then**
    $\hat{E} \leftarrow E$; $\hat{l} \leftarrow l$;
return $\hat{l}$

Algorithm 3 Spectrum Detector Location Adjustment

Input : $\mathcal{V}$: Set of all candidates locations,
$\mathcal{S}$: Set of candidate locations with spectrum detectors,
Ψ: DCT basis.
Output: $\mathcal{S}_{ad}$: Set of deployment locations after adjustment

$\mathcal{S}_{ad} \leftarrow \mathcal{S}$;
repeat
  $error_{min} \leftarrow +\infty$; $l_{min} \leftarrow 0$;
  foreach $l \in \mathcal{S}_{ad}$ **do**
    $\mathcal{L} \leftarrow \mathcal{S}_{ad}\setminus\{l\}$;
    $\boldsymbol{y}_{\mathcal{L}} \leftarrow Measurement\ result\ under\ \Phi_{\mathcal{L}}$;
    $\boldsymbol{y}_l \leftarrow Measurement\ result\ under\ \Phi_l$;
    $\sigma_{ini}^2 \leftarrow var(\boldsymbol{y}_{\mathcal{L}})/100$;
    $[\boldsymbol{\mu}, \Sigma, \sigma^2] \leftarrow BCS(\boldsymbol{y}_{\mathcal{L}}, \Psi, \Phi_{\mathcal{L}}, \sigma_{ini}^2)$;
    $E(\boldsymbol{x}) \leftarrow \Psi\boldsymbol{\mu}$;
    $error \leftarrow \|\boldsymbol{y}_l - \Phi_l E(\boldsymbol{x})\|_2/\|\boldsymbol{y}_l\|_2$;
    if $error < error_{min}$ **then**
      $error_{min} \leftarrow error$; $l_{min} \leftarrow l$; $\Sigma_{min} \leftarrow \Sigma$;
  $\mathcal{S}_{ad} \leftarrow \mathcal{S}_{ad}\setminus\{l_{min}\}$;
  $\hat{l} \leftarrow NextDetector(\mathcal{V}, \mathcal{S}_{ad}, \Psi, \Sigma_{min})$;
  $\mathcal{S}_{ad} \leftarrow \mathcal{S}_{ad}\cup\{\hat{l}\}$;
until $l_{min} = \hat{l}$;
return $\mathcal{S}_{ad}$;

Our incremental spectrum detectors deployment method does not rely on any training data. Since the first r detectors are randomly deployed, the locations selected may be not optimal, especially when the number of spectrum detectors is small. In order to improve the system performance, we adjust locations of some spectrum detectors after the incremental deployment. The detailed adjustment method is shown in Algorithm 3. $\mathcal{S}_{ad}$ stands for the set of deployment locations after adjustment, and is initialized as $\mathcal{S}_{ad} \leftarrow \mathcal{S}$. For each location $l \in \mathcal{S}_{ad}$, we estimate the TV channels' signal strengths at it only based on the knowledge of the other locations with

spectrum detectors deployed $\mathcal{L} = \mathcal{S}_{ad} \setminus \{l\}$. Then, we compare the estimated values with the measured ones at location l, and calculate the relative error (line 11), where for any vector $v \in \mathbb{R}^n$, $\|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$. A small error indicates that the channels at location l can be accurately estimated based on the readings at the other locations $\mathcal{L}$, and thus l is not an essential location. If we take away the spectrum detector at location l, and only use those at $\mathcal{L}$, we can also get a good estimation result. In contrast, a large relative error means that l is an essential location, because we cannot correctly reconstruct the values at location l based on the other locations' readings. Hence, we calculate the relative reconstruction error for every location in $\mathcal{S}_{ad}$, choose the location l_{min} with the minimum relative error, and remove it from $\mathcal{S}_{ad}$. After that, we add a new location $\hat{l}$ based on our incremental deployment method (lines 15-16). We repeat the above process until the location we remove from $\mathcal{S}_{ad}$ is the same as the location that we add based on the incremental deployment method, which means $l_{min} = \hat{l}$.

D. Proper Number of Spectrum Detectors

In the previous subsection, we presented the incremental algorithm for deploying all the M spectrum detectors, such that the differential entropy is minimized in each iteration. However, in practice, to reduce the cost of system deployment, it is always desirable to deploy an appropriate and possibly smaller number of spectrum detectors, if the expected reconstruction accuracy can still be acceptable. Therefore, in this part, we study the problem of determining a proper number of spectrum detectors.

In our work, we determine the proper number of spectrum detectors based on the error bars of the reconstructed data $\boldsymbol{x}$. In Algorithm 1, we randomly deploy a number of detectors, and then incrementally deploy the remaining detectors according to the intermediate data reconstruction result. In the incremental deploying process, before deploying each detector, we perform a Bayesian compressive sensing based data reconstruction (line 10 in Algorithm 1), and get $\boldsymbol{\omega}$'s covariance Σ_d. According to equation (25), the covariance of $\boldsymbol{x}$ is $\mathrm{Cov}(\boldsymbol{x}) = \Psi \Sigma_d \Psi^T$. We define the error bar as the diagonal of $\mathrm{Cov}(\boldsymbol{x})$

$$errbar = \mathrm{diag}(\Psi \Sigma_d \Psi^T), \tag{31}$$

to measure the uncertainty of $\boldsymbol{x}$. We set $AvgErr$ as the average error that we can tolerate in reconstructing $\boldsymbol{x}$. In each iteration, after getting the reconstructed data, we calculate $errbar$. If the average value in $errbar$ is less than $AvgErr$, we stop the deployment, otherwise we keep on repeating the above process. We note that the above presented spectrum detector location adjustment algorithm can also be applied to the allocation reached in this part. For example, we can stop the adjustment when $errbar$ meets the requirements.

V. Performance Evaluation

In this section, we perform extensive experiments to evaluate the performance of TIME. First, we compare the performance of TIME with two state-of-the-art training-based indoor white space exploration mechanisms, WISER [9] and FIWEX [10]. Then, we evaluate the performance of TIME at different locations with different channels. Next, we study the performance of the incremental spectrum detectors deployment method by comparing with a random deployment. We also show the feasibility of our deployment method. After that, we evaluate the algorithm determining the proper number of spectrum detectors. At last, we combine TIME with the geo-location database approach, and study the corresponding performance improvement.

