Robust Cooperative Sensing via State Estimation in Cognitive Radio Networks

Alexander W. Min,[†] Kyu-Han Kim,[‡] and Kang G. Shin[†]

[†]Real-Time Computing Laboratory, EECS Dept., The University of Michigan, Ann Arbor, MI 48109-2121
[‡] Deutsche Telekom Inc., R&D Laboratories USA, Los Altos, CA 94022-1530
{alexmin, kgshin}@eecs.umich.edu, kyu-han.kim@telekom.com

***Abstract*—Cooperative sensing, a key enabling technology for dynamic spectrum access, is vulnerable to various sensing-targeted attacks, such as the primary user emulation or spectrum sensing data falsification. These attacks can easily disrupt the primary signal detection process, thus crippling the operation of dynamic spectrum access. While such sensing-targeted attacks can be easily launched by an attacker, it is very challenging to design a robust cooperative spectrum sensing scheme due mainly to the practical constraints inherent in spectrum sensing, particularly the shared/open nature of the wireless medium and the unpredictability of signal propagation. In this paper, we develop an efficient, yet simple attack detection framework, called** `IRIS` **(*robust cooperatIve sensing via iteRatIve State estimation*), that safeguards the incumbent detection process by checking the consistency among sensing reports via the estimation of system states, namely, the primary user's transmit-power and path-loss exponent. The key insight behind the design of** `IRIS` **is that the sensing results are governed by the network topology and the law of signal propagation, which cannot be easily compromised by an attacker. Consequently, the sensing reports must demonstrate consistency among themselves in estimating system states. Our analytical and simulation results show that, by performing consistency-checks,** `IRIS` **provides high attack-detection capability, and preserves satisfactory performance in estimating the system states even under very challenging attack scenarios. Based on these observations, we propose a new incumbent detection rule that can further improve the spectrum efficiency.** `IRIS` **can be readily deployed in infrastructure-based cognitive radio networks, such as IEEE 802.22 WRANs, with manageable processing and communication overheads.**

I. Introduction

Solving the expected spectrum scarcity problem becomes increasingly important to accommodate emerging wireless services and ever-increasing wireless spectrum demands [1], [2]. Cognitive radio (CR) is a key technology to mitigate the overcrowding of spectrum space, e.g., cellular and ISM bands, by enabling unlicensed users to *opportunistically* communicate over the licensed spectrum bands left idle by the incumbents [3]. Among the various challenges in realizing this new concept of opportunistic spectrum access, spectrum sensing—the detection of primary signals and spectrum whitespaces—has been considered the key enabler [4], [5].

While cooperative sensing has proven to be a viable means to improve the primary detection performance to meet the stringent detectability requirements imposed by the regulatory body (e.g., FCC) [4]–[7], it is vulnerable to various critical security attacks. In the absence of attack, cooperative sensing can significantly improve the incumbent detection performance by carefully choosing a suitable set of cooperative sensors [5], [8], [9]. However, sensors are often deployed in open hostile environments, and can be compromised by an attacker or exposed to external interferences that can distort the measurement results. Therefore, their sensing reports cannot be fully trusted. Moreover, the vulnerability to attacks is exacerbated further by several unique features in opportunistic spectrum access, such as spatial and temporal variations of primary signal characteristics, and easy accessibility and high reconfigurability of the low-layer protocol stacks in software-defined radio-based CRs (e.g., USRP [10] or Sora [11]). To cripple the operation of cooperative sensing, an attacker can (i) create an illusion of a primary signal by simply broadcasting a falsified primary signal, i.e., the primary user emulation attack (PUEA) [12], [13], or (ii) physically compromise the sensor and manipulate its sensing reports, i.e., the spectrum sensing data falsification (SSDF) attack [14]–[16]. Accordingly, these attacks can significantly impair the incumbent detection process, resulting in either excessive interference to primary communications or waste of spectrum opportunities, denying the basic premise of opportunistic spectrum access.

The design of robust cooperative spectrum sensing is, however, a challenging problem, and existing approaches have their own limitations. It is intrinsically difficult to detect manipulated sensing reports due mainly to the absence of "ground truth" in sensing reports. That is, there is no convenient way to determine whether the sensing reports are the true estimates of a primary signal strength or distorted by attackers (directly in the case of SSDF or indirectly in the case of PUEA). Moreover, most existing approaches share two key shortcomings. First, attack detection and filtering schemes are designed heuristically without any rigid attack-detection criterion. As a result, they may not work properly against various network and attack scenarios [13], [14], [16]. Second, the design of most existing schemes is tailored to a specific attack type, e.g., PUEA [13] or SSDF [15], and thus may be unable to cope with general types of attack. These shortcomings limit the applicability of existing schemes in real network environments under unpredictable attacks. Therefore, there is a clear need for developing a robust cooperative sensing framework that can preserve high incumbent detection performance even under challenging attack scenarios.

In this paper, we propose a robust cooperative spectrum sensing framework, called `IRIS` (*robust cooperatIve sensing via iteRatIve State estimation*), that withstands sensing-targeted attacks by weeding out any abnormal sensing reports regardless of the actual cause (or attack type) of such deviations. The design of `IRIS` is motivated by the observation that the measured primary signal strengths at sensors are governed by the topology of the network and the law of signal propagation at the PHY-layer, which cannot be easily compromised by attackers. IRIS estimates the system states—i.e., the transmit-power of the primary transmitter and the path-loss exponent—based on the sensing reports, and monitors the measurement residual, which indicates how close the sensing reports are to the normal value of the received signal strengths. Thus, any sensing report with a large deviation (or attack strength) will result in a large measurement residual. Once IRIS detects the existence of an abnormal sensing report(s), it can accurately pinpoint the manipulated sensing reports and remove them from the incumbent detection process. To evade such detection, attackers must lower their attack strengths, and thus, the lowered attack strength makes a negligible impact on primary detection accuracy.

A. Contributions

The main contributions can be summarized as follows.

- Introduction of *joint* cooperative sensing and system state estimation for robust incumbent detection. This is very different from most previous sensing schemes that focus only on the detection of a primary signal [5]. The state estimation introduced in IRIS provides a useful criterion for detecting the existence of attacks and further pinpointing and removing the manipulated sensing reports.
- Analysis of the attack-tolerance of IRIS. In particular, we show that it is infeasible for an attacker to completely evade the detection rule in IRIS, unless the attacker compromises all but one cooperating sensor, and simultaneously controls the sensing reports, demonstrating its high attack-tolerance.
- In-depth evaluation of the performance of IRIS under various attack scenarios. Our evaluation results show that IRIS successfully detects and removes the manipulated sensing reports by checking the consistency among sensing reports through the measurement residual. We also study the interesting tradeoff in the selection of attack detection threshold.
- Proposal of a new approach to incumbent detection based on the estimated transmit-power level. We show that IRIS accurately estimates primary users' transmit-power level even under very challenging attack scenarios where the majority of sensors are compromised, making this approach highly attractive.

B. Organization

The remainder of this paper is organized as follows. Section II reviews the existing approaches to robust cooperative sensing in cognitive radio networks (CRNs). Section III introduces the network and attack models and assumptions that we will use throughout the paper. Section IV describes the IRIS framework and formulates the attack detection problem as a hypothesis testing. Section V proposes methods for attack detection and identification, analyzes their attack detection capability, and describes the IRIS algorithm. Section VI evaluates the performance of IRIS under various types of attacks in realistic wireless environments, and Section VII concludes the paper.

II. Related Work

The design of robust cooperative sensing has recently received considerable attention. For example, statistics-based anomaly detection has been proposed [12], [14], [17], [18]. Chen *et al.* [12] proposed a sensor reputation management framework that assigns different weights to sensing reports based on their reputation achieved from a previous sensing history. Kaligineedi *et al.* [14] proposed a simple method to filter out any outliers among the sensing reports. However, this method is agnostic of the existence or type of attack, and its attack detection rule is not adaptive to different attack scenarios.

An effort has also been made to design robust cooperative sensing schemes by exploiting various aspects of primary users' characteristics [13], [15], [16], [19]–[21]. Chen *et al.* [13] proposed to exploit the primary user's location information to verify the identity of the signal source in order to defeat the primary emulation attack. However, their scheme, called LocDef, requires a separate dense sensor network for localization. Moreover, it is designed only to defeat PUEA. In contrast, IRIS targets robust detection of primary signals in the presence of *any* type of attacks that can potentially affect the sensing resorts. Recently, Min and Shin [15] proposed to exploit shadow fading correlation in the received primary signal to detect and remove abnormal sensing reports. However, their scheme, called ADSP, assumes the existence of sensor clusters, which might not always hold in practice.

Recently, primary signal propagation characteristics at the PHY-layer have been exploited to detect abnormality in CRNs [19], [20]. Liu *et al.* [19] proposed an anomaly detection framework, called ALDO, that monitors the path-loss exponent in signal propagation to detect selfish secondary users violating spectrum etiquette, i.e., using spectrum bands without authorization. Lie *et al.* [20] exploited the location-dependent link signature (i.e., multipath fading profile) along with conventional cryptographic authentication to detect a falsified primary signal. Unlike the above approaches, we propose to exploit the consistency among sensing reports in regard to physical characteristics, such as network topology and signal propagation, to detect and remove any manipulated sensing reports.

In a broader context, our paper is related to secure data aggregation [22], [23], insider attack detection [24], and secure localization and target tracking [25] in wireless sensor networks. However, the problem considered in this paper differs from them since (i) it focuses on a unique case in CRNs where attackers manipulate the sensor reports to disrupt the cooperative sensing process, and (ii) in CRNs, no modification should be required to the primary system, and thus, the measured received primary signal strengths obtained via spectrum sensing is the only available information to the secondary system.

