This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2009 proceedings.

ALDO: An Anomaly Detection Framework for Dynamic Spectrum Access Networks

Song Liu*, Yingying Chen†, Wade Trappe* and Larry J. Greenstein*
*WINLAB, Rutgers University
North Brunswick, New Jersey 08902
Email: {song, trappe, ljg}@winlab.rutgers.edu
†Stevens Institute of Technology
Hoboken, New Jersey 07030
Email: yingying.chen@stevens.edu

***Abstract*—Dynamic spectrum access has been proposed as a means to share scarce radio resources, and requires devices to follow protocols that use resources in a proper, disciplined manner. For a cognitive radio network to achieve this goal, spectrum policies and the ability to enforce them are necessary. Detection of an unauthorized (anomalous) usage is one of the critical issues in spectrum etiquette enforcement. In this paper, we present a network structure for dynamic spectrum access and formulate the anomalous usage detection problem using statistical significance testing. The detection problem is classified into two subproblems. For the case where no authorized signal is present, we describe the existing cooperative sensing schemes and investigate the impact of signal path loss on their performance. For the case where an authorized signal is present, we propose three methods that detect anomalous transmissions by making use of the characteristics of radio propagation. Analytical models are formulated for two special cases and, due to the intractability of the general problem, we present an algorithm using machine learning techniques to solve the general case. Our simulation results show that our approaches can effectively detect unauthorized spectrum usage with high detection rate and low false positive rate.**

I. Introduction

The openness of the lower-layer protocol stacks renders cognitive radios (CR) an appealing solution to dynamic spectrum access (DSA). Its open nature will increase the flexibility of spectrum utilization and promote spectrally-efficient communication. Nevertheless, due to the exposure of the protocol stacks to the public, CR platforms can become a tempting target for adversaries or irresponsible secondary users [1]. A misuse of a CR can significantly compromise the benefits of DSA and threaten the privileges of incumbent users. Therefore, having the ability to enforce spectrum etiquettes is critical to effectiveness and correctness of a DSA system.

Identification of a malicious or reckless spectrum usage is an essential component of etiquette enforcement functions. This is basically a problem of distinguishing bad (unauthorized) transmissions from good (authorized) ones. While sophisticated signal processing techniques have been designed for detecting a desired signal from interference [2], [3], they are of little help in this new paradigm of spectrum access. In many DSA systems (e.g., spectrum leasing), there can be a heterogeneous collection of authorized users and it is impractical to enumerate all of their signal structures. Even if the authorized signal is known (e.g., TV signals in IEEE 802.22), unauthorized users can disguise themselves by emulating authorized signals [4]. Therefore, an effective detection mechanism should not rely on programmable features, such as signal patterns. Fortunately, there is one aspect of the problem that cannot be easily modified—the propagation channel. This motivates us to pursue a reliable detection approach by making use of the characteristics of radio propagation. Specifically, in this paper, our detection will be based on the measurement of received signal powers at a collection of monitoring stations.

Distinguishing unknown signals based on power (or energy) measurements is not a trivial issue. We need to first define the objective of our detection task. In the context of anomalous detection, our goal is to identify unauthorized usage. Although it is hard to find a comprehensive description about anomalous behaviors, we should at least clearly describe the characteristics of the power measurements in the normal usage case. In this work, we limit the normal usage to a single transmitter, static or mobile, that uses each portion of the spectrum resource (e.g., a channel). That is, anomalous usage will correspond to the presence of unauthorized signals in a radio resource that is either idle or being used by an authorized transmitter. Although detecting unknown signals in noise has been well studied [5], [6], little research exists that addresses the problem of detecting whether power measurements are from a single source or multiple sources. In [7], though, the authors proposed to verify the location of the transmitter to detect a false primary user. This requires knowledge of the location of the primary transmitter, which has to be fixed and at a great distance from the sensing area, and thus does not work when the transmitter is mobile, or in more general scenarios.

In this paper, we will investigate the spectrum anomaly detection problem in a broader context, where the authorized transmitter can be mobile within the sensing area. In general, we know that the received signal power (in dB) decays approximately linearly with the logarithmic distance from the source. Hence, we propose three methods that capture the difference between measured power and its estimation by making use of this linear relationship. In particular, the spatially distributed power measurements are obtained via cooperative

978-1-4244-3513-5/09/$25.00 ©2009 IEEE

spectrum sensing in a centralized sensor network. Further, by exploiting the linearity with log-distance inherent in the normal usage case, we formulate our detection problem as a statistical significance test. We derive analytical solutions for the case where additional information (e.g., location) about the authorized transmitter is available and propose a generalized approach using support vector machine techniques.