A. Methodology

The evaluation is based on our indoor white space measurement experiment. In Section III, we measure the signal strengths of 45 TV channels at 66 indoor locations for a period of two weeks, and get 14 spectrum measurement data sets (X and $\boldsymbol{x}$). Although due to limitations of time, we can only consider 66 locations in the experiment, which may not cover all the rooms and corridors in our complex, the evaluation results can still give a good indication of the performance of our mechanism. In practice, if a sufficient number of candidate locations are considered, TIME is expected to have satisfactory performance. Furthermore, TIME is not limited to any specific indoor environment, because it explores and utilizes the general correlations of indoor white spaces [9], [10]. In the evaluation, we feed the first day's measurement results into our mechanism for the seek of spectrum detectors deployment, and evaluate its performance using the following 13 data sets. We randomly deploy the initial 2 spectrum detectors ($r = 2$), and then incrementally deploy the remaining spectrum detectors based on the intermediate sensing results. The protection range PR is set to -0.7 by default. We consider FA Rate, WS Loss Rate, and Reconstruction Error as the metrics of system performance. Their definitions are as follows.

- *False Alarm Rate (FA Rate):* is the ratio between the number of channels mis-identified as vacant and the total number of vacant channels identified by the system.
- *White Space Loss Rate (WS Loss Rate):* is the ratio between the number of channels mis-identified as occupied and the total number of actually vacant channels.
- *Reconstruction Error:* is defined as

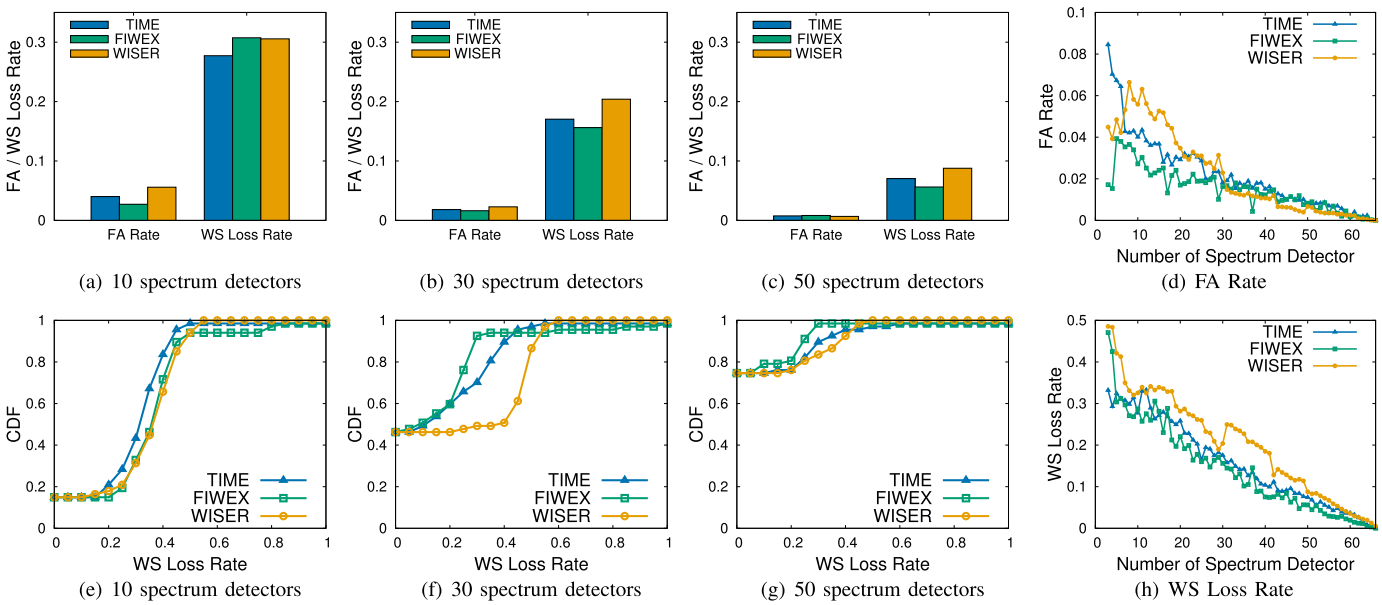
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2,$$

where $\boldsymbol{x} \in \mathbb{R}^{NC}$ is the real indoor white space information, and $\hat{\boldsymbol{x}} \in \mathbb{R}^{NC}$ refers to the reconstructed value.

B. Comparison With Existing Mechanisms

In this set of evaluations, we compare the performance of TIME with the existing training-based indoor white space exploration mechanisms (WISER and FIWEX). We train WISER and FIWEX based on 4 data sets, and evaluate their performance using the remaining 10. We also use the same 10 data sets to evaluate TIME.

We first run the 3 mechanisms with 3 different numbers of spectrum detectors deployed: 10, 30 and 50. Evaluation results on FA Rate and WS Loss Rate are shown in Fig. 6(a)-(c). For the FA Rate, FIWEX performs the best with 2 different



(a) 10 spectrum detectors (b) 30 spectrum detectors (c) 50 spectrum detectors (d) FA Rate

(e) 10 spectrum detectors (f) 30 spectrum detectors (g) 50 spectrum detectors (h) WS Loss Rate

Fig. 6. Comparison results in terms of FA Rate and WS Loss Rate.

number of spectrum detectors (10 and 30), while WISER gets the largest FA Rates. When there are 50 spectrum detectors, FIWEX gets the largest FA Rate, while WISER gets the smallest. TIME has the smallest WS Loss Rate when there are 10 spectrum detectors, while FIWEX performs the best in terms of WS Loss Rate, when the number of spectrum detectors are 30 and 50.