In summary, IRIS differs from most previous work in three key aspects. First, IRIS exploits network topology and signal propagation characteristics to validate the integrity of sensing reports. Second, IRIS accurately identifies those sensing reports affected by attackers (either directly or indirectly) since the detection focuses on the consistency among sensing reports, which must always be maintained in the absence of attacks. Third, IRIS detects an incumbent signal by estimating the transmit-power based on sensing reports. This is very different from most existing fusion schemes, such as k-out-of-N rule and equal gain combining, which rely solely on sensing reports.

III. System and Attack Models

In this section, we first describe the network, cooperative sensing, signal propagation, and system state estimation model, and assumptions that we will use throughout the paper. We then describe the attack model.

A. Network and Cooperative Spectrum-Sensing Model

We consider a CRN where both primary and secondary users coexist in the same geographical area, as shown in Fig. 1. We assume an infrastructure-based secondary network, e.g., 802.22 WRANs, where each cell consists of a single base station (BS) and multiple secondary users (sensors).[1] Since the BS is maintained by an expert, we assume that BSes are trusted. Each BS coordinates the opportunistic spectrum access of secondary users in its cell by directing a (sub)set $\mathcal{N}$ of sensors to perform spectrum sensing periodically for primary signal detection. At the end of each sensing period, cooperative sensors report their measurement results (i.e., sensing output) to the BS to make a final decision on the presence or absence of a primary signal. Finally, the BS broadcasts the final decision to the secondary users within the cell. The sensing reports and final decisions are communicated through a reliable, dedicated control channel. We assume that the BS knows the location of the primary transmitter and sensors.[2]

For spectrum sensing, the energy detector [28] is used as the physical-layer sensing technology mainly because of its simple design and low overhead. Cooperative sensors simply measure the primary signal power on a target frequency band using the energy detector and reports the sensing results to the BS for the detection of a primary signal. We assume that sensors do not move together in close proximity, and thus, they produce independent measurement results.

B. Signal Propagation Model

The received primary signal strength at a cooperative sensor $i \in \mathcal{N}$ can be expressed as [6]:

$$P_{R,i} = P_o + \alpha 10 \log_{10}(d_o/d_i) + w_i, \quad \text{(dB)} \tag{1}$$

where P_o is the received power at the reference distance d_o, α the path-loss exponent, which typically ranges from 2 to 5; α depends on a network environment and we assume that it is not known *a priori* to the secondary system, d_i the distance between the primary transmitter and sensor i, and w_i the measurement error, which accounts for the errors in the energy detector, multi-path fading, and shadow fading. We use this signal propagation model as the "ground truth". Note that, although we use a simple signal propagation model, our proposed attack-detection method is generic, and does not rely on any specific choice of model.

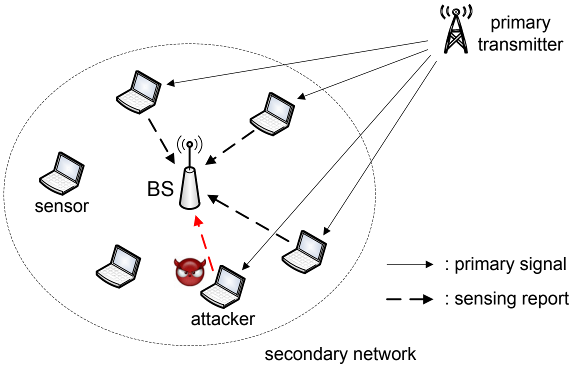


Fig. 1. **A CRN model with primary-secondary coexistence**: During each sensing period, cooperative sensors perform energy detection to measure the received primary signal strength, and report the sensing results to the BS to determine the presence/absence of a primary signal as well as to estimate its states, e.g., transmit-power. Sensing reports could be contaminated by, for example, attacks or hardware/software faults, and `IRIS` at the BS detects such abnormal sensing reports via *iterative* state estimation.

We assume that multi-path fading can be ignored when sensors perform spectrum sensing on a wide channel bandwidth for a long period of time, e.g., coherence time [29]. For example, in 802.22 WRANs, the impact of multi-path fading can be ignored when sensing the entire 6 MHz TV channel [30]. The shadow fading gain is location-dependent and it can thus be estimated at each sensor location. So, the fading gain for each sensor can be considered as a specific realization of a normal random variable [5].

C. System State Estimation Model

We define the problem of state estimation as that of estimating the primary transmitter's power P_o and the path-loss exponent α, i.e.,

$$\mathbf{x} \triangleq [P_o \, \alpha]^T. \tag{2}$$

Let $\mathbf{P_R} = [P_1, \ldots, P_N]^T$ denote the vector of the primary's signal strength received at N cooperative sensors. Then, the primary signal power (i.e., the energy detector output) received at cooperative sensors can be expressed as:

$$\mathbf{P_R} = \mathbf{H}\mathbf{x} + \mathbf{w}, \quad \text{(dB)} \tag{3}$$

where $\mathbf{H}$ is the channel gain matrix between the primary transmitter and the cooperating sensors, i.e.,

$$\mathbf{H} \triangleq \begin{bmatrix} 1 & 10\log_{10}(d_o) - 10\log_{10}(d_1) \\ \vdots & \vdots \\ 1 & 10\log_{10}(d_o) - 10\log_{10}(d_N) \end{bmatrix}_{N \times 2}, \tag{4}$$

where N is the number of cooperating sensors. Note that $\mathbf{H}$ is determined by the network topology, i.e., $\{d_i\}_{i=1}^{N}$. We assume that noise $\mathbf{w} = [w_1, \ldots, w_N]^T$ at cooperating sensors can be approximated as Gaussian distribution, i.e., $w_i \sim \mathcal{N}(0, \sigma_{w,i}^2)$.

Given that the noise $\mathbf{w}$ follows a Gaussian distribution, the maximum likelihood estimator (MLE) of the state variables can be expressed as [31]:[3]

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{\Sigma}_w \mathbf{H})^{-1} \mathbf{H}^T \mathbf{\Sigma}_w \mathbf{P_R}, \tag{5}$$

[1] We use the terms *secondary users* and *sensors* interchangeably as we focus on spectrum sensing functionality of secondary users.

[2] It is relatively easy for the BS to obtain the location of large-scale primary transmitter, e.g., TV transmitters, via geo-location database [3], [26], [27].

[3] The estimate in Eq. (5) is also the solution to the *weighted least-square criterion* and the *minimum variance criterion*.

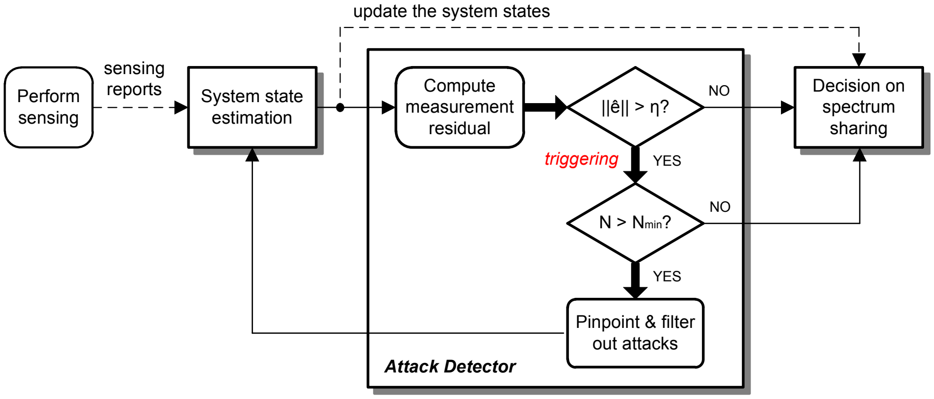


Fig. 2. **The** IRIS **framework**: IRIS resides at the BS and safeguards the cooperative sensing by filtering out abnormal sensing reports via iterative estimation of the system state parameters (i.e., the transmit-power P_o and the path-loss exponent α). The estimated state parameters will be used for the detection of a primary signal and spectrum reuse planning.

where $\mathbf{\Sigma}_w$ is a diagonal matrix whose elements are reciprocals of the variances of the measurement errors:

$$\mathbf{\Sigma}_w \triangleq \begin{bmatrix} \sigma_{w,1}^{-2} & & & \\ & \sigma_{w,2}^{-2} & & \\ & & \ddots & \\ & & & \sigma_{w,N}^{-2} \end{bmatrix}, \tag{6}$$

where $\sigma_{w,i}^2$ is the variance of the cooperative sensor i. For the ease of presentation, we assume that $\sigma_{dB,i} = \sigma_{dB}$ and $\sigma_{m,i} = \sigma_m \; \forall i$.

Then, the state vector $\mathbf{x} = [P_o \; \alpha]^T$ can be expressed as:

$$\begin{aligned} \hat{\mathbf{x}} = \begin{bmatrix} \hat{P}_o \\ \hat{\alpha} \end{bmatrix} &= (\mathbf{H}^T \mathbf{\Sigma}_w \mathbf{H})^{-1} \mathbf{H}^T \mathbf{\Sigma}_w \mathbf{P_R} \\ &= (\mathbf{H}^T \mathbf{\Sigma}_w \mathbf{H})^{-1} \mathbf{H}^T \mathbf{\Sigma}_w (\mathbf{H}\mathbf{x} + \mathbf{w}) \\ &= \mathbf{x} + (\mathbf{H}^T \mathbf{\Sigma}_w \mathbf{H})^{-1} \mathbf{H}^T \mathbf{\Sigma}_w \mathbf{w}. \end{aligned} \tag{7}$$

The estimated state vector $\hat{\mathbf{x}}$ will be used for calculating the measurement residual (in Eq. (10)), based on which IRIS detects and filters out abnormal sensing reports. We will elaborate on the design of the attack detector in Section IV.

We define the *estimation error* of the state vector $\mathbf{x}$ as:

$$\phi \triangleq \|P_o - \hat{P}_o\|_2 + \|\alpha - \hat{\alpha}\|_2, \tag{8}$$

where $\| \bullet \|_2$ represents the L_2 norm. We will henceforth omit the subscript for brevity. The estimation error will be used to evaluate the estimation performance of IRIS in Section VI.