The rest of the paper is organized as follows. Section II reviews previous research in spectrum sensing. Section III presents our DSA network structure and spectrum access policy. In Section IV, we present the generalized theoretical model for anomalous spectrum usage detection. Section V discusses the problem of detecting unknown signals from noise and the performance of existing cooperative sensing strategies is addressed. We propose solutions and present results for detecting unknown signals from a variable signal strength in Section VI. Section VII concludes our work.

II. Related Work

In the context of dynamic spectrum access, spectrum sensing is an essential function to learn the characteristics of spectrum environment, such as the presence of spectrum holes or current signal transmission modes (i.e., bandwidth or modulation techniques). Depending on the level of a-priori knowledge about primary signal characteristics, signal processing techniques used for spectrum sensing include: matched filter, energy detector and cyclostationary feature detector [8], [9]. Due to the simplicity and ability to detect unknown signals, the energy detector has extensively been studied since early days of radar [5]. In this classic paper, assuming deterministic signals over a flat band-limited Gaussian noise channel, the received signal energy was modeled via chi-square statistics and both the probability of detection and the probability of false alarm were derived. The detection problem was recently extended to a variety of fading environments [6], including Rayleigh, Rican and Nakagami channels.

To combat the uncertainties caused by fading channels and other spectrum irregularities, cooperative sensing is regarded as an effective solution, which can be implemented either in a centralized or in a distributed manner. Based on the systematic study in [10], [11], recent progress has been made both at the signal level and at the system level [12]–[16] in the context of cognitive radios. In [15], the licensed TV signal power received among all secondary users was modeled as being correlated Gaussian, and the optimal decision space was obtained based on the Neyman-Pearson criterion [3]. The hard decision combining strategy, "k-out-of-n rules", was also investigated. Although the analysis showed that soft decision combining can achieve a lower miss detection probability than hard cooperation, [14], [17] found that this gain is limited by physical noise uncertainty. Based on the energy detection model developed by [6], [12] investigated the asymptotic performance of the cooperation under independent fading channels and uncertain noise. The related study under correlated log-normal shadowing was given in [18]. Most of the above work addressed detection performance by assuming the same distribution for the primary signal at all sensors. This is generally true when the primary transmitter is a great distance away from the sensing area, such as a TV broadcasting tower. In such a case, the link distance between the transmitter and all secondary users are approximately equal. Relay-based cooperation is proposed in [16], which takes into account the effect of different link distances while assuming a known location of the transmitter.

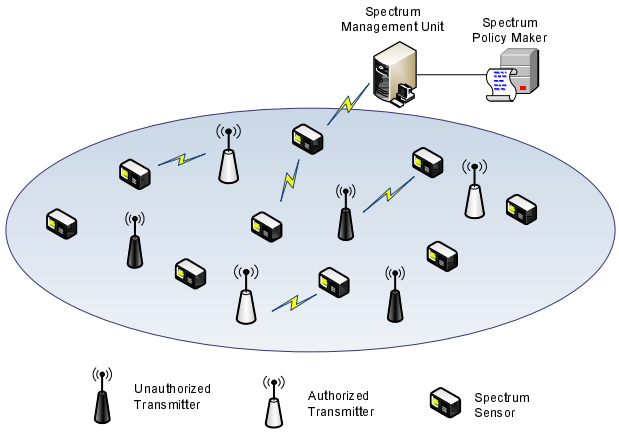


Fig. 1. Spectrum sensing infrastructure.

Our work is distinguished by the fact that we utilize the characteristics of radio propagation for anomaly detection in spectrum usage. Specifically, the path loss effect will play a critical role in our studies, which has been ignored in most of the previous work. Also, we provided a generalized solution in detecting unauthorized spectrum usage using machine learning techniques.

III. A System Model of Dynamic Spectrum Access

A. DSA Network Structure and Access Policy

We consider a dynamic spectrum access network as illustrated in Fig. 1, where licensed (i.e., primary) and unlicensed (i.e., secondary) transmitters are scattered in an area filled with auxiliary spectrum sensors. In the paradigm of DSA, secondary users should access spectrum resources without interfering with incumbent users. Thus, a spectrum access policy is necessary to enforce the usage etiquettes. The policies are defined by spectrum policy makers and are broadcast by a management unit. For spectrum agility, the policies can change dynamically and users should be able to interpret them without human intervention. Interpreted languages, such as XG Policy Language (XGPL), have been proposed to formalize the policies [1]. To enforce spectrum policies, the sensor network is responsible for collecting spectrum usage data and reporting them to the spectrum management unit. The management unit applies appropriate detection rules to identify anomalous usage and performs localization to locate anomalous transmitters.