When different white space exploration mechanisms are used, we may have different distributions of WS Loss Rate over the indoor locations. Fig. 6(e)-(g) illustrate the Cumulative Distribution Function (CDF) curves of WS Loss Rates at the 66 locations corresponding to the 3 different number of spectrum detectors deployed. We can observe that the performance of TIME is competitive to that of WISER and FIWEX generally, and is even superior to the other two when there are 10 spectrum detectors available. As shown in Fig. 6(e), there are 83.6% locations in TIME having WS Loss Rates less than 40%. The numbers are 71.6% and 65.7% for FIWEX and WISER, respectively. Furthermore, the number of locations with WS Loss Rate 0 increases as the increment of deployed spectrum detectors: 10, 30, 50.

Fig. 6(d) shows the FA Rates when we vary the number of spectrum detectors from 3 to 66. When the number of spectrum detectors is small (< 7), the FA Rates of TIME are higher than the others. This is because we randomly choose the beginning 2 spectrum detectors. When the number of spectrum detectors is small, there exists randomness among the selected spectrum detectors deployment locations, and thus the best locations may not be chosen. As the number of spectrum detectors increases, the influence of the incremental spectrum detector deployment method becomes more significant, and makes the FA Rate of TIME decreases quickly. Actually, the average FA rate (from 3 spectrum detectors to 66 spectrum detectors) of TIME is 2.21%, which is almost the same as WISER (2.18%), and only a little bit higher than FIWEX (1.49%).

We show the WS loss Rates in Fig. 6(h). TIME's WS Loss Rates are almost the same as FIWEX. Both of TIME's and FIWEX's WS Loss Rates are lower than those of WISER. The average WS Loss Rate is 19.9% for WISER, 14.2% for FIWEX, and 15.6% for TIME.

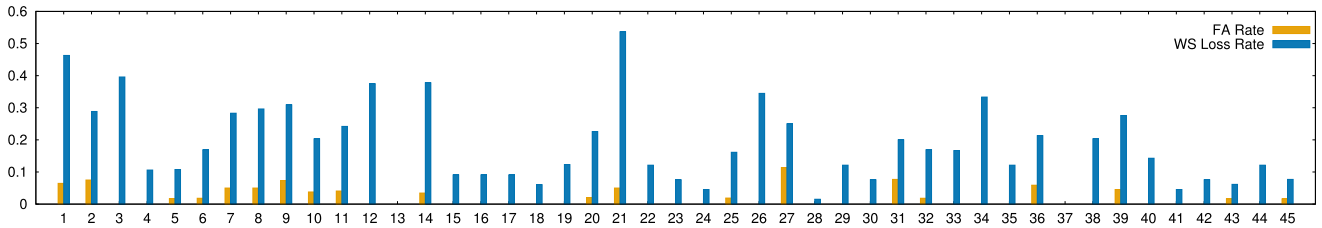
The above results show that TIME achieves competitive performance to the two representative existing training-based mechanisms. This is because that although TIME does not know the detailed correlations among white spaces at different indoor locations, it utilizes the general redundancy that exists in the indoor white spaces.

Yet, another observation from Fig. 6 is that the FA Rates of TIME (2.21% on average) are much smaller than its WS Loss Rates (15.6% on average). This is caused by the protection range PR. In this way, the licensed signal transmissions can be protected.

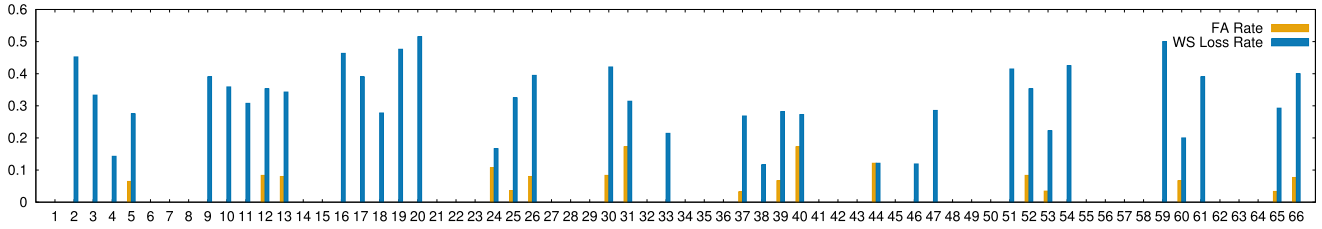
C. Differences Among Channels and Locations

In this set of evaluations, we evaluate the performance of TIME on different channels and at different locations. We use TIME to deploy 30 spectrum detectors, and calculate its FA Rates and WS Loss Rates on different channels at various locations.

Fig. 7(a) illustrates the FA Rates and WS Loss Rates on different channels. We observe that the FA Rates and WS Loss Rates differ on different channels. For example, the FA Rates and WS Loss Rates of channel 13 and 37 are all 0. This is because channel 13 and 37 are always occupied at all the locations, and thus their availabilities can be correctly predicted. Channel 1 and 21 suffer relative large FA Rates and WS Loss Rates, since the signal strengths on them are near to the white space threshold, making their availabilities differs at different locations. Hence, it is more difficult to perform accurate prediction on them. Another important observation is that more than half of the channels have an FA Rate of 0,

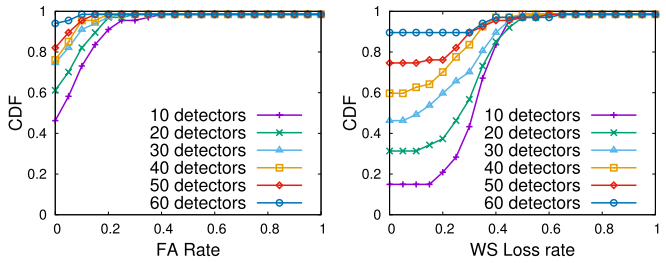


(a) FA Rate and WS Loss Rate on different channels



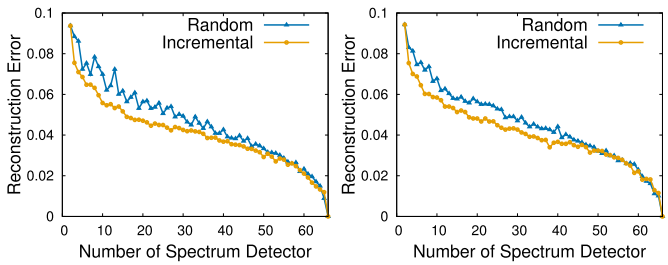
(b) FA Rate and WS Loss Rate at different locations

Fig. 7. Performance of different channels and locations when there are 30 spectrum detectors.