D. Attack Model

1) Attack Scenarios: We consider the following scenarios.

- The sensors may be *compromised* by the attackers, and their sensing reports are manipulated by the attackers,
- The sensors may be *exposed* to a falsified primary signal emitted by an attacker, and their measurements are biased,
- The sensors may be *faulty*, thus sending erroneous sensing reports with a non-zero offset.

A common consequence of the above attack/fault scenarios is that the sensing results, i.e., measured received signal strengths (RSSs), reported to the BS will be distorted somewhat, thus affecting the final state estimation results. So, we focus on the first two attack scenarios, which are more challenging. In particular, we consider the following attack scenarios: Attackers

- can compromise multiple (e.g., up to k) sensors, and craft their sensing reports simultaneously, and
- are intelligent enough to know the presence of a primary signal and attack detection rule of the fusion center.

2) Attacker's Goal: The goal of an attacker is to mislead the BS to make a wrong decision on the primary users' channel usage activity. For example, when no primary signal exists, the attacker may inject positive offsets to the sensing reports to create an illusion of a primary signal, causing unnecessary channel vacation of secondary users, and vice versa. These attacks will ultimately result in either excessive interference to primary communications or waste of spectrum opportunities. To achieve this goal, the attacker must be able to incur a significant error in the state estimation, e.g., transmit-power $\hat{P}_o$, while not being detected by the fusion center. We study the feasibility of bypassing the attack detector via analysis in Section V-B and via simulations in Section VI.

3) Final Sensing Reports: The final sensing reports from the cooperative sensors can be expressed as follows:

$$\mathbf{P_R^a} = \mathbf{P_R} + \mathbf{a}, \tag{9}$$

where $\mathbf{P_R}$ is the received primary signal strength vector in Eq. (3) and $\mathbf{a} = [a_1, \ldots, a_N]^T$ is an attack vector where $a_i \in \mathbb{R}$ is an attack strength. Let $\mathcal{M} \subseteq \mathcal{N}$ denote the set of sensors whose sensing reports affected by attackers either directly (i.e., spectrum-sensing data falsification) or indirectly (i.e., primary emulation). If sensor i's report is neither affected by the attack, i.e., $i \notin \mathcal{M}$, nor faulty, then $a_i = 0$; otherwise, the attacker can introduce an arbitrary attack strength.

IV. The Proposed Approach

In this section, we overview the IRIS framework, discuss the design rationale of the IRIS's attack detector, and formulate the attack detection problem as a hypothesis testing.

A. IRIS *Framework*

The IRIS framework, residing at the BS, consists of the following three functional building blocks, as shown in Fig. 2, which closely interact with each other:

- **State estimator** that estimates the system states based on the collected sensing reports at the BS,
- **Attack detector** that detects the existence of abnormal sensing reports, and then pinpoints and filters them out, and
- **Decision-maker** that makes a final decision on the presence of a primary signal based on the estimated transmit-power of the primary transmitter.

First, the *state estimator* and the *attack detector* constitute the core of the IRIS framework. These two components offer three main benefits. They enable the BS to:

- accurately and promptly detect and filter out abnormal sensing reports without requiring any information about attack types or strategies,
- greatly reduce computational overhead since they require the BS to check only a single parameter, i.e., the measurement residual, instead of validating every sensing report as in [14], [15], and
- efficiently coexist with primary users by providing accurate estimation of the transmit-power level of the primary transmitter.

Next, the *decision-maker* in IRIS adopts the threshold-based spectrum access decision rule, as shown in Fig. 3. For example, if the estimated transmit-power $\hat{P}_o$ is above the upper threshold TH_{upper}, all of the secondary users in the system must vacate the channel; if the transmit-power is below the lower threshold TH_{lower} $(\leq TH_{upper})$, then the secondary users can fully utilize the channel without any limit. Otherwise, they can utilize the channel with reduced transmit-power to meet the interference constraints to the primary communications. In Section VI, we demonstrate the feasibility of this approach by showing that IRIS accurately estimates the transmit-power even under challenging attack scenarios. The detection thresholds must be carefully chosen by the system designer so as to maximize incumbent detection performance.

Note that IRIS is designed to stop filtering sensing reports either when the measurement residual (in Eq. (10)) drops below a predefined threshold or when the number of remaining sensing reports becomes less than or equal to N_{min}, whichever comes first. This prevents over-filtering of sensing reports in the case where the measurement residual becomes large due to the smaller set of sensing samples.

B. Design Rationale of the Attack Detector

Here we elaborate on the design principle of the attack detector in IRIS. The key insight behind the attack detection is that the sensing reports must demonstrate *consistency* among themselves since the received primary signal strengths are governed by the physical aspects of the network environments. In particular, we show that the measurement residual in state estimation follows the χ^2-distribution in the absence of attacks (in Section IV-C). To exploit this relationship, IRIS monitors the measurement residual as a criterion for consistency-check among sensing reports. A measurement residual residing outside the expected range will be interpreted as an indication of the existence of abnormal sensing reports. Consequently, any sensing report with a large deviation, regardless of the attack type, will be easily detected by the attack detector and will be filtered out in the state estimation process. Therefore, an attacker must lower the attack strength to evade the detection, thereby being able to make only a negligible impact on the primary detection process. One important feature of IRIS is that it focuses on inconsistency among sensing reports, and thus the attack type, e.g., PUEA or SSDF, makes no difference. The detailed description of the proposed attack detection rule in IRIS that demonstrates this feature will be discussed in Section V.

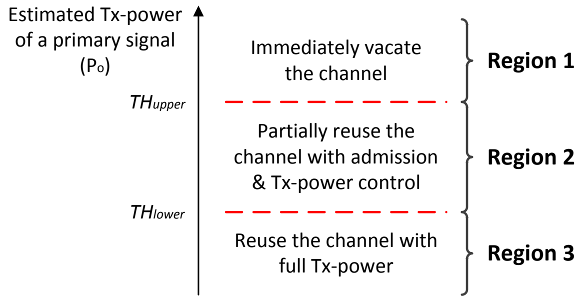

Estimated Tx-power of a primary signal (Pₒ)
THupper
THlower
Immediately vacate the channel
Partially reuse the channel with admission & Tx-power control
Reuse the channel with full Tx-power
Region 1
Region 2
Region 3

Fig. 3. **The proposed incumbent detection rule**: IRIS makes a final decision based on the estimated transmit-power ($\hat{P}_o$) of the primary transmitter.

C. Problem Formulation

We first formulate the attack detection problem as a binary hypothesis testing, and then characterize the distribution of the measurement residual, which will form the basis of our attack detector.

1) Hypothesis Testing: The estimation of measurement residual, i.e., the difference between the measured RSSs and the estimated RSSs, denoted as $\hat{\mathbf{e}}$, can be expressed as:

$$\begin{aligned}\hat{\mathbf{e}} &= \mathbf{P_R} - \mathbf{H}\hat{\mathbf{x}} \\ &= \mathbf{Hx} + \mathbf{w} - \mathbf{H}(\mathbf{x} + (\mathbf{H}^T\boldsymbol{\Sigma}_w\mathbf{H}^{-1})\mathbf{H}^T\boldsymbol{\Sigma}_w\mathbf{w}) \\ &= \mathbf{w} - \mathbf{H}(\mathbf{H}^T\boldsymbol{\Sigma}_w\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\Sigma}_w\mathbf{w} \\ &= (\mathbf{I} - \mathbf{H}(\mathbf{H}^T\boldsymbol{\Sigma}_w\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\Sigma}_w)\mathbf{w}. \end{aligned} \tag{10}$$

The measurement residual evaluates the closeness of the measurements to the truth. Eq. (10) indicates that the measurement residual $\hat{\mathbf{e}}$ depends on the uncertainty in RSSs induced by the measurement noise, i.e., $\mathbf{w}$, and thus attacker-injected deviations in sensing reports will increase the variance of the error.

Given that the noise power $\mathbf{w}$ in Eq. (3) follows a Gaussian distribution, i.e., $\mathbf{w} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_w)$ where $\boldsymbol{\Sigma}_w$ is defined in Eq. (6), the measurement residual vector $\hat{\mathbf{e}}$ in Eq. (10) follows a multivariate Gaussian distribution, i.e.,

$$\hat{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e), \tag{11}$$

where $\boldsymbol{\Sigma}_e = \boldsymbol{\Pi}\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Pi}$, and $\boldsymbol{\Pi} = \mathbf{I} - \mathbf{H}(\mathbf{H}^T\boldsymbol{\Sigma}_w\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\Sigma}_w$.

Then, the attack detection problem can be cast into a binary hypothesis testing problem where the observed measurement residual $\hat{\mathbf{e}}$ belongs to one of two classes, $\mathcal{H}_0$ or $\mathcal{H}_1$, where:

$$\begin{aligned} &\mathcal{H}_0 : \hat{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e) \quad \text{(attack does not exist)} \\ &\mathcal{H}_1 : \hat{\mathbf{e}} \nsim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e) \quad \text{(attack exists).} \end{aligned}$$

The above hypothesis testing indicates that IRIS in the BS will assume the existence of abnormal sensing report(s) if the measurement residual does not belong to the expected distribution under $\mathcal{H}_0$, which can be approximated as χ^2-distribution as we discuss next.