In this work, we propose an anomaly detection framework for enforcing spectrum etiquettes in outdoor DSA networks. As illustrated in Fig. 2, the entire network (e.g., that covers a

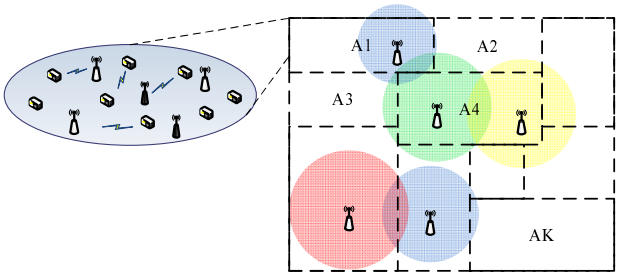


Fig. 2. The DSA network is divided into several zones. Sensors are sensing the spectrum usage in the zone where they reside. Shaded circles denote the signal coverage of authorized transmitters.

state) is divided into several zones (e.g., counties). Analogous to cellular systems, we assume that the spectrum resources are assigned to each zone so that there is no inter-zone interference at any time. That is, users in adjacent zones must use different frequencies at the same time. Also, users in the same zone will be assigned to different frequency bands at any time so that there is no intra-zone interference either.

An example of the spectrum policy can be expressed as:

- *User U_m is allowed to use frequency band W_i from time T_1 to T_2, as long as the power levels do not go above λ dBm in zone A_k.*

To enforce a spectrum policy, the sensors in each zone will periodically measure spectrum levels at protected bands and report to the management unit. For instance, in Fig. 2 suppose there are N sensors in zone A_1. A measurement will produce a vector $\mathbf{Y}(f,t) = (Y_1, Y_2, \ldots, Y_N)$, where Y_n is the reading of the n-th sensor[1]. For convenience, we hereafter suppress the argument (f,t). Without specifying the structure of authorized and unauthorized signals, in this work we shall apply energy detectors. A general-purpose energy detector consists of a low-pass filter, Nyquist sampling A/D converter, square-law device and integrator, as depicted in Fig. 3.

B. Anomalous Spectrum Usage

According to our spectrum access policy, an anomalous usage occurs in the following conditions:

- The transmission power of an authorized user is above the specified power level. This will cause interference with users in neighboring zones;
- A transmitter is using a portion of the spectrum which is supposed to be idle. In this case, the transmitter is an unauthorized user;
- A transmitter is using a portion of the spectrum that is granted to another authorized transmitter.

Since our anomalous detector is based on energy measurement, we may use an energy threshold to determine the unauthorized spectrum usage in the first two cases. The challenge to detect anomalous behaviors in the second case is that due to noise we may incorrectly declare an anomalous usage when the spectrum is indeed idle (i.e., a false alarm). From the viewpoint

[1]Throughout the paper, we denote received signal power by P, and denote an energy measurement by Y that consists of signal and noise.

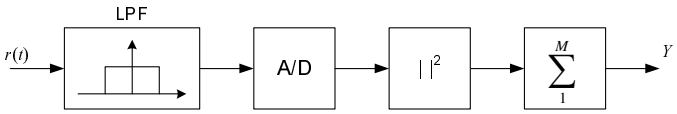


Fig. 3. A square-law based energy detector.

of detection analysis, the first case can be regarded as a subset of the second where the probability of the false alarm is not taken into consideration. Therefore, we will focus on the last two cases in our investigation.

C. Propagation Model

We assume the received signal power at a sensor is given by the following propagation model [19]:

$$P_n = P_0 - 10\gamma \log_{10}(d_n/d_0) + S_n \text{ (dB)}, n = 1, \ldots, N, \quad (1)$$

where N is the number of sensors. In this model, P_0 is the signal strength, in dB, measured at the reference distance d_0 ($d_0 = 1$ m throughout this study[2]); d_n is the distance between the source and the n-th sensor; γ is the path loss exponent with a typical value of $3 \sim 6$ in suburban environments; and S_n accounts for the shadow fading, which is widely modeled as a Gaussian random process across space with zero mean and standard deviation σ_{dB}. In addition, $\mathbf{S} = (S_1, S_2, \ldots, S_N)^T$ follows a correlated multivariate Gaussian distribution, $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Here, we neglect the effect of multipath, which can be averaged out by sufficiently wideband measurements. The model parameters, γ and $\mathbf{\Sigma}$, can be estimated through some training measurements.