(a) FA Rate (b) WS Loss Rate

Fig. 8. CDF curves at different locations.



(a) First day (b) Average of the other days

Fig. 9. Incremental deployment vs random deployment.

which is caused by the protection range we set (equation (23)) to protect the licensed users.

Fig. 7(b) shows the FA Rates and WS Loss Rates at different locations. Similar to Fig. 7(a), the FA Rates and WS Loss Rates differ at different locations, since the correlations among different locations are different. For example, the FA Rate and WS Loss Rate are relatively small at location 38. This may be because that location 38 is tightly correlated to the locations with spectrum detectors deployed. On the contrary, the correlations between location 40 and the locations with spectrum detectors may be relatively weak, which lead to the high FA Rate and WS Loss Rate. Specially, the FA rates and WS loss rates at locations with spectrum detectors deployed are always 0.

We also evaluate the FA Rates and WS Loss Rates at different locations with different numbers of deployed spectrum detectors. We vary the number of spectrum detectors from 10 to 60 with a step of 10, and calculate the CDF curves of FA Rates and WS Loss Rates. As shown in Fig. 8, the CDF curves are "higher", when more spectrum detectors are deployed. This means that the number of locations with high FA Rate or WS Loss Rate decreases as increment of the number of deployed spectrum detectors.

D. Performance of Spectrum Detector Deployment

In this set of evaluations, we study the performance of the incremental spectrum detector deployment method. We deploy spectrum detectors based on the first day's measurement, and compare TIME's performance with a random spectrum detector deployment using the first day's data. Fig. 9(a) shows the reconstruction errors of our incremental deployment and the random deployment, when number of spectrum detectors is from 2 to 66. The initial 2 spectrum detectors of TIME and the random deployment method are the same, which leads to the same reconstruction error. As the number of spectrum detectors increases, the reconstruction error of our incremental deployment method decreases quicker than that of the random way. The average reconstruction error (from 2 spectrum detectors to 66 spectrum detectors) of our incremental deployment method is 0.0398, which is 12.1% lower than that of the random deployment.

Then, we evaluate the feasibility of our spectrum detector deployment method. Since the signal strengths of indoor white spaces vary over time, a feasible deployment method should guarantee that if we deploy spectrum detectors according to the current sensing result, the deployed detectors should also lead to a good system performance in the future. In Section III, we have shown the stability of correlations among indoor white spaces, which means that the locations of coefficients in $\boldsymbol{\omega}$ with large absolute amplitudes are stable over time. If we deploy spectrum detectors at candidate locations which collect the information of "large" coefficients in $\boldsymbol{\omega}$, the deployed spectrum

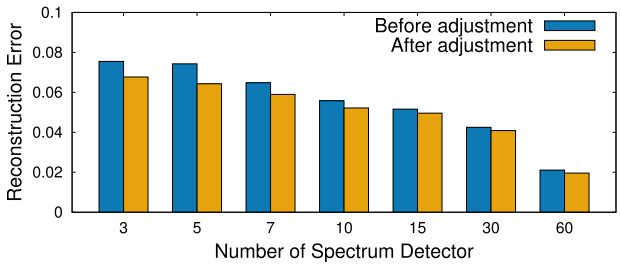
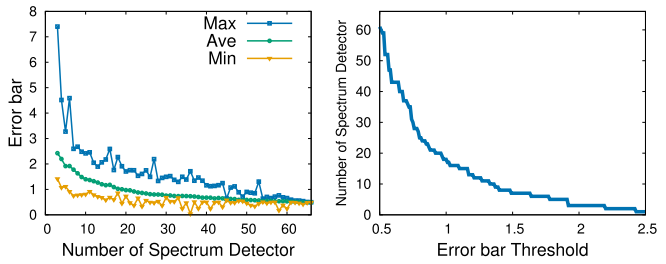


Fig. 10. Spectrum detector location adjustment.



(a) Error bar (b) Number of spectrum detectors

Fig. 11. Proper number of spectrum detectors.

detectors would also yield "large" coefficients at different times. Hence, we believe that our incremental spectrum detector deployment method is feasible. Here, we deploy spectrum detectors according to the first day's data set, and calculate the reconstruction errors under other data sets, and then draw their average in Fig. 9(b). The average reconstruction error (from 2 spectrum detectors to 66 spectrum detectors) of our incremental deployment method is 0.04, while 0.0446 for the random deployment, which means that our incremental deployment get a relatively 10.3% lower reconstruction error.

We also study the effect of the spectrum detector location adjustment. We compare the reconstruction errors before and after adjustment with typical numbers of spectrum detectors. Fig. 10 illustrates this comparison when the numbers of spectrum detectors are 3, 5, 7, 10, 15, 30, and 60. When there are only a small number of spectrum detectors, the adjustment has an obvious impact on the reconstruction error. For example, when there are 5 spectrum detectors, the reconstruction error is 0.0743 before adjustment, and drops to 0.0643 after adjustment. Similarly, when there are 7 spectrum detectors, the reconstruction errors are 0.0648 and 0.0590, respectively. The impact of the spectrum detector location adjustment algorithm decreases as the number of spectrum detectors increases, since the incremental deployment method performs better when there are a larger number of spectrum detectors. For example, when there are 60 spectrum detectors, the reconstruction error is 0.0211 before adjustment and 0.0196 after that.

After that, we evaluate the algorithm determining the proper number of spectrum detectors. Fig. 11(a) shows the error bars with different numbers of spectrum detectors. We show the maximum value, minimum value, and mean of $errbar$ (equation (31)). We observe that the mean of error bar decreases while the number of spectrum detectors increases. Fig. 11(b) shows the error bar threshold and the corresponding number of spectrum detectors. The number of spectrum detectors needed decreases as the the error bar threshold increases. We can determine a proper number of spectrum detectors based on the average error that can be tolerated.

