Detection of Primary User Emulation Attacks Based on Compressive Sensing in Cognitive Radio Networks

Manman Dang*, Zhifeng Zhao*, and Honggang Zhang*†
*York-Zhejiang Lab for Cognitive Radio and Green Communications
Zhejiang University, Zheda Road 38, Hangzhou 310027, China
Email: {mmdang, zhaozf}@zju.edu.cn
†Université Européenne de Bretagne (UEB) & Supélec, CS 47601, Cesson-Sévigné Cedex, France
Email: honggang.zhang@supelec.fr

Abstract—In this paper, we introduce the idea of compressive sensing (CS) into primary user emulation attack (PUEA) detection in cognitive radio networks. We can distinguish whether the signal transmitters are primary users or PUEAs by obtaining the locations of transmitters through processing the received signal strength (RSS) readings. Since the RSS has redundancy in spatial domain, we employ CS theory to save the number of measurement sensors and messages need to be collected. Additionally, the number of measurements is dominated by the number of PUEA being sensed, the conventional algorithms with fixed measurement number have poor performances in physical situations where number of PUEA is unknown and changing. So, we propose an adaptive orthogonal matching pursuit algorithm (AOMP) to adapt to the changing cases of PUEA. Simulation results show that the location reconstruction of changing PUEA by AOMP algorithm outperforms traditional OMP algorithm with better accuracy. As a result, the channel utilization can be greatly improved.

Index Terms—cognitive radio users, primary user emulation attacks, compressive sensing, adaptive orthogonal matching pursuit

I. INTRODUCTION

Cognitive radio is introduced as an effective method to solve the ever-growing spectral resource demands of the telecommunications world by allowing secondary users (SUs) accessing to the under utilized licensed spectrum which belongs to the primary users (PUs) at the moment [1][2]. Two cornerstones of cognitive radio are spectrum sensing and dynamic spectrum access. Spectrum sensing aims at finding and utilizing the vacant spectrum at present moment while avoiding interference to the primary network at the same time. After obtaining the usage conditions of the spectral resources, cognitive radio users choose to dynamically access to the vacant licensed spectrum for secondary transmission or implement spectrum mobility to vacate the needed licensed spectrum [3]. So, detection results must be accurate enough to guarantee cognitive radio networks make the most use of available resources as well as cause interferences to the primary network as little as possible.

The methods for detecting primary users turn out to be inaccurate when cognitive radio operates in hostile environment. In particular, an adversary’s CR (cognitive radio equipment) transmits signals whose characteristics (e.g., signal energy, signal features and transmitter locations) emulate those of incumbent signals, and these are defined as primary user emulation attacks. PUEA will mislead cognitive radio users mistake the adversary CRs as primary users. As a result, this will lead to wastage of spectrum resources and severe interference to the spectrum management of cognitive radio networks.

There are several methods proposed to detect the PUEA. Existing technologies can be generally classified as: location aware and location unaware [4]-[7]. Location aware methods such as feature detection, energy detection assume that the location of PUEA is known. Feature detection is based on the assumption that PUEA can’t emulate the certain feature of the primary user’s signals [4], while energy sensing is based on the difference between the transmit energy of PUEA signals and primary user’s signals [5][6]. The localization of the wireless location techniques relies on the measurements of certain distance-dependent parameters at BS or processing center. These distance-dependent parameters can be further sorted as: Received Signal Strength (RSS), Angle of Arrival (AOA), and Time of Arrival (TOA) [4][7]. Among these methods, the using of AOA and TOA are restricted because of the request of directional antennas and accurate timing clock at each cognitive radio user respectively. Though impact of the attacks can be affected by the hardware constraints of attackers, it is assumed that they can emulate signal characteristics and transmit powers of primary users in most cases. So, location based techniques are more robust. A transmitter verification scheme based on RSS was proposed in [4] which includes three steps: verification of signal characteristics, measurement of received signal’s energy level, and the localization of the signal source. However, the underlying sensor network requires hundreds and thousands sensors to be deployed which brings in heavy overhead to the network.

In this paper, we introduce compressive sensing into the localization of PUEA to reduce the number of measurement sensors needed, through which the cost of configuration of
sensor network and the computation involved at the fusion center can be greatly saved. There have been many researches on localization based on compressed sensing, among which the localization of transmitters is what we considered [8]-[11]. Primary user detection is performed by multi-resolution Bayesian compressed sensing in [8], and multiple targets localization in wireless networks is discussed in details in [9] [10]. WLAN access points positioning using RSS based compressive sensing is discussed in [11]. However, both of these methods proposed take no consideration of the changing of number of detection targets, which is the physical truth in most cases, especially for PUEA detection. In this paper, by using CS theory for localization of signal sources we propose an adaptive orthogonal matching pursuit algorithm (AOMP), which has a better performance for detecting changing and unknown sources.