IV. Modeling Anomalous Detection Using Significance Testing

In general, we only have the information in the normal situation and thus the detection of anomalous usage can be formulated as a statistical significance testing problem. We define the received signal at each sensor as:

$$\mathcal{H}_0 : r(t) = s(t) + w(t), \quad \text{normal usage}, \quad (2a)$$
$$\mathcal{H}_1 : r(t) = s(t) + x(t) + w(t), \quad \text{anomalous usage}. \quad (2b)$$

where $s(t)$ is the signal from an authorized transmitters complying with the spectrum policy, $x(t)$ is an unknown unauthorized signal, and $w(t)$ is noise that we assume to be additive and white Gaussian with zero mean. Note that the three terms are all complex. The normal spectrum usage is defined as the null hypothesis $\mathcal{H}_0$.

A significance testing problem consists of the following key components:

- Test statistic $\mathbf{T}$: a measure of the observed data.
- Acceptance region Ω: if $\mathbf{T} \in \Omega$, we accept the null hypothesis $\mathcal{H}_0$.
- Significance level α: the probability of incorrectly rejecting the null hypothesis, i.e., the probability of false alarm.

[2]In general, this propagation model holds when the link distance $d_n > d_0$, so we assume a sensor is always located at the distance greater than d_0 from a transmitter.

In our detection problem, the observed data is a series of energy measurements, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$. The output of the energy detector at the n-th sensor is given by

$$Y_n = \sum_{i=1}^{M} |r_i|^2 = \sum_{i=1}^{M} [\mathrm{Re}(r_i)^2 + \mathrm{Im}(r_i)^2], \tag{3}$$

where M is the number of samples. Without loss of generality, we assume the real and imaginary components of w_i both follow $\mathcal{N}(0,1)$, so the noise power is 2. The strengths of the received signals s_i and x_i follow the propagation model (1) and we assume they add noncoherently.

For different statistics of $\mathbf{Y}$ in what follows, we will define $\mathbf{T}$ and Ω so that, for a specified false alarm probability α, $Prob(\mathbf{T} \notin \Omega|\mathcal{H}_0) \leq \alpha$, where $\mathbf{T}$ not in Ω declares the presence of the anomalous behaviors in the network.

V. Detecting Unauthorized Signals in Noise

In this section, we describe our detection methods when an unauthorized transmitter is using a portion of the spectrum, which is supposed to be idle.

The detection problem in this case becomes detecting unauthorized signals in noise, which is similar to the classic problem of detecting primary signals, and several cooperative methods have been proposed [6], [12], [20]. Although their performance has been well studied under multipath fading and shadowing, little research has considered the effect of path loss. Since in our study the unauthorized transmitter is located within the sensing area, the variation of link distances from the transmitter to the sensors should not be ignored. Thus, we will assess the performance of some popular cooperation methods under the effect of path loss.

When there is no authorized signal (i.e., $s(t) = 0$), the energy detector output follows a chi-square distribution under the normal situation (no unauthorized transmitter) and a non-central chi-square distribution with the presence of the unauthorized users. Their probability density functions (PDF) are [6]

$$f_Y(y|\mathcal{H}_0) = \frac{1}{2^M \Gamma(M)} y^{M-1} e^{-y/2}, \tag{4a}$$

$$f_Y(y|\mathcal{H}_1) = \frac{1}{2} e^{-(\lambda+y)/2} \left(\frac{y}{\lambda}\right)^{M/2-1/2} I_{M-1}(\sqrt{\lambda y}), \tag{4b}$$

where $I_\alpha(x)$ denotes the modified Bessel function of the first kind, $\lambda = \sum_{i=1}^{M} |x_i|^2$ and $\lambda/(2M)$ is the average signal-to-noise ratio (SNR) over one measurement duration.