TABLE I
COMPARISON OF THE INDOOR AND OUTDOOR MEASUREMENT

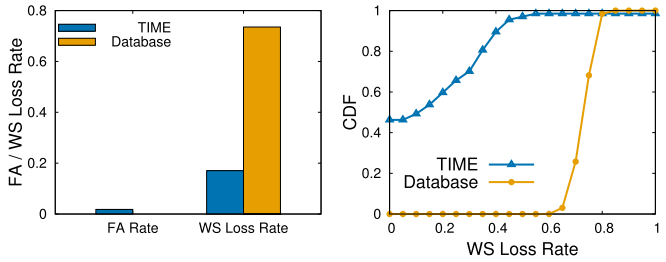
	Outdoor	Indoor	Indoor bonus
Number of vacant channels	20.57	27.62	7.05
White space rate (%)	45.71	61.38	15.67
Vacant Spectrum (MHz)	164.57	220.97	56.4

E. Combining With Geo-Location Database

In this set of evaluations, we first study the performance of the geo-location database approach in exploring indoor white spaces. Then, we show that the geo-location database could help TIME to improvement its performance.

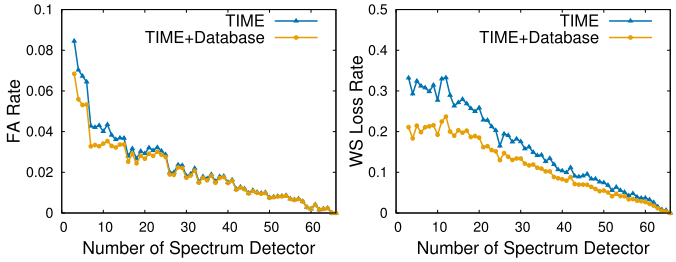
As we have mentioned before, the geo-location database estimates the signal strengths of TV channels at different outdoor locations based on the signal propagation model. However, the indoor obstacles (e.g., walls) are not considered by the geo-location database when calculating the signal strengths. This means that directly applying the geo-location database to explore indoor white spaces would lead to the overly conservative results. In order to study the performance of the geo-location database, we first compare the results of geo-location database with the indoor measurement results, and then compare the geo-location database approach with TIME.

Since there does not exist a publicly accessible geo-database in the country we perform the indoor white space measurement, we alternatively use the outdoor measurement results as a substitution of the geo-location database. Actually, we have measured the outdoor TV channels at the same time with the indoor white space measurement as shown in Section III. Apart from the 66 indoor locations, we also perform the measurement on the rooftop of our measured building, and get 14 vectors containing the corresponding signal strengths of the 45 TV channels. TABLE I shows the comparison of the indoor and outdoor measurement results. We calculate the number of vacant TV channels in the 14 data sets, and then average them. Specially, for the indoor measurement results, we also take the average of the 66 indoor locations. In the indoor environment, there are 7.05 more vacant channels than the outdoor. The 7.05 extra vacant channels are about 15.67% of all the 45 channels, and lead to the 56.4 MHz extra spectrum. The above results tell us that there are more white spaces in the indoor environment than the outdoor. If we directly apply the outdoor results (i.e., geo-location database results) to the indoor environment, we may lose a lot of white spaces. We also compare the geo-location database with TIME. We deploy 30 spectrum detectors for TIME, and compare its reconstruction results at the 66 indoor locations with the geo-location database approach, which considers that the white space availabilities at the 66 indoor locations are the same as the outdoor results. The results are illustrated in Fig. 12. Fig. 12(a) shows the FA Rates and WS Loss Rates. We observe that the FA Rate of the geo-location database approach is 0 whereas its WS Loss Rate (73.5%) is much



(a) 30 spectrum detectors (b) CDF curves

Fig. 12. Comparison between TIME and geo-location database.



(a) FA Rate (b) WS Loss Rate

Fig. 13. Combination of TIME and the geo-location database.

higher that TIME (17.0%). This means that the geo-locations database loses a lot of indoor white spaces. Fig. 12(b) illustrates the corresponding CDF curves about the WS Loss Rates. The results are consistent with Fig. 12(a).

Although directly applying the geo-location database approach into the indoor scenario would lead to a very conservative result, it can help us to improve the performance of TIME. Intuitively, if a channel is vacant in the outdoor scenario, it must be vacant in the indoor scenario. Hence, we can combine TIME with the geo-location database in the following way: according to the geo-location database or the outdoor measurement results, TIME considers the outdoor vacant channels always vacant at different indoor locations, and only focus on the following channels. We compare the FA Rates and WS Loss Rate of TIME before and after considering the results of the geo-location database (TIME+Database). As shown in Fig. 13, both of the FA Rate and WS Loss of TIME decrease after using the geo-location database results. On average, TIME gets a relatively 10.8% smaller FA Rate and a relatively 27.7% smaller WS Loss Rate after considering the results of the geo-location database.

VI. Discussion

In this section, we discuss limitations of the training-free indoor white space exploration mechanism, TIME, and introduce the corresponding future works. We first discuss the indoor white space exploration in the relatively small indoor environment, and then talk about the candidate locations. After that, we introduce the necessity of a dynamic indoor white space exploration system. At last, we discuss the quality of the white spaces.

A. White Space Exploration of Small Indoor Environment

Our training-free indoor white space exploration mechanism, TIME, mainly focuses on the large indoor environments, such as shopping malls and office buildings. This is because the training process in such a large indoor environment requires considerable time, devices, as well as human power. Hence, a training-free indoor white space exploration mechanism is highly attractive, even if its performance may be not as good as the training-based mechanism.

However, in a relatively small indoor environment, such as a single home scenario, we may prefer to indoor white space exploration with high accuracy, since it is much easier to collect the training data. Hence, in a relative small indoor environment, a training-based indoor white space exploration mechanism is more appropriate.

B. Spectrum Detectors and Candidate Locations

The RF spectrum detectors used to identify white spaces are usually expensive. For example, the measurement device we used in Section III, which consists of a USRP N210 with SBX daughter-board and a log periodic PCB antenna, costs around 2K US dollars. Moreover, if we utilize the devices with a higher accuracy (e.g., ThinkRF WSA5000 [26]), the costs would be much higher and is out of our budget. Hence, reducing the total device cost should be an important design consideration of TIME.