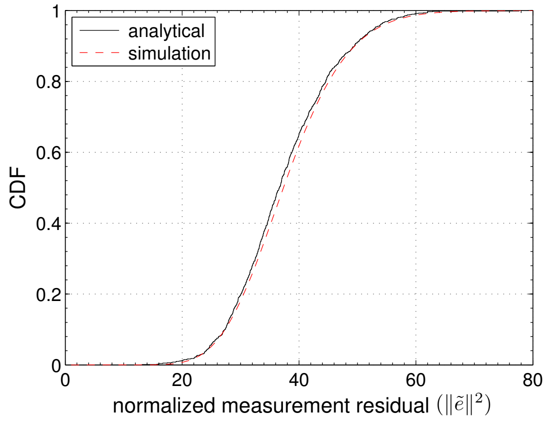


Fig. 4. **Distribution of normalized measurement residual of analysis vs. simulation**: The χ^2 approximation of the measurement residual closely matches the simulation results over 10^3 random network topologies with $N=40$, $P_o=4\,\text{kW}$ and $\sigma_w=0.3\,\text{dB}$.

2) Characterization of the Measurement Residual: Let $\|\tilde{\mathbf{e}}\| = \sqrt{\hat{\mathbf{e}}^T \mathbf{\Sigma}_w \hat{\mathbf{e}}}$ denote the *normalized* L_2 norm of the measurement residual. Then, $\|\tilde{\mathbf{e}}\|^2$ follows χ^2-distribution with $k=N-M$ degrees of freedom where N is the number of cooperative sensors and M is the number of state variables (i.e., $M=2$) [32]. Thus, the cumulative distribution function (c.d.f.) of the normalized L_2 norm of the measurement residual $\|\tilde{\mathbf{e}}\|^2$ is given as:

$$\mathbb{F}_{\chi^2}(x;k) = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)}, \tag{12}$$

where $\Gamma(\bullet)$ is the Gamma function, and $\gamma(\bullet,\bullet)$ is the lower incomplete Gamma function.

Fig. 4 compares the empirical c.d.f. of the normalized measurement residual (dotted line) and that of the approximated χ^2-distribution (solid line). The figure shows that the simulation results closely match the analytical results. Although it is not shown in the figure, we observed that the approximation is very accurate even with a smaller number of sensors (i.e., $N<40$).

V. Attack Detection

In this section, we design a rule for detecting attacks and a method for pinpointing abnormal sensing reports. We then analyze the possibility of evading the detection rule, and finally present algorithms for state estimation and attacker detection.

A. Design of Attack-Detection Rule

Based on Eq. (12), IRIS checks the consistency among the sensing reports by comparing the measurement residual with a predefined threshold value. Specifically, IRIS uses the following rule for attack detection:

$$\delta_{L_2} = \begin{cases} 1, & \|\tilde{\mathbf{e}}\| > \eta \qquad \mathcal{H}_0 \\ 0, & \text{otherwise}, \qquad \mathcal{H}_1 \end{cases} \tag{13}$$

where the attack-detection threshold η is chosen to meet a desired level of false-alarm (i.e., triggering attack detection and filtering process) probability P^*_{FA}.

The probability of attack false-alarm with the decision threshold $\eta \in \mathbb{R}$ is given as:

$$\begin{aligned} P^a_{FA} &\triangleq Pr(\|\tilde{\mathbf{e}}\| > \eta \,|\, \mathcal{H}_0) \\ &= 1 - \mathbb{F}_{\chi^2}(\eta^2;k), \end{aligned} \tag{14}$$

where $\mathbb{F}_{\chi^2}(\cdot)$ is the c.d.f. of the measurement residual defined in Eq. (12).

In the absence of attack, the probability that the attack detector filters out $n\in\mathbb{N}$ legitimate sensing reports is $(P^a_{FA})^n$. For example, when $P^a_{FA}=0.1$, the probability that IRIS will mistakenly filters out 5 legitimate sensing reports is 10^{-5}.

Based on Eq. (14), the decision threshold η to achieve a desired level of false-alarm rate P^*_{FA} is given as:

$$\eta = \sqrt{\mathbb{F}^{-1}_{\chi^2}(1-P^*_{FA};k)}. \tag{15}$$

Based on Eqs. (14) and (15), the attack mis-detection probability (i.e., no triggering of the attack detection even in the presence of manipulated sensing reports) can be expressed as:

$$P^a_{MD} \triangleq Pr(\|\tilde{\mathbf{e}}\| < \eta \,|\, \mathcal{H}_1, \mathbf{a} \neq \mathbf{0}). \tag{16}$$

The above attack-detection rule in Eq. (13) offers two main benefits as follows.

- Its detection performance depends only on the number of cooperative sensors, N, and thus, it works well under both hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$.
- It is lightweight, requiring the BS to check only a single parameter, i.e., the measurement residual, instead of validating each sensing report as in [14], [15].

These advantageous features make the attack-detection rule in Eq. (13) suitable for various network environments and attack scenarios with minimal processing overheads.

Remark: A key problem is the choice of attack-detection threshold η, by which IRIS can strike a balance between attack false-alarm (P^a_{FA}) and mis-detection rate (P^a_{MD}). However, our simulation study indicated that the threshold η set based on the desired level of false-alarm rate may work well only for the case of a small number of compromised sensors. When multiple sensing reports are compromised, η needs to be set aggressively to achieve accurate estimation results. The impact of η on the performance of IRIS will be detailed in Section VI-C.

B. An Attack Strategy to Evade the Detector

We study the attack-detection performance by analyzing the conditions under which the attack detection rule in Eq. (13) can be evaded. In particular, we consider the worst-case scenarios where an attacker compromises a set $\mathcal{M} \in \mathcal{N}$ of sensors and controls their sensing reports simultaneously in order to maximize the chance to evade the detection rule. Note that this is much stealthier than PUEA since the attacker has a much finer-grained control over sensing reports.

We assume that the attack vector, introduced in Eq. (9), is $\mathbf{a} \neq \mathbf{0}$ and let $\hat{\mathbf{x}}_a = \hat{\mathbf{x}} + \mathbf{c}$ denote the vector of estimated state variables, i.e., $\hat{P}_o$ and $\hat{\alpha}$, where $\mathbf{c}$ is the vector of state estimation errors induced by the attack vector $\mathbf{a}$. Therefore, if

$\mathbf{a}\neq\mathbf{0}$, then $\mathbf{c}\neq\mathbf{0}$. The measurement residual in the presence of a non-zero attack vector $\mathbf{a}$ can be expressed as:

$$\begin{aligned}\|\tilde{\mathbf{e}}_a\| &= \sigma_w^{-1}\left\|\mathbf{P}_{\mathbf{R}}^{\mathbf{a}} - \mathbf{H}\hat{\mathbf{x}}_a\right\| \\ &= \sigma_w^{-1}\left\|\mathbf{P}_{\mathbf{R}} + \mathbf{a} - \mathbf{H}(\hat{\mathbf{x}} + \mathbf{c})\right\| \\ &= \sigma_w^{-1}\left\|\mathbf{P}_{\mathbf{R}} - \mathbf{H}\hat{\mathbf{x}} + \mathbf{a} - \mathbf{H}\mathbf{c}\right\| \\ &\leq \sigma_w^{-1}\left\|\mathbf{P}_{\mathbf{R}} - \mathbf{H}\hat{\mathbf{x}}\right\| + \sigma_w^{-1}\left\|\mathbf{a} - \mathbf{H}\mathbf{c}\right\|. \end{aligned} \tag{17}$$

Recall that the goal of an attacker is to keep the measurement residual below the detection threshold, i.e., $\|\tilde{\mathbf{e}}_a\| < \eta$, to evade detection, while affecting the estimation results.

Lemma 1 *Attacks will not be detected by the attack-detector in Eq.* (13) *when the attack vector is set to* $\mathbf{a}=\mathbf{H}\mathbf{c}$.

Proof: Assume that the original measurement $\mathbf{P}_{\mathbf{R}}$ can evade the attack detector, i.e., $\|\tilde{\mathbf{e}}\| = \sigma_w^{-1}\left\|\mathbf{P}_{\mathbf{R}} - \mathbf{H}\hat{\mathbf{x}}\right\| \leq \eta$. If the attack vector satisfies the condition $\mathbf{a}=\mathbf{H}\mathbf{c}$, then from Eq. (17), we have:

$$\|\tilde{\mathbf{e}}_a\| \leq \sigma_w^{-1}\left\|\mathbf{P}_{\mathbf{R}} - \mathbf{H}\hat{\mathbf{x}}\right\| + \sigma_w^{-1}\underbrace{\left\|\mathbf{a} - \mathbf{H}\mathbf{c}\right\|}_{\mathbf{0}} \leq \eta. \tag{18}$$

Thus, the lemma follows. ∎

The lemma indicates that the attacker must be able to find an attack vector $\mathbf{a}$ such that $\mathbf{a} = \mathbf{H}\mathbf{c}$, i.e., the attack vector $\mathbf{a}$ must be a linear combination of column vectors of $\mathbf{H}$, to ensure it will evade the attack detector. Here we examine the feasibility of constructing such an attack vector $\mathbf{a}$ s.t. $\mathbf{a}=\mathbf{H}\mathbf{c}$.