A. Equal Gain Combining (EGC)

A decision rule is usually called soft-decision combining when it is based on continuous-valued samples or accurate digitized versions thereof. In fading channels, equal gain combining (EGC) is known to have a near-optimal performance without channel estimation [20]. By summing over energy measurements from all sensors N, the test statistic is

$$Y_\Sigma = \sum_{n=1}^{N} Y_n. \tag{5}$$

When Y_n contains only i.i.d. Gaussian noise, $Y_\Sigma \sim \mathcal{X}^2$. The acceptance region is $\{\Omega : Y_\Sigma < T\}$, where T is the threshold for energy detection. Then, the false alarm probability Q_F can be expressed as

$$Q_F = Prob(Y_\Sigma > T|\mathcal{H}_0) = \frac{\Gamma(NM, T/2)}{\Gamma(NM)}, \tag{6}$$

where $\Gamma(a)$ and $\Gamma(a,b)$ are gamma and incomplete gamma function, respectively.

B. Hard-Decision Combining, k-out-of-n

Although the optimal decision can be approached by perfect knowledge of the measurements, it requires a high communication overhead in order to transmit the large number of bits needed for the required precision. In practice, hard-decision combining is more appealing in cases where the sensors can only report their 1-bit individual decisions. Each sensor performs detection only based on the power measured by itself and the final decision is based on a combination of all individual ones. The acceptance region for each individual test is $\{\Omega : y_i < T\}$ for $i = 1, \ldots, N$. Assuming the distribution of the measurement is identical, all the sensors apply the same threshold. The false alarm probability at each sensor is

$$P_F = Prob(Y > T|\mathcal{H}_0) = \frac{\Gamma(M, T/2)}{\Gamma(M)}. \tag{7}$$

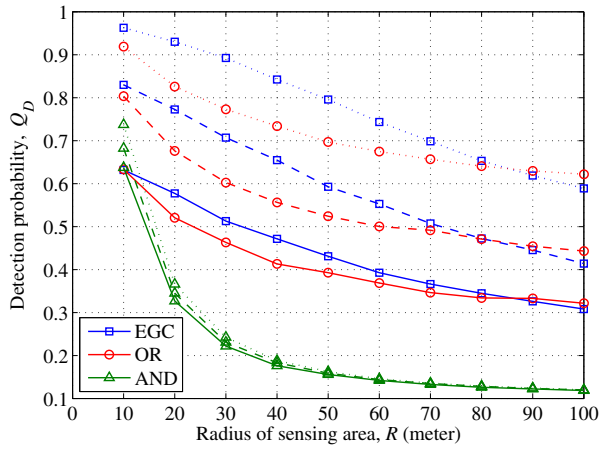
If the measured energy is independent among the sensors, it is known the optimal combining belongs to one of the k-out-of-n rules [12]. That is, if there are no less than K sensors having detected the signal, the system declares the signal presents. The combined false alarm probability is

$$Q_F = \sum_{n=K}^{N} \binom{N}{n} P_F^n (1 - P_F)^{N-n}. \tag{8}$$

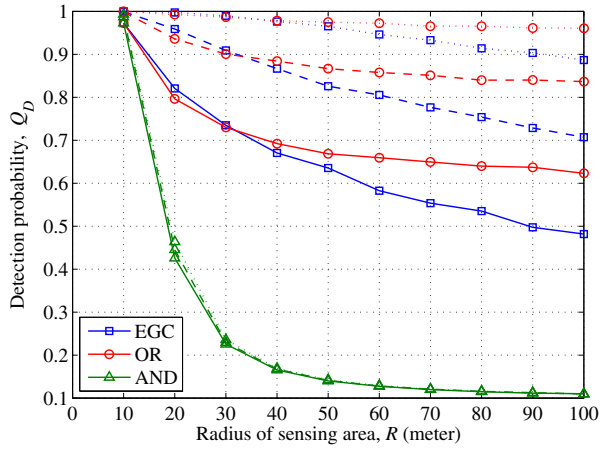
As a point of practical interest, the most frequently used combinations are two extremes: 1-out-of-n (a.k.a. OR rule) and n-out-of-n (a.k.a. AND rule).

C. Simulation Evaluation

We study the above two cooperative strategies by simulations in our anomaly detection problem, where an unauthorized radio source resides within the sensing area. Ideally, we assume sensors are uniformly distributed in a circular area centered at the unauthorized transmitter, with the radius R. For the same density of sensors, it can be proved that this ideal deployment yields the highest detection rate. To investigate the effect of path loss, we assume there is no shadowing (i.e., $S_n = 0$ in (1)), which is known to degrade detection performance. Without considering the path loss, previous work has shown that increasing the number of cooperative sensors can improve the detection rate. However, this is no longer true when the link distance between the transmitter and the sensors are significantly different. Fig. 4 presents the detection probability Q_D for the OR, AND and EGC, respectively, under different radius R. We fix the density of sensors so that there are 1, 2 and 4 sensors, respectively, when $R = 10$ meters.