Just like prior works [9], [10], the spectrum detectors can only be deployed at the candidate locations. Hence, the positions and number of the candidate locations would affect the number of needed spectrum detectors. Since all prior works as well as TIME assumed that the candidate locations are given beforehand, it will be an interesting future work to study how to determine the positions and number of the candidate locations.

C. Dynamic Indoor White Space Exploration

As shown in [9], the correlations among TV channels of the licensed user are stable in a long period of time (e.g., years). Hence, we could find a set of candidate locations based on the current measurement results, and deploy static spectrum detectors at them. The stability of the correlations makes the static deployment of the spectrum detectors a feasible approach. However, in a near future, when secondary users can hopefully coexist in the TV bands, the spectrum will have more temporal variations indicating the need of a dynamic mechanism.

In a dynamic indoor white space exploration system, the locations of the spectrum detectors must be dynamically adjusted according to the current measurement results. The location adjustment is difficult in a dynamic system, because the location correlation and channel correlation are not stable. Besides, we cannot get the complete correlations, since we can only measure a part of the indoor environment in each time slot. We thus set the design of a dynamic indoor white space exploration system as another future work.

D. Quality of TV White Spaces

In practice, the secondary users care about not only whether a channel is vacant or not, but also the quality of the

white spaces (or TV channels). Intuitively, there should be little interference when communicating through a high quality vacant channel. However, due to the limitation of budget and run time cost, existing indoor white space exploration mechanisms (including TIME) only deploy spectrum detectors at a part of the indoor locations, and recover the white space information of the whole indoor environment using the data reconstruction techniques. This means that there exist errors in the reconstruction results. For instance, it is possible that an occupied channel is incorrectly identified as vacant. The unlicensed signal transmissions in such mis-identified vacant channels would be interfered with the licensed TV signal transmissions.

Different from the prior works, TIME not only provides the values of the reconstruction but also provides their uncertainties, which can be described by the variance of the reconstruction results. Actually, the variance is equal to the error bar we defined in equation (31). A channel with a smaller variance means that we are more confident about its status, which reduces the potential interference with the licensed users. Hence, the variance of the reconstruction results can be set as a criterion of the quality of the white spaces.

VII. Related Work

In this paper, we study the training-free indoor white space exploration based on the Bayesian compressive sensing technique. Indoor white space exploration and Bayesian compressive sensing are two attractive research topics, which have been widely studied in recent years.

A. Bayesian Compressive Sensing

The theory of compressive sensing has been widely studied [16], [17], [30]–[32] and utilized in different fields, such as network traffic estimation [12], time-varying signals estimation [14], localization in mobile networks [33], soil moisture sensing [15], and data gathering [13], [34]. The sampling matrices in traditional compressive sensing theory [16], [17] are certain random matrices. Bayesian Compressive sensing [18] is a new type of data reconstruction technique, which presents an incremental way to design the sampling matrix. Ji *et al.* [19] proposed the multitask compressive sensing with an empirical Bayesian procedure for the estimation of hyperparameter. Qi *et al.* [20] considered the Dirichlet Process Priors in the multitask compressive sensing. As an efficient data reconstruction technique, Bayesian compressive sensing has been widely utilized in many realms, such as image representation [35], wideband spectrum sensing [36], encoding and decoding acceleration [37], and so on.

B. White Spaces

In recent years, TV white spaces have been receiving a lot of attentions from researchers. For example, Deb *et al.* [38] proposed a dynamic white space allocation system. Bahl *et al.* [39] designed a white space Wi-Fi like network. Radunovic *et al.* [40] designed a mechanism enabling low-power nodes to coexist with high-power nodes in white space networks. Radunovic *et al.* [41] studied the dynamic channel, rate selection and scheduling for white spaces. Vaze and Murthy [42] proposed a white space identification method using random sensors. Ding *et al.* [43] studied the cellular-base-station assisted device-to-device communications in TV white space. Zhang *et al.* [8] proposed a vehicle-based enhancing for white space spectrum database.

Most of the prior works focus on the outdoor white spaces. Spectrum sensing is an important approach to explore white spaces. For example, Ding *et al.* [44] studied the robust spectrum sensing with crowd sensors. Zou *et al.* [45] analyzed the impact of spectrum sensing overhead. Cooperative spectrum sensing has been shown to be a more effective way to improve the detection performance, and has been widely studied in past years [46]–[51]. However, most white space devices and standards are designed based on the geo-location database approach [52]–[54] proposed by FCC, because it is energy efficient and easy to implement.

Recently, indoor white space receives more and more attentions, because there are more indoor white spaces than outdoor and the geo-location database exploration approach cannot be directly applied to the indoor scenarios. In 2013, Ying *et al.* [9] proposed the first indoor white space exploration system, WISER, using a correlation based clustering algorithm. Then, Zhang *et al.* [55] designed the first framework for indoor multi-AP white space network, namely WINET. After that, Liu *et al.* [10] proposed a cost-efficient indoor white space exploration mechanism FIWEX by carefully studying the dependence between channels and locations. However, both of them need a training process, where indoor white space information should be collected before deploying spectrum detectors. In this paper, we propose the first training-free indoor white space exploration mechanism, namely TIME, based on the Bayesian compressive sensing technique. Our training-free approach, TIME, has competitive performances against the existing training-based indoor white space exploration mechanisms.

VIII. Conclusion

In this paper, we have performed indoor white space measurement in a complex of offices and labs. The measurement results demonstrate the correlations of indoor white spaces, as well as their stability over time. Based on the measurement observations, we propose a Training-free Indoor white space exploration MEchanism, namely TIME. TIME mainly contains a Bayesian compressive sensing based data reconstruction part and a method of incremental spectrum detectors deployment, which achieves an average of 10.3% improvement compared with the random deployment. Compared with the state-of-the-art training-based indoor white space exploration mechanisms, TIME can achieve competitive performances without training.

Acknowledgment

The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies or the government.

REFERENCES

[1] E. Hossain, D. Niyato, and Z. Han, *Dynamic Spectrum Access and Management in Cognitive Radio Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[2] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, 2006.

[4] J. Li, Q. Yang, L. Hanzo, and K. S. Kwak, "Over-booking approach for dynamic spectrum management," in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1–5.