The rest of this paper is organized as follows. In Section II, the PUEA detection problem is described in details as well as compressed sensing based localization method. We then introduce the problem of PUEA location with unknown and changing transmitter number. Then we propose AOMP algorithm for increasing the localization accuracy and adapting to the changes of network, and guarantee the location error in each sensing period. The simulation results and analysis are given in section III. Section IV describes the simulation results and analysis. Conclusion and future work is presented at section V.

II. PUEA LOCALIZATION USING COMPRESSED SENSING
A. System Model and Problem Formulation
Since using of TV White Space is one of the most important applications of CR technology, we consider the scenario where CR network area is under the coverage of TV transmit tower, and CR users dynamically access to the vacant TV sub-channels. There are also some adversary CR users that emulate the characteristics of primary users which are called as PUEA in the CR network. There is an underlying sensor network to catch the signals’ snapshot of the CR network. Our aim is to locate the transmitters correctly with appropriate number of measurement sensors.

This is a multiple transmitters’ localization and counting problem, which can be described in Fig.1. $P$ transmitters transmit signals on $K$ TV sub-channels in an isotropic area, and we need to find the locations of the signal sources. CR users and measurement sensors which in the transmit range of transmitter $i$ can receive the signals of it. We divide this area into $N$ girds of the same size. Each of the $P$ transmitters locates at the center of one certain gird. Assume there are $M$ arbitrary measurement sensor nodes (SNs) to collect the RSS measurements from these transmitters and each of them is also located at the center of the gird. We are aiming at determining the locations of these transmitters simultaneously and correctly with only a small number of collected noisy RSS measurements. Note that with $P \ll N$ this localization problem has a sparse nature. Besides, as a result of the spatial correlation of RSS measurements, the measurements number can be much smaller than the number of girds, that is $M \ll N$.

As a novel approach to reconstruct signals accurately with far fewer measurements, compressive sensing is based on the signals’ sparse nature or compressibility under certain basis. In this paper the sparsity is reflected in the few locations of signal sources.

Consider the Rayleigh energy decay model in [10], the received signal energy at grid $j$ from a transmitter at grid $i$ is given by:

$$R_{i,j} = P_0 A_{i,j} / d_{i,j}^\alpha$$

where, $P_0$ is the power density at the transmitter, $A_{i,j}$ represents the Rayleigh fading of the transmit signals, and $\alpha \in [2, 5]$ is the path loss exponent determined by the environment.

Assume the location of transmitters over the gird is denoted by a matrix $S_{N \times K}$, which is

$$S = [s_1, ..., s_k, ..., s_K]$$

where each $s_k$ is a $N \times 1$ vector with all elements equal to zeros except $s_k(n) = 1$, where $n \in C_k$, $C_k$ is collection of the grid index of transmitter locations on sub-channel $k$. Usually, the total number of $C_k$ is a small integer. Since there are only $P$ transmitters, which means $S$ has only $P$ non-zero elements and $S$ is $P$-sparse.

To count and localize the transmitters, a conventional way is to place a sensor networks at the monitored area to collect the snapshots of RSS. For example, we can place one sensor at the center of each gird. Thus, the received RSS $X$ is a $N \times K$ matrix,

$$X = \Psi S$$

(1)

where, $\Psi$ is a $N \times N$ matrix represents transmitter energy decay which is defined by Eq. (2):

$$\Psi = P_0 \begin{bmatrix}
A_{11} & A_{12} & \ldots & A_{1N} \\
A_{21} & A_{22} & \ldots & A_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
A_{N1} & A_{N2} & \ldots & A_{NN}
\end{bmatrix}$$

(2)
Rather than collecting the measurements of received signals over all of the grid, compressive noisy RSS measurements $Y$ in an $M \times K$ dimensional space collected through the $M$ measurement sensors according to CS theory:

$$Y = \Phi X$$  \hspace{1cm} (3)

where $\Phi_{M \times N}$ is the measurement matrix defined as:

$$\Phi = [\Phi(1), \Phi(2), \ldots, \Phi(M)]^T$$

in above equation, $\Phi(i)$ is an $1 \times N$ vector represents the location of the $i$th $(1 \leq i \leq M)$ measurement sensor, with all elements equal to zero except $\Phi(i,j) = 1$, where $j(1 \leq j \leq N)$ is the grid index at which the $i$th measurement sensor is located.