The attacker can find an attack vector satisfying the condition $\mathbf{a} = \mathbf{H}\mathbf{c}$ as follows. First, let us define the projection matrix of $\mathbf{H}$ as $\mathbf{P}=\mathbf{H}(\mathbf{H}^T\mathbf{\Sigma}_w\mathbf{H})^{-1}\mathbf{\Sigma}_w\mathbf{H}^T$ and $\mathbf{B}=\mathbf{P}-\mathbf{I}$. Then, we can obtain an equivalent representation of $\mathbf{a}=\mathbf{H}\mathbf{c}$ as [32]:

$$\begin{aligned}\mathbf{a} = \mathbf{H}\mathbf{c} &\Leftrightarrow \mathbf{P}\mathbf{a} = \mathbf{P}\mathbf{H}\mathbf{c} \Leftrightarrow \mathbf{P}\mathbf{a} = \mathbf{H}\mathbf{c} \Leftrightarrow \mathbf{P}\mathbf{a} = \mathbf{a} \\ &\Leftrightarrow \mathbf{P}\mathbf{a} - \mathbf{a} = \mathbf{0} \Leftrightarrow (\mathbf{P}-\mathbf{I})\mathbf{a} = \mathbf{0} \\ &\Leftrightarrow \mathbf{B}\mathbf{a} = \mathbf{0}. \end{aligned} \tag{19}$$

Let us consider the case where an attacker compromises k specific sensors. Then, the attack vector is $\mathbf{a} = (0,\ldots,0,a_1,0,\ldots,0,a_i,0,\ldots,0,a_k,0,\ldots,0)^T$ where $a_i\neq 0$ if $i \in \mathcal{M}$. Let $\mathbf{a}' = [a_1',\ldots,a_k']^T$ where a_j' is the j-th non-zero element in the original attack vector $\mathbf{a}$ and let $\mathbf{B}' = (\mathbf{b}_1',\ldots,\mathbf{b}_k')$ the columns of $\mathbf{B}$ corresponding to the non-zero elements of $\mathbf{a}$. Then, we can simplify the condition $\mathbf{a}=\mathbf{H}\mathbf{c}$ as:

$$\mathbf{B}\mathbf{a} = \mathbf{0} \Leftrightarrow \mathbf{B}'\mathbf{a}' = \mathbf{0}. \tag{20}$$

The following proposition derives the sufficient condition on the minimum number of sensors that an attacker must compromise to find an attack vector s.t. $\mathbf{a}=\mathbf{H}\mathbf{c}$.

Proposition 1 *Assuming a full-rank matrix* $\mathbf{H}$*, an attacker must compromise at least k specific sensors, where* $k \geq N-M+1$*, to find attack vectors* $\mathbf{a}=\mathbf{H}\mathbf{c}$ *such that* $\mathbf{a}\neq\mathbf{0}$ *and* $a_i=0$ *for* $i\notin\mathcal{M}$.

Proof: Based on Eq. (19), $\mathbf{a} = \mathbf{H}\mathbf{c} \Leftrightarrow \mathbf{B}\mathbf{a} = \mathbf{0}$ where $\mathbf{B} = \mathbf{P} - \mathbf{I} = \mathbf{H}(\mathbf{H}^T\mathbf{\Sigma}_w\mathbf{H})^{-1}\mathbf{\Sigma}_w\mathbf{H}^T - \mathbf{I}$. The channel gain matrix $\mathbf{H}$ (defined in Eq. (4)) is an $N\times 2$ matrix, and thus its rank is $\texttt{rank}(\mathbf{H})=2$. $\mathbf{P}$ is the projection matrix of $\mathbf{H}$, and its rank is $\texttt{rank}(\mathbf{P}) = \texttt{rank}(\mathbf{H}) = 2$. Since $\mathbf{B} = \mathbf{P} - \mathbf{I}$, its rank is $\texttt{rank}(\mathbf{B}) = N-2$, and $\texttt{rank}(\mathbf{B}') \leq N-2$. When $k \geq N-M+1$, $\texttt{rank}(\mathbf{B}') \leq N-2 \leq k-1$, and thus, $\texttt{rank}(\mathbf{B}') < k$. This indicates that $\mathbf{B}'$ is a rank-deficient matrix when $k \geq N-M+1$, and thus, there exist an infinite number of non-zero solutions for $\mathbf{a}'$ that satisfies the relation $\mathbf{B}'\mathbf{a}' = \mathbf{0}$. Therefore, an attacker can always find an attack vector $\mathbf{a}$ that can evade the detection rule by compromising at least $N-M+1$ sensors. ∎

Proposition 1 indicates that it is possible for an attacker to launch a powerful attack that can completely evade the detection, while affecting the estimation results of the system states. However, launching such a stealthy attack requires the attacker to physically capture and compromise all but one cooperative sensor, i.e., $N-1$, since $M=2$ in our case, which is infeasible even for a capable attacker as the sensors are likely to be distributed in a large geographical area. For example, in 802.22 WRANs, the cell radius is typically 33 km (up to 100 km) [8]. We can thus conclude that the L_2-norm detector in `IRIS` can reliably detect the existence of attacks even in challenging attack scenarios where a significant fraction of the sensing reports are manipulated.

The assumption of a full-rank of matrix $\mathbf{H}$ in Proposition 1 is reasonable, since sensors are highly likely to be located at different distances from the primary transmitter, i.e., $d_i \neq d_j\ \forall i \neq j \in \mathcal{M}$ in Eq. (4). If some of the sensors are co-located in close proximity and are thus of about same distance to the primary transmitter, then the rank of $\mathbf{H}$ may be reduced, making it easier to find an attack vector that satisfies $\mathbf{a}=\mathbf{H}\mathbf{c}$.

C. Pinpointing Abnormal Sensing Reports

Once `IRIS` detects the existence of manipulated sensing reports using the detection rule in Eq. (13), it proceeds to identify the most suspicious sensing report and filters it out. `IRIS` repeats this process until the remaining sensing reports pass the detection rule. To accurately pinpoint the manipulated sensing reports without incurring extra overhead, `IRIS` excludes the sensing report with the largest normalized residual, i.e., `IRIS` filters out the sensor i^*'s report such that $i^* = \max_{i\in\mathcal{S}}\{|\mathrm{e}_i|\}$. This is also known as the *largest normalized residual criterion* [33].

D. `IRIS` *Algorithm*

Algorithm 1 describes the three-step approach of `IRIS`.

S1. Once the BS collects the sensing reports, `IRIS` estimates the state variables, i.e., $\hat{P}_o$ and $\hat{\alpha}$, and calculates the normalized measurement residual, i.e., $\|\tilde{\mathbf{e}}\|$. Then, `IRIS` compares the measurement residual with the attack detection threshold η set to achieve a desired attack false-alarm rate.

S2. If the measurement residual exceeds the threshold, `IRIS` assumes that there exists at least one abnormal sensing report, and pinpoints the sensing report that contributes to the measurement residual most, as described in **Algorithm 2**. `IRIS` repeats this detection and filtering process until the measurement residual drops below the threshold η or the number of remaining sensing reports hits the lower threshold, i.e., `min_num_sensor`.

S3. After the attack detection and filtering process, `IRIS` determines the presence/absence of a primary signal based on the estimated transmit-power level, as we discussed

Algorithm 1 ALGORITHM FOR COOPERATIVE SENSING WITH ITERATIVE STATE ESTIMATION

```
At the end of each sensing period, IRIS performs the following steps
$\mathcal{S} \longleftarrow$ Cooperative sensor set
num_sensor $\longleftarrow |\mathcal{S}|$
min_num_sensor $\longleftarrow f_a \times$num_sensor
// $f_a \in [0, 1]$: fraction of compromised sensing reports
$\eta \longleftarrow$ Set the attack detection threshold
// Step 1. Perform state estimation
$\hat{\mathbf{x}} = (\hat{P}_o, \hat{\alpha}) \longleftarrow$ Update the state estimates based on the sensing reports reported by the sensors in $\mathcal{S}$
$||\tilde{\mathbf{e}}|| \longleftarrow$ Compute the normalized $L_2$ norm of the measurement residual using Eq. (10)
// Step 2. Perform attack detection
while ($||\tilde{\mathbf{e}}|| > \eta$) and (num_sensor $\geq$ min_num_sensor) do
    $i^* \longleftarrow \mathrm{IRIS}_{det}(\mathcal{S}, \tilde{\mathbf{e}})$ // Pinoint the compromised sensor
    $\mathcal{S} \longleftarrow \mathcal{S} \setminus \{i^*\}$ // Filter out the sensor
    num_sensor $\longleftarrow$ num_sensor $-1$
    $\hat{\mathbf{x}} \longleftarrow$ Update the state estimates with the updated sensor set $\mathcal{S}$
    $||\tilde{\mathbf{e}}|| \longleftarrow$ Calculate the $L_2$ norm based on the updated $\hat{\mathbf{x}}$
end while
// Step 3. Perform incumbent detection
if $\hat{P}_o > TH_{upper}$ then
    Primary user exists and vacate the channel
else if $\hat{P}_o < TH_{lower}$ then
    Primary does not exist
else
    Primary user exists and use the channel with reduced transmit-power level
end if
```

Algorithm 2 ALGORITHM FOR PINPOINTING THE ATTACKER

```
Procedure $\mathrm{IRIS}_{det}(\mathcal{S}, \tilde{\mathbf{e}})$
$i^* \longleftarrow \arg\max_{i \in \mathcal{S}}\{|\tilde{e}_i|\}$ // Pinpoint the compromised sensor
return $i^*$
```

in Section IV-A. Finally, IRIS makes a decision on spectrum access based on the estimated transmit-power level.

In essence, IRIS successfully tolerates attacks by accurately detecting the existence of abnormal sensing reports, and pinpointing the compromised sensing reports. The estimated transmit-power level is not only used to detect the incumbent signal, but also to enable more efficient coexistence between primary and secondary systems.

VI. PERFORMANCE EVALUATION

We now evaluate the performance of IRIS via MATLAB-based simulations. We first describe the simulation setting, and then demonstrate IRIS's attack-tolerance under various network environments and attack scenarios.

A. Simulation Setup

We consider a CRN where primary and secondary users coexist and the radius of the secondary cell is 1 km and the primary transmitter is located 5 km away from the secondary BS. The power of the primary transmitter is set to $P_o = 4$ kW. The number of cooperative sensors is $N = 40$ and $N_{min} = 5$, unless otherwise specified. Table I lists the system parameters used in our simulation.

TABLE I
THE SYSTEM PARAMETERS IN SIMULATIONS

Parameter	Value	Comments
D_p	5 km	Distance between primary and secondary BS
R_s	1 km	Radius of secondary network
d_o	5 m	Reference distance
P_o	4 kW	Transmission power
α	4	Path-loss exponent
σ_w	0.3 dB	Noise power variance
N	40	Number of cooperative sensors
N_{min}	5	Minimum number of cooperative sensors

To demonstrate the attack-tolerance of IRIS, we consider two representative attack scenarios as follows.