[5] Y. Liao, T. Wang, K. Bian, L. Song, and Z. Han, "Decentralized dynamic spectrum access in full-duplex cognitive radio networks," in *Proc. IEEE ICC*, Jun. 2015, pp. 7552–7557.

[6] R. Balamurthi, H. Joshi, C. Nguyen, A. K. Sadek, S. J. Shellhammer, and C. Shen, "A TV white space spectrum sensing prototype," in *Proc. IEEE DySPAN*, May 2011, pp. 297–307.

[7] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl, "SenseLess: A database-driven white spaces network," *IEEE Trans. Mobile Comput.*, vol. 11, no. 2, pp. 189–203, Feb. 2012.

[8] T. Zhang, N. Leng, and S. Banerjee, "A vehicle-based measurement framework for enhancing whitespace spectrum databases," in *Proc. ACM MobiCom*, 2014, pp. 17–28.

[9] X. Ying, J. Zhang, L. Yan, G. Zhang, M. Chen, and R. Chandra, "Exploring indoor white spaces in metropolises," in *Proc. ACM MobiCom*, 2013, pp. 255–266.

[10] D. Liu, Z. Wu, F. Wu, Y. Zhang, and G. Chen, "FIWEX: Compressive sensing based cost-efficient indoor white space exploration," in *Proc. ACM MobiHoc*, 2015, pp. 17–26.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.

[12] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and Internet traffic matrices," in *Proc. ACM SIGCOMM*, 2009, pp. 267–278.

[13] X. Wu, P. Yang, T. Jung, Y. Xiong, and X. Zheng, "Compressive sensing meets unreliable link: Sparsest random scheduling for compressive data gathering in lossy WSNs," in *Proc. ACM MobiHoc*, 2014, pp. 13–22.

[14] D. Angelosante, G. B. Giannakis, and E. Grossi, "Compressed sensing of time-varying signals," in *Proc. IEEE DSP*, Jul. 2009, pp. 816–823.

[15] X. Wu and M. Liu, "*In-situ* soil moisture sensing: Measurement scheduling and estimation using compressive sensing," in *Proc. ACM IPSN*, 2012, pp. 1–12.

[16] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[17] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[18] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[19] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 92–106, Jan. 2009.

[20] Y. Qi, D. Liu, D. Dunson, and L. Carin, "Multi-task compressive sensing with Dirichlet process priors," in *Proc. ACM ICML*, 2008, pp. 768–775.

[21] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[22] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tut.*, vol. 11, no. 1, pp. 116–130, 1st Quart., 2009.

[23] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, Apr. 1967.

[24] A. Bagheri, P. C. Sofotasios, T. A. Tsiftsis, A. Shahzadi, and M. Valkama, "AUC study of energy detection based spectrum sensing over η-μ and α-μ fading channels," in *Proc. IEEE ICC*, Jun. 2015, pp. 1410–1415.

[25] FCC. *Second Memorandum Option and Order*, accessed date Sep. 2015. [Online]. Available: https://apps.fcc.gov/edocs_public/attachmatch/FCC-10-174A1.pdf

[26] *ThinkRF*, accessed on Jul. 2016. [Online]. Available: http://www.thinkrf.com/

[27] M. E. Tipping and A. Faul, "Fast marginal likelihood maximization for sparse Bayesian models," in *Proc. AISTSTS*, Jan. 2003, pp. 1–8.

[28] D. J. C. MacKay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, 1992.

[29] S. Ji, B. Krishnapuram, and L. Carin, "Variational Bayes for continuous hidden Markov models and its application to active learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 522–532, Apr. 2006.

[30] M. A. Khajehnejad, J. Yoo, A. Anandkumar, and B. Hassibi, "Summary based structures with improved sublinear recovery for compressed sensing," in *Proc. IEEE ISIT*, Jul./Aug. 2011, pp. 1427–1431.

[31] M. Cheraghchi, A. Hormati, A. Karbasi, and M. Vetterli, "Compressed sensing with probabilistic measurements: A group testing solution," in *Proc. IEEE Allerton*, Sep./Oct. 2009, pp. 30–35.

[32] A. Karbasi, A. Hormati, S. Mohajer, and M. Vetterli, "Support recovery in compressed sensing: An estimation theoretic approach," in *Proc. IEEE ISIT*, Jun./Jul. 2009, pp. 679–683.

[33] S. Rallapalli, L. Qiu, Y. Zhang, and Y.-C. Chen, "Exploiting temporal stability and low-rank structure for localization in mobile networks," in *Proc. ACM MobiCom*, 2010, pp. 161–172.

[34] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. ACM MobiCom*, 2009, pp. 145–156.

[35] M. Zhou *et al.*, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.

[36] Z. Zhang, H. Li, D. Yang, and C. Pei, "Space-time Bayesian compressed spectrum sensing for wideband cognitive radio networks," in *Proc. IEEE DySPAN*, Apr. 2010, pp. 1–11.

[37] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[38] S. Deb, V. Srinivasan, and R. Maheshwari, "Dynamic spectrum access in DTV whitespaces: Design rules, architecture and algorithms," in *Proc. ACM MobiCom*, 2009, pp. 1–12.

[39] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh, "White space networking with Wi-Fi like connectivity," in *Proc. ACM SIGCOMM*, 2009, pp. 27–38.

[40] B. Radunović, R. Chandra, and D. Gunawardena, "Weeble: Enabling low-power nodes to coexist with high-power nodes in white space networks," in *Proc. ACM CoNEXT*, 2012, pp. 205–216.

[41] B. Radunović, A. Proutiere, D. Gunawardena, and P. Key, "Dynamic channel, rate selection and scheduling for white spaces," in *Proc. ACM CoNEXT*, 2011, Art. no. 2.

[42] R. Vaze and C. R. Murthy, "On whitespace identification using randomly deployed sensors," in *Proc. IEEE COMSNETS*, Jan. 2014, pp. 1–7.

[43] G. Ding, J. Wang, Q. Wu, Y.-D. Yao, F. Song, and T. A. Tsiftsis, "Cellular-base-station-assisted device-to-device communications in TV white space," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 107–121, Jan. 2016.

[44] G. Ding *et al.*, "Robust spectrum sensing with crowd sensors," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3129–3143, Sep. 2014.