The compressed sensing problem can be expressed by combing Eq. (1) to Eq. (3).

$$Y = \Phi \Psi S = \Theta S$$  \hspace{1cm} (4)

From Eq. (4), the noisy measurement value matrix $Y$ can be expressed as,

$$Y = \Theta S + N$$  \hspace{1cm} (5)

where $N$ is the additive Gaussian white noise matrix.

As stated in CS theory, reconstruction matrix $\Theta_{M \times N}$ obeys RIP rule is the sufficient condition for the successful recovery of a signal by CS. And it is provable that $\Theta$ here obeys RIP rules [10], this transmitters localization problem is solvable by CS.

The CS problem in Eq. (5) is a $l_1$ -minimization process, which is,

$$\min |S|_{l_1}, \text{ s.t. } \|\Theta S - Y\|_2 < \varepsilon$$  \hspace{1cm} (6)

There are many methods for solving Eq. (6), Such as: Basis Pursuit (BP) which is computationally intensive, Orthogonal Matching Pursuit (OMP) which provides fast solutions by iteratively selecting the optimal candidates. If there are prior information about $S$ available in some applications, Bayesian Compressed Sensing (BCS) can also be utilized [12]. Among these methods, OMP algorithm is preferable because of its fast convergence and good reconstruction performance.

In this paper, we first employ OMP algorithm to solve Eq. (5), we can locate accurately for a certain number of transmitters with number of measurements $M = O(P \cdot \log(N/P))$. However, the number of transmitters is unknown and changing with time, OMP with constant measurement number can't guarantee the accuracy for all the time. So we propose Adaptive Orthogonal Matching Pursuit (AOMP) algorithm to deal with the changing circumstances to be detected.

### B. Adaptive orthogonal matching pursuit algorithm (AOMP)

Since the recovery accuracy of OMP is dependent on the number of measurements $M$ and transmitters $P$ to be located. We have to find an optimal changing $M$ to ensure the accuracy along with the changing of $P$ at each time. By running OMP algorithm, we can get the recovery location matrix $Y$ and the averaged Mean Square Error of measurement value matrix $Y$ (YAMSE). YAMSE is defined as:

$$YAMSE = \frac{1}{M} \|Y - \Theta \hat{S}\|_2$$  \hspace{1cm} (7)

YAMSE represents the difference of real measurement values and recovery measurement values at each measurement sensors. We choose YAMSE as the stopping criteria when running OMP algorithm to search for appropriate $M$. The details of AOMP algorithm is described as follows:

#### Algorithm 1 AOMP

**Input:**
- A $M \times N$ reconstruction matrix $\Theta$.
- Initial number of measurements: $M = M_0$.
- A non-zero integer $L$ which represents the times of running to obtain a map of YAMSE versus measurement number.

**Output:**
- A $N \times K$ location matrix $\hat{S}$.

**Steps:**
1. Sensor nodes collect RSS measurements, and the fusion center generates the $M \times K$ measurement value matrix $Y$.
2. For the $i$th $(i \geq 1)$ sensing, if $i \leq L$, run step 3 with $M_i = M_0$. Meanwhile when $i = L$, compute the map of YAMSE versus the number of measurements $M$ based on the computation results before, and find the optimal $M_o$, with which the corresponding YAMSE nearest to the predefined stopping criteria $Er$. When $i > L$, run step 3 with $M_i = M_o$.
3. Run OMP algorithm with $M_i$to reconstruct the sparse location matrix $\hat{S}$. Calculate YAMSE and compare it with criteria $Er$. If YAMSE is bigger than $Er$, we then increase $M_i : M_i = M_i + 1$, else if YAMSE does not meet the predefined $Er$ then decrease $M_i : M_i = M_i - 1$.
4. Repeat step 3 until YAMSE meets the predefined stopping criteria $Er$.

### III. Simulation Results And Analysis

The scenario we considered here is shown in Fig.2, in which cognitive radio network area is 10 km away from TV transmitter tower. The TV transmitter has 10 MHz channels and there are $Q = 50$ CR users dynamically access to the vacant TV sub-channels. The duty cycle of each TV sub-channel is fixed at 0.2. There are also 8 adversary CR users active as PUEAs in the CR network. Each CR has a transmission range of 100 m, and each PUEA has an attack range of 250 m. We divide the CRN area into $50 \times 50$ grids, and the length of each grid is $l = 10 m$. The total grid is composed of the grids in the CRN area together with the one where primary user is located at. The total number of grid is:
The average available bandwidth (MHz) is defined as:

$$ABW = \frac{6 \times \sum_{i=1}^{Q} a_i}{10 \times Q} \times \alpha$$  \hspace{1cm} (8)$$

where, $a_i$ is the number of available sub-channels at the $i$th CR, and $\alpha$ is the duty cycle of TV sub-channels.