- *Spectrum Sensing Data Falsification Attack*: An attacker compromises a specific set of sensors, and manipulates the measurement results by injecting arbitrary values to sensing reports.
- *Primary User Emulation Attack*: An attacker broadcasts a falsified primary signal, which affects the neighboring sensors' measurements.

The following two performance metrics are used in our evaluation of IRIS.

- *Estimation error (ϕ)*: we measure the state estimation error defined in Eq. (8). The accuracy of the estimated transmit-power can be translated to the incumbent detection performance.
- *Attack false-alarm (P_{FA}^a) and mis-detection (P_{MD}^a) rate*: we evaluate the probabilities of falsely-triggering and mis-triggering the attack detection defined in Eqs. (14) and (16).

Each simulation is run 10^4 times and their average values are taken as the performance measures.

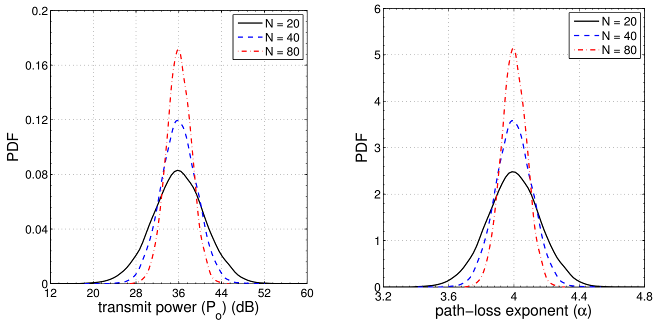
B. Estimation Performance of IRIS

Before delving into the attack scenarios, we first characterize the estimation error performance of IRIS in the *absence* of attacks for various network parameters. In particular, we identify the following three key factors that can affect the estimation performance:

- Number of cooperative sensors (N),
- Measurement error in spectrum sensing (σ_w), and
- Transmit-power of the primary user (P_o).

We study how these network parameters affect the performance of state estimation, which is crucial in the proposed estimated transmit-power-based incumbent detection.

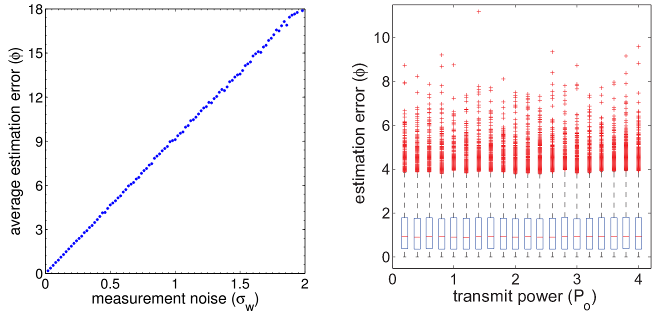
First, a simple way to improve the estimation performance is to increase the number of cooperative sensors. Fig. 5 plots the empirical probability distribution function (p.d.f.) of the estimated parameters, i.e., $\hat{P}_o$ and $\hat{\alpha}$. As expected, the estimation becomes more accurate as the number of cooperative sensors increases. While this will allow the secondary users to efficiently reuse the spectrum based on the estimated primary's transmit-power level, a larger number of cooperative sensors will incur sensing overhead, e.g. time and energy. Therefore, the number of cooperative sensors must be carefully



(a) Transmit-power ($\hat{P}_o$)

(b) Path-loss exponent ($\hat{\alpha}$)

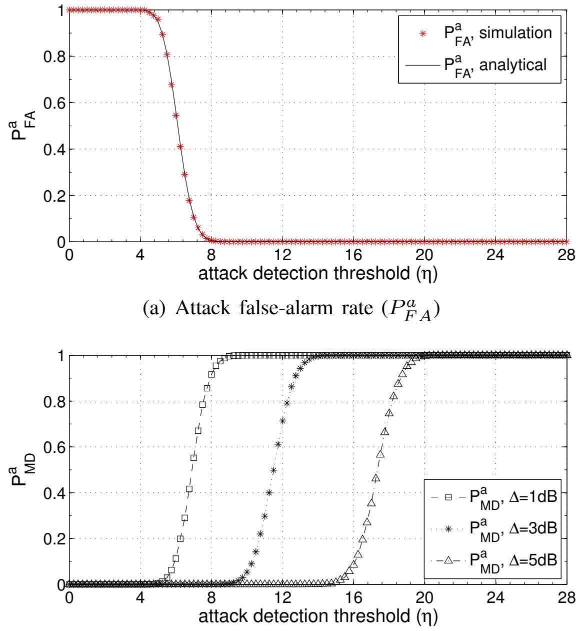
Fig. 5. **Impact of number of cooperative sensors on estimation performance**: State estimation becomes more accurate (i.e., the distribution becomes narrower) as the number of cooperative sensors increases.



(a) Impact of measurement noise

(b) Impact of transmit-power

Fig. 6. **Impact of network parameters on estimation performance**: Estimation error (a) increases almost linearly with the measurement noise, and (b) is independent of the transmit-power of the primary signal.



(a) Attack false-alarm rate (P^a_{FA})

(b) Attack mis-detection rate (P^a_{MD})

Fig. 7. **Impact of the detection threshold η on attack detection performance**: As the attack detection threshold η increases, (a) attack detection false-alarm (P^a_{FA}) decreases and (b) attack mis-detection (P^a_{MD}) increases. For the mis-detection rate, we assume a single compromised sensor $N_a = 1$ and attack strengths in $\Delta \in \{1, 3, 5\}$ dB.

chosen to strike a balance between the sensing overhead, and the spectrum efficiency that can be gained from improved incumbent detection performance [5], [29].

Second, the variance of the measurement in spectrum sensing is also an important factor that can affect the estimation accuracy. The measurement variance depends mainly on (i) shadow and multi-path fading and (ii) inaccuracy of the energy detector. Fig. 6(a) shows that the average estimation error ϕ grows almost linearly as the measurement error (in terms of σ_w) increases. This implies that minimization of the measurement variance is the key to enhance the estimation performance. Fortunately, in stationary sensor environments where sensors do not move, e.g., sensors (also called as CPEs) in IEEE 802.22 WRANs, the shadow fading gains between the primary transmitter and sensors can be estimated via the measurement at the time of sensor deployment [5]. Moreover, the measurement error of the energy detector can be controlled by enlarging the spectrum sensing time [29]. Thus, we assume that the standard deviation of the measurement error is $\sigma_w = 0.3$ dB throughout the simulation.

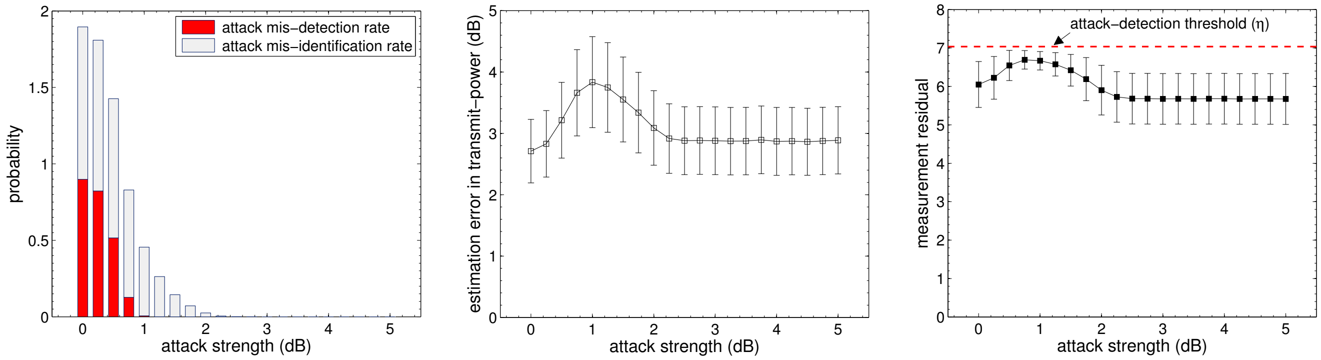
Third, one might think that the estimation performance would be better with a higher transmit-power. However, Fig. 6(b) shows that the average estimation error remains very small (i.e., $\phi < 1$), and does not vary much with the transmit-power level. In fact, this confirms our observation that the estimation error is independent of the transmit-power level, as shown in Eq. (10) in Section III. Thus, the proposed attack detection scheme can be used in a wide range of network environments, regardless of the transmit-power of the primary user or network topology, i.e., relative distances between the primary transmitter and sensors. More importantly, this implies the plausibility of our proposed incumbent detection based on estimated transmit-power. In Section VI-D, we further demonstrate IRIS's high accuracy in estimating the transmit-power even when a significant fraction of sensing reports are compromised.

C. Attack Detection Performance

As discussed in Section IV, a key problem in the design of IRIS is the choice of the attack-detection threshold η. We first investigate the tradeoff between attack false-alarm and mis-detection rates for various attack-detection thresholds. We then study the attack-detection performance in the presence of single and multiple manipulated sensing reports.