[45] Y. Zou, Y.-D. Yao, and B. Zheng, "Outage probability analysis of cognitive transmissions: Impact of spectrum sensing overhead," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2676–2688, Aug. 2010.

[46] G. Ding, Q. Wu, Y.-D. Yao, J. Wang, and Y. Chen, "Kernel-based learning for statistical signal processing in cognitive radio networks: Theoretical foundations, example applications, and future directions," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 126–136, Jul. 2013.

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

[48] Y. Liao, K. Bian, L. Ma, and L. Song, "Robust cooperative spectrum sensing in full-duplex cognitive radio networks," in *Proc. IEEE ICUFN*, Jul. 2015, pp. 66–68.

[49] B. Selim, O. Alhussein, G. K. Karagiannidis, and S. Muhaidat, "Optimal cooperative spectrum sensing over composite fading channels," in *Proc. IEEE ICCW*, Jun. 2015, pp. 520–525.

[50] Y. Zou, Y.-D. Yao, and B. Zheng, "A selective-relay based cooperative spectrum sensing scheme without dedicated reporting channels in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1188–1198, Apr. 2011.

[51] L. Mohjazi *et al.*, "Unified analysis of cooperative spectrum sensing over generalized multipath fading channels," in *Proc. IEEE PIMRC*, Aug./Sep. 2015, pp. 370–375.

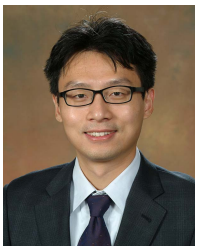
[52] M. Gibson, "TV white space geolocation database," in *Proc. IEEE 802 Plenary Meeting Workshop*, Jul. 2010, pp. 1–14.

[53] 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 *Proc. IEEE DySPAN*, Oct. 2008, pp. 1–9.
[54] X. Feng, J. Zhang, and Q. Zhang, "Database-assisted multi-AP network on TV white spaces: Architecture, spectrum allocation and AP discovery," in *Proc. IEEE DySPAN*, May 2011, pp. 265–276.
[55] J. Zhang, W. Zhang, M. Chen, and Z. Wang, "WINET: Indoor white space network design," in *Proc. IEEE INFOCOM*, Apr./May 2015, pp. 630–638.

Dongxin Liu received the B.S. degree in computer science from Shanghai Jiao Tong University, China, in 2014, where he is currently pursuing the degree with the Department of Computer Science and Engineering. His research interests encompass indoor white space exploration, compressive sensing and mobile wireless networks. He is a Student Member of the ACM and CCF.

Fan Wu (S'06–M'10) received the B.S. degree in computer science from Nanjing University in 2004, and the Ph.D. degree in computer science and engineering from the State University of New York at Buffalo in 2009. He is currently an Associate Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. He has visited the University of Illinois at Urbana–Champaign as a Post-Doctoral Research Associate. His research interests include wireless networking and mobile computing, algorithmic game theory and its applications, and privacy preservation. He has authored over 100 peer-reviewed papers in technical journals and conference proceedings. He is a recipient of the first class prize for Natural Science Award of China Ministry of Education, NSFC Excellent Young Scholars Program, the ACM China Rising Star Award, the CCF-Tencent Rhinoceros bird Outstanding Award, the CCF-Intel Young Faculty Researcher Program Award, and the Pujiang Scholar. He has served as the Chair of the CCF YOCSEF Shanghai, and serves on the Editorial Board of the Elsevier *Computer Communications*, and a member of Technical Program Committee of over 60 academic conferences.

Linghe Kong (S'09–M'13) received the bachelor's degree in automation with Xidian University in 2005, the master's degree in telecommunication with TELECOM SudParis in 2007, and the Ph.D. degree in computer science with Shanghai Jiao Tong University in 2012. He was a Post-Doctoral Fellow with Columbia University and McGill University. He is currently an Associate Professor with Shanghai Jiao Tong University. His research interests include wireless communications, sensor networks and mobile computing.



Shaojie Tang (S'09–M'14) received the Ph.D. degree in computer science from the Illinois Institute of Technology in 2012. He is currently an Assistant Professor with the Naveen Jindal School of Management, University of Texas at Dallas. His research interest includes social networks, mobile commerce, game theory, e-business and optimization. He received the best paper awards in the ACM MobiHoc 2014 and the IEEE MASS 2013. He also received the ACM SIGMobile service award in 2014. He served in various positions (as a Chair and the TPC member) at numerous conferences, including the ACM MobiHoc and the IEEE ICNP. He is an Editor of the *Elsevier Information Processing in Agriculture* and the *International Journal of Distributed Sensor Networks*.

Yuan Luo (M'06) received the Ph.D. degree in probability and mathematical statistics from Nankai University, China, in 1999. From 1999 to 2001, he held a post-doctoral position with the Institute of Systems Science, Chinese Academy of Sciences, China. From 2001 to 2003, he held a post-doctoral position with the Institute for Experimental Mathematics, University of Duisburg-Essen, Germany. He is currently a Full Professor of Computer Science with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. His research interests include information theory, coding theory, and computer security.

Guihai Chen (M'06–SM'13) received the B.S. degree in computer software from Nanjing University in 1984, the M.E. degree in computer applications from Southeast University in 1987, and the Ph.D. degree in computer science from the University of Hong Kong in 1997. He is currently a Distinguished Professor with Shanghai Jiao Tong University. He had been invited as a Visiting Professor from the Kyushu Institute of Technology, Japan, the University of Queensland, Australia, and Wayne State University, USA. He has a wide range of research interests with focus on parallel computing, wireless networks, data centers, peer-to-peer computing, high-performance computer architecture and data engineering. He has authored over 350 peer-reviewed papers, and over 200 of them are in well-archived international journals, such as the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON COMPUTERS, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE/ACM TRANSACTIONS ON NETWORKING and the *ACM Transactions on Sensor Networks*, and also in well-known conference proceedings, such as HPCA, MOBIHOC, INFOCOM, ICNP, ICDCS, CoNext, and AAAI. He has won several best paper awards including ICNP 2015 best paper award. His papers have been cited over 10 000 times according to Google Scholar.