According to the coverage range of PUEAs, we can see that one or at most three PUEAs in each channel can lead to serious damage to the available bandwidth at each CR. In the simulation, we assume that there are at most 2 PUEAs in each TV sub-channel. Fig.3 shows that the averaged available bandwidth at each CR when there are different number of PUEA in CR network.

The accuracy of the localization in our paper is defined by normalized localization error (NLE):

$$NLE = \frac{\sum_{i=1}^{P} \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{P \times l}$$  \hspace{1cm} (9)$$

where $(x_i, y_i)$ is the real location of the $i$th transmitter, and $(\hat{x}_i, \hat{y}_i)$ is the corresponding recovered location. From Eq. (9) we know that NLE is the averaged location error of all transmitters and is normalized by cell length $l$.

The accuracy of the localization in our paper is defined by normalized localization error (NLE):

$$NLE = \frac{\sum_{i=1}^{P} \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{P \times l}$$  \hspace{1cm} (9)$$

where $(x_i, y_i)$ is the real location of the $i$th transmitter, and $(\hat{x}_i, \hat{y}_i)$ is the corresponding recovered location. From Eq. (9) we know that NLE is the averaged location error of all transmitters and is normalized by cell length $l$.

The detection environments. When the number of transmitters increases, AOMP can find the optimal measurement number $M$ adaptively, which is bigger than that in OMP algorithm, thus leading to better location reconstruction. Otherwise, when the number of transmitters decreases, AOMP algorithm pursues the optimal measurement number $M$ adaptively, which is smaller than that in OMP algorithm. Therefore, AOMP algorithm can also save the computation cost at the fusion center for a certain level.

The accuracy of the localization in our paper is defined by normalized location error (NLE):

$$NLE = \frac{\sum_{i=1}^{P} \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{P \times l}$$  \hspace{1cm} (9)$$

where $(x_i, y_i)$ is the real location of the $i$th transmitter, and $(\hat{x}_i, \hat{y}_i)$ is the corresponding recovered location. From Eq. (9) we know that NLE is the averaged location error of all transmitters and is normalized by cell length $l$.

The accuracy of the localization in our paper is defined by normalized location error (NLE):

$$NLE = \frac{\sum_{i=1}^{P} \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{P \times l}$$  \hspace{1cm} (9)$$

where $(x_i, y_i)$ is the real location of the $i$th transmitter, and $(\hat{x}_i, \hat{y}_i)$ is the corresponding recovered location. From Eq. (9) we know that NLE is the averaged location error of all transmitters and is normalized by cell length $l$.

Fig.5 shows the averaged $YAMSE$ (averaged MSE of measurement matrix $Y$) and NLE (normalized localization error) of PUEA versus measurement number $M$ under a case of 8 constant PUEAs. It can be seen from Fig.5 that $YAMSE$ decreases as the measurement number increases. Besides, NLE of PUEA has the same decreasing trend as $YAMSE$. So, $YAMSE$ can be used as a stopping criteria to tell if measurement number currently is large enough to locate correctly. This simulation results validate the rationality of our proposed AOMP algorithm consequently.

Fig.6 plots the effects of different stopping criteria under two constant cases of 10 PUEAs and 18 PUEAs with stopping criteria changing from 0.1 to 0.4. We can see that when the stopping criteria is small, more measurements are needed to achieve the same NLE. On the contrary, when stopping criteria is large, less measurements can meet the requirement, which means bigger differences in the real locations and recovery locations. Fig. 7 shows the NLE of OMP and AOMP algorithms with $M_0 = 30$ and $Er = 0.2$ for different SNR scenarios. The number of PUEA is changing from 1 to 20.
We can see that AOMP algorithm has better performance that OMP. Moreover, the NLE received by AOMP algorithm within two grids can be reached with SNR less than 20 dB, which is a considerable saving compare to 40 dB when OMP algorithm is used. And the better performance of AOMP algorithm owes to the adaptivity to the changing detection environment.

IV. CONCLUSION AND FUTURE WORK

In this paper, aimed at the detection of primary emulation attacks in cognitive radio networks, we introduced CS into the localization of PUEA transmitters. The locations of transmitters can be reconstructed correctly with lower computation based on CS theory. Furthermore, we proposed AOMP algorithm for dealing with more practical cases where transmitter number and locations are unknown and changing with time. In AOMP algorithm, we provided a stopping criteria to determine if the current measurement number is sufficient enough for required detection precision. Because of the good adaptability of AOMP algorithm, small location errors of transmitters can always be obtained with different environments. In the future, the changing features of PUEA will be explored to improve the detection performance of AOMP algorithm. Moreover, the chosen of running times \( L \) to establish a map of YAMSE versus measurements is also worth further consideration.

REFERENCES