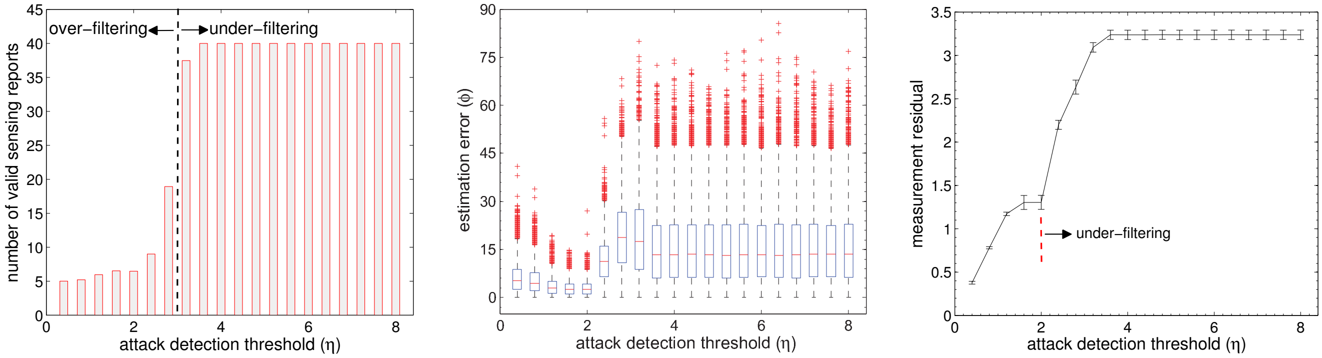
1) Tradeoff in Selecting Attack Detection Threshold: Here we study the impact of attack detection threshold η on attack detection performance. η must be chosen carefully to strike a balance between the attack false-alarm and mis-detection rates. Fig. 7 plots the attack detection performance, i.e., P^a_{FA} and P^a_{MD}, for various attack-detection threshold values in the range $\eta \in [0, 28]$. Fig. 7(a) indicates that, when the threshold is too low, IRIS becomes too aggressive in attack detection, thus suffering from high attack false-alarm rate. The figure also shows that the simulation results closely match the analytical results obtained from Eq. (14) in Section IV.

Fig. 7(b) represents attack mis-detection rate given an assumption that there is a single manipulated sensing report (i.e., $N_a = 1$) with different attack strengths at $\Delta \in \{1, 3, 5\}$ dB. Unlike the false-alarm rate, the attack mis-detection rate increases as η increases. Although the mis-detection rate is



(a) Attack mis-detection/identification rate (b) Estimation error in transmit power (c) Measurement residual

Fig. 8. **Performance of** `IRIS` **with** $N_a=1$: (a) `IRIS` accurately detects the manipulated sensing reports with as weak attack-strength as 2 dB; the attack mis-detection (red bar) and mis-identification rates (gray bar) decrease drastically as the attack strength increases; (b) the estimation error in transmit-power, i.e., $|P_o - \hat{P}_o|$, varies with the attack-detection threshold; (c) measurement residual remains below the attack-detection threshold $\eta = 7.04$.



(a) Number of valid sensing reports (b) Estimation error (c) Measurement residual

Fig. 9. **Tradeoff in selecting the attack detection threshold** (η) **with** $N_a = 5$: Too small a value of η will cause the estimation error to suffer from *over-filtering*, whereas too large a value will cause *under-filtering*. (a) The number of sensing samples that pass the detector increases with η; (b) there exists an optimal η that minimizes the estimation error (i.e., $\eta = 3$); and (c) the measurement residual ($\|\tilde{\mathbf{e}}\|$) increases drastically beyond a certain threshold value. The attack strength is assumed to be fixed at $\Delta = 5$ dB.

relatively high for weak attack strengths, i.e., $\Delta = 1$ dB, such a small deviation in sensing reports makes only a negligible impact on state estimation, and can thus be ignored. The figure also shows that `IRIS` suffers lesser mis-detection with stronger attack strengths.

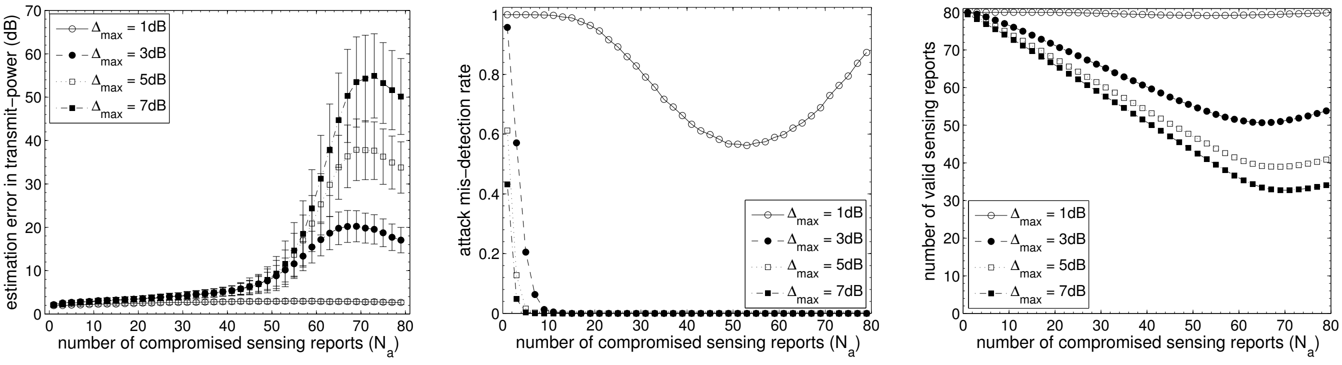
2) **Case I** *(A Single Compromised Sensing Report):* We now demonstrate `IRIS`'s effectiveness in pinpointing the attacker. In the simulations, we assume a single manipulated sensing report with various attack strengths ranging from 0 to 5 dB. Here we assume a fixed attack detection threshold at $\eta = 7.04$, which corresponds to the attack false-alarm rate of $P^a_{FA} = 0.1$ (see Eq. (15) in Section V).

Fig. 8(a) depicts the *attack mis-detection* and *attack mis-identification* rates. The attack mis-detection rate is defined as the probability that the measurement residual remains below the detection threshold η, and thus the attack detection process is not triggered even in the presence of manipulated sensing report(s). The attack mis-identification rate is defined as the average fraction of instances that `IRIS` mistakenly identifies and filters out a legitimate sensing report. The figure shows that `IRIS` accurately detects and pinpoints the manipulated sensing report with as weak a attack strength as 2 dB. When the attack strength is lower than 2 dB, the measurement residual will remain below the threshold, as shown in Fig. 8(c), and thus the attack detection will not be triggered. As the attack strength increases, however, the measurement residual is highly likely to exceed the attack-detection threshold, and thus the mis-detection rate is close to 0.

Fig. 8(b) plots the average as well as $\pm 0.25\,\sigma$ of the error in the estimation of transmit-power. It shows that `IRIS` maintains a small error under various attack strengths. The estimation error is maximized when the attack strength is around 1 dB at which the measurement residual hits the attack-detection threshold η, as shown in Fig. 8(c). When the attack strength is weak (i.e., $<$ 1 dB), the impact of the attack is negligible even though manipulated sensing reports can evade the attack detector. On the other hand, `IRIS` can easily detect the sensing reports with large deviations (i.e., $>$ 1 dB), and thus, the estimation error remains low.

Fig. 8(c) shows that the measurement residual has a pattern similar to that of estimation error, as shown in Fig. 8(b). Moreover, the measurement residual is always below the detection threshold $\eta = 7.04$ (the dashed line), which was set to achieve a desired attack false-alarm rate of $P^a_{FA} = 0.1$.

3) **Case II** *(Multiple Compromised Sensing Reports):* Next, we examine `IRIS`'s attack-detection performance when multiple sensing reports are manipulated (e.g., due to PUEA or SSDF attacks). Specifically, we study the impact of attack-



(a) Attack mis-detection rate

(b) Estimation error in transmit-power

(c) Number of valid sensing samples

Fig. 10. **Impact of number of manipulated sensing reports on estimation performance**: (a) The attack mis-detection rate decreases as the number of manipulated sensing reports increases and (b) the number of valid sensing reports that passed the detector also decreases. (c) The estimation error in transmit-power is maintained small regardless of the attack strengths, until an almost half of the sensing reports are contaminated.

detection threshold η on the attack-detection and estimation error performance. In the simulations, we assume that attack strength is fixed at $\Delta = 5\,$dB. Fig. 9(a) shows the tradeoff in the design of the detection threshold, i.e., when the detection threshold is set too low (i.e., $\eta < 2$), IRIS tends to *over-filter* sensing reports; otherwise, it tends to *under-filter* sensing reports. Fig. 9(b) shows that such over- or under-filtering degrades estimation error performance due to either the lack of enough samples or the presence of manipulated sensing reports in the state estimation. The figure shows that the optimal detection threshold that minimizes the average estimation error is $\eta^* \approx 2$. Fig. 9(c) shows the sharp increase in measurement residual as the attack-detection threshold exceeds 2 where IRIS starts to over-filtering the sensing reports.

Although we found the optimal attack detection threshold η^* for the specific case of $N_a = 5$, the optimal threshold may vary with the attack scenario, such as the number of compromised sensing reports or attack strengths. We observed, however, that the attack detection performance does not critically depend on the above factors, and $\eta = 2$ works reasonably well in various attack scenarios, as we will observe in the next subsection. Henceforth we set $\eta = 2$.

D. Impact of Attack Population

We now consider IRIS's attack-tolerance by studying the impact of number of compromised sensors on state estimation performance. We assume that the BS employs $N = 80$ cooperative sensors, and the number of manipulated sensing reports ranges from 1 to 79. Each compromised sensor launches attacks with the strength uniformly distributed in $[0, \Delta_{max}]\,$dB, i.e., $a_i = \texttt{rand}() \cdot \Delta_{max} \;\; \forall i \in \mathcal{M}$, where $\Delta_{max} \in \{1, 3, 5, 7\}\,$(dB). We set the attack-detection threshold at $\eta = 2$, based on the observation made in Fig. 9.

Fig. 10(a) plots the average as well as $\pm 0.25\,\sigma$ of the estimation error in transmit-power. It shows that the error is kept small, i.e., less than 5 dB, until almost half of the sensing reports are manipulated, regardless of the attack strength. When the number of manipulated sensing reports exceeds 50 % of the entire set of sensing samples, the estimation error behaves differently according to the attack strength, indicating that the attack is effective only when the majority of the sensors are compromised.

Fig. 10(b) shows that IRIS suffers from a large attack mis-detection rate under weak attacks, i.e., $\Delta_{max} = 1\,$dB. However, as shown in Fig. 10(a), such a weak attack fails to cause a significant estimation error. On the other hand, as the attack becomes stronger, the attack mis-detection rate decreases drastically as the number of compromised sensors increases. For example, when $\Delta_{max} = 5\,$dB, the attack mis-detection rate drops below 10 % as the number of compromised sensing reports exceeds $N_a = 4$, i.e., $4/80 = 5\,\%$ of the total sensing reports.

Fig. 10(c) plots the average number of sensing reports that passed the detector. It decreases initially as the number of manipulated sensing reports increases, but it starts to increase when a significant fraction of the sensing reports are compromised, because identifying contaminated sensing reports, which constitutes the majority, becomes very difficult.

E. Performance Comparison

Fig. 11 compares the error in estimating transmit-power among three schemes: (i) IRIS *with the attack detector*, (ii) IRIS *without the attack detector*, and (iii) statistics-based heuristic method in [14] (denoted as Outlier in Fig. 11). In Outlier, the BS filters out the sensing reports that fall outside the range $[e_1 - \delta \cdot e_{iqr}, e_3 + \delta \cdot e_{iqr}]$ where e_1 (e_3) represents the first (third) quartile of the sensing reports, and $e_{iqr} = e_3 - e_1$ is the interquartile range. δ is a control knob that adjusts the aggressiveness of the detector, which is a design parameter. In the simulations, we assume $N = 80$, $N_a = 40$ and the attack strength is in the range of $\Delta_{max} \in [1, 20]\,$(dB). In the absence of the attack detector, the estimation error almost linearly increases with the attack strength. On the other hand, the performance of Outlier depends critically on δ. For example, when $\delta = 0.5$, the performance suffers from large estimation error due to the over-filtering of sensing reports. When $\delta = 1$, the error decreases, but the estimation performs worse than IRIS without the attack detector. This is because the Outlier method mistakenly filters out too many legitimate sensing reports since its design does not take into account the inherent heterogeneity in sensing reports due to their geographical locations. By contrast, IRIS with the attack detector maintains small estimation error, thanks to its ability to accurately detect abnormal sensing reports based on the consistency check we introduced in Section V.

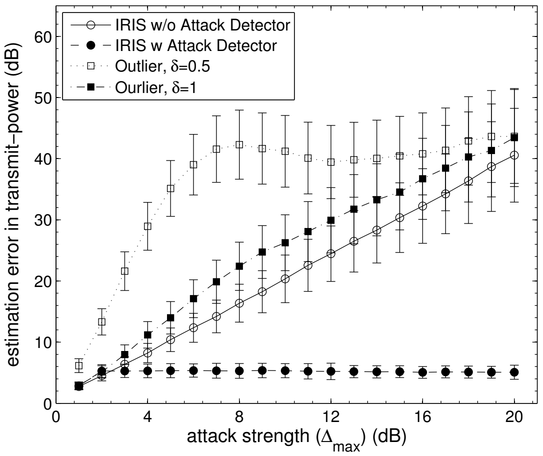


Fig. 11. **Comparison of transmit-power estimation performance**: IRIS with the attack detector maintains a small transmit-power estimation error, outperforming the other detection schemes, i.e., IRIS without the attack detector and method Outlier.

In summary, we can conclude that IRIS is (i) highly robust even when a significant fraction of the sensors are compromised, and (ii) highly accurate in estimating system states even in challenging attack scenarios. These two characteristics make the incumbent detection rule based on estimated transmit-power very attractive.

VII. Conclusion

In this paper, we proposed a robust spectrum sensing framework, called IRIS, to enable efficient opportunistic spectrum access in cognitive radio networks. The key insight behind IRIS is that the received primary signal strengths are mainly governed by the network topology as well as the PHY signal-propagation property that attackers cannot easily compromise. IRIS checks the consistency among the sensing reports with estimated transmit-power and path-loss exponent to safeguard the cooperative sensing. By checking the consistency among the sensing reports, IRIS accurately detects the presence of abnormal sensing reports, and then pinpoints and removes them. As a result, IRIS greatly reduces the impact of attacks on the effectiveness of cooperative sensing. Using in-depth analysis and simulation, we demonstrated IRIS's high attack-tolerance even under very challenging scenarios, such as when a significant fraction of sensors are compromised. IRIS can be readily implemented and deployed in infrastructure-based CRNs, such as IEEE 802.22 WRANs, while incurring only small computation and communication overheads.

References

[1] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh, "White Space Networking with Wi-Fi like Connectivity," in *Proc. ACM SIGCOMM*, Aug 2009.

[2] Mobile Broadband Capacity Constraints and the Need for Optimization, http://www.rasavy.com/Articles/2010_02_Rysavy_Mobile_Broadband_Capacity_Constraints.pdf.

[3] FCC, "Second Memorandum Opinion and Order," FCC 10-174, Sep 2010.

[4] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative Sensing among Cognitive Radios," in *Proc. IEEE ICC*, June 2006.

[5] A. W. Min and K. G. Shin, "An Optimal Sensing Framework Based on Spatial RSS-profile in Cognitive Radio Networks," in *Proc. IEEE SECON*, June 2009.

[6] A. Ghasemi and E. S. Sousa, "Collaborative Spectrum Sensing for Opportunistic Access in Fading Environments," in *Proc. IEEE DySPAN*, Nov 2005.

[7] E. Visotsky, S. Kuffner, and R. Peterson, "On Collaborative Detection of TV Transmissions in Support of Dynamic Spectrum Sharing," in *Proc. IEEE DySPAN*, Nov 2005.

[8] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar, "IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radio," *J. Commun.*, vol. 1, no. 1, pp. 38–47, April 2006.

[9] Y. Selén, H. Tullberg, and J. Kronander, "Sensor Selection for Cooperative Spectrum Sensing," in *Proc. IEEE DySPAN*, Oct 2008.

[10] USRP: Universal Software Radio Peripheral, http://www.ettus.com.

[11] K. Tan *et al.*, "Sora: High Performance Software Radio Using General Purpose Multi-core Processors," in *Proc. USENIX NSDI*, April 2009.

[12] R. Chen, J.-M. Park, and K. Bian, "Robust Distributed Spectrum Sensing in Cognitive Radio Networks," in *Proc. IEEE INFOCOM*, April 2008.

[13] R. Chen, J.-M. Park, and J. H. Reed, "Defense against Primary User Emulation Attacks in Cognitive Radio Networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 25–37, Jan 2008.

[14] P. Kaligineedi, M. Khabbazian, and V. K. Bharava, "Secure Cooperative Sensing Techniques for Cognitive Radio Systems," in *Proc. IEEE ICC*, May 2008.

[15] A. W. Min, K. G. Shin, and X. Hu, "Attack-Tolerant Distributed Sensing for Dynamic Spectrum Access Networks," in *Proc. IEEE ICNP*, Oct 2009.

[16] O. Fatemieh, R. Chandra, and C. A. Gunter, "Secure Collaborative Sensing for Crowdsourcing Spectrum Data in White Space Networks," in *Proc. IEEE DySPAN*, April 2010.

[17] W. Wang, H. Li, Y. Sun, and Z. Han, "CatchIt: Detect Malicious Nodes in Collaborative Spectrum Sensing," in *Proc. IEEE Globecom*, Nov 2009.

[18] H. Li and Z. Han, "Catch Me if You Can: An Abnormality Detection Approach for Collaborative Spectrum Sensing in Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3554–3565, Nov 2010.

[19] S. Liu, Y. Chen, W. Trappe, and L. J. Greenstein, "ALDO: An Anomaly Detection Framework for Dynamic Spectrum Access Networks," in *Proc. IEEE INFOCOM*, April 2009.

[20] Y. Liu, P. Ning, and H. Dai, "Authenticating Primary Users' Signals in Cognitive Radio Networks via Integrated Cryptographic and Wireless Link Signatures," in *Proc. IEEE Symposium on Security and Privacy*, May 2010.

[21] O. Fatemieh, A. Farhadi, R. Chandra, and C. A. Gunter, "Using Classification to Protect the Integrity of Spectrum Measurements in White Space Networks," in *Proc. NDSS*, Feb 2011.

[22] W. Zhang, S. K. Das, and Y. Liu, "A Trust Based Framework for Secure Data Aggregation in Wireless Sensor Networks," in *Proc. IEEE SECON*, Sep 2006.

[23] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks," in *Proc. ACM MobiHoc*, May 2006.

[24] F. Liu, X. Cheng, and D. Chen, "Insider Attacker Detection in Wireless Sensor Networks," in *Proc. IEEE INFOCOM*, May 2007.

[25] L. Lazos and R. Poovendran, "SeRLoc: Secure Range-Independent Localization for Wireless Sensor Networks," in *Proc. ACM WiSe*, Oct 2004.

[26] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch, "Geo-Location Database Techniques for Incumbent Protection in the TV White Space," in *Proc. IEEE DySPAN*, Oct 2008.

[27] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl, "SenseLess: A Database Driven White Spaces Network," MSR, Tech. Rep. MSR-TR-2010-127, Sep 2010.

[28] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the Energy Detection of Unknown Signals over Fading Channels," in *Proc. IEEE ICC*, May 2003.

[29] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, April 2008.

[30] S. Shellhammer, S. Shankar, R. Tandra, and J. Tomcik, "Performance of Power Detector Sensors of DTV Signals in IEEE 802.22 WRANs," in *Proc. ACM TAPAS*, Aug 2006.

[31] T. K. Moon and W. C. Stirling, *Mathmatical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.

[32] Y. Liu, P. Ning, and M. K. Reiter, "False Data Injection Attacks against State Estimation in Electronic Power Grids," in *Proc. ACM CCS*, Nov 2009.

[33] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad Data Analysis for Power System State Estimation," *IEEE Trans. Power App. Syst.*, vol. 94, no. 2, pp. 329–337, Mar 1975.