(a)



(b)

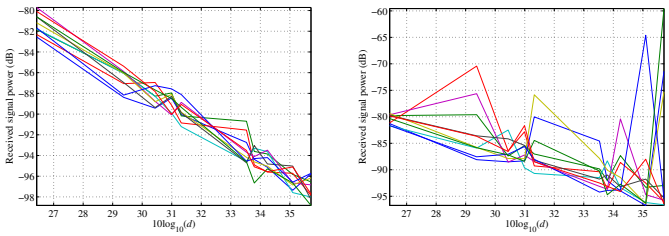
Fig. 4. The detection probability vs. the radius of sensing area, R, where (a) the path loss $\gamma = 2$ and the SNR at the reference distance is 15 dB; and (b) the path loss $\gamma = 4$ and the SNR at the reference distance is 40 dB. The false alarm probability $Q_F = 0.1$. The number of samples $M = 8$. The density is such that, when $R = 10$ m, $N = 1$ (solid curves), 2 (dashed curves), and 4 (dotted curves), respectively.

By increasing R, more sensors will collaborate on the same detection task, whereas the detection rate decreases. Intuitively, the path loss effect will render a measurement, which is closer to the source, more reliable than a measurement that is far away. Thus, a cooperation scheme that takes all measurements equally may no longer be maximally efficient. For the same reason, we see that the OR rule outperforms the EGC in an environment with a large path loss (i.e., $\gamma = 4$), although the latter is near-optimal without considering the path loss effect.

There can be multiple unauthorized transmitters simultaneously. We argue that, however, the received power from a single source will dominate at the sensors close by. Therefore, we can make representative conclusions by examining a single source case.

VI. Detecting Unauthorized Signals with an Authorized Signal Present

In general, it is impractical to apply a threshold-based detection method when the authorized signal is present. The



(a) $\mathcal{H}_0$ (b) $\mathcal{H}_1$

Fig. 5. Received power vs. logarithmic distance between an authorized transmitter and $N = 10$ sensors. In an 100-meter$\times$100-meter area, the sensors are uniformly located. Path loss exponent $\gamma = 3.5$ and $\sigma_{dB} = 1$ dB in (1). The authorized transmitter is located at the center. In the $\mathcal{H}_1$ case, one unauthorized transmitter is randomly located. Both the transmitters use the same transmission power in this example.

received signal strength can be time-variant because of several effects, such as power control and transmitter mobility. This power variation can render the estimation useless. In this section, we investigate the problem of detecting the presence of unknown signals plus an authorized signal of variable strength. Based on propagation channel characteristics and some knowledge about the authorized signal, we propose three detection methods.

A. Linearity Check-given-Location (LCL)

Following the discussion in Section III, there should be only one authorized transmitter at any time and for any frequency. Thus, the detection problem becomes distinguishing between single and multiple transmissions in the same spectral resource. To do this, we need a decision statistic that captures the characteristics of the radiation power in the case of a single transmission. The propagation model (1) shows that the received signal power (in dB) is a linear function of the log-distance (i.e., $\log_{10}(d)$) plus a random term in the case of a single source[3]. This is not true when the received power consists of signal strengths from multiple transmitters. As depicted in Fig. 5, where we measure the received signal power at 10 sensors for 10 distinct time points, the power from a single transmitter shows distinct linear decay with the log-distance, whereas the power from two transmitters does not present the similar pattern. Thus, by examining the linearity of the power measurements with log-distance, we may distinguish the case of a single transmission (i.e., normal usage) from the case of multiple overlapped transmissions (i.e., anomalous usage). Note that we do not specify the direction of a propagation link, as depicted in the model (1). Therefore, the authorized transmitter is assumed to use omnidirectional antennas.

Further, the distance d between the transmitter and a sensor can be obtained in two ways: (a) the authorized transmitter periodically announces its location, using a signal format that is decodable at the sensors; or (b) the sensors cooperatively estimate the transmitter location based on measured power. In either case, a sensor knows its own location.

[3]In the following study, we neglect the noise term in the energy measurements, assuming sufficiently large number of samples.

Given the distance d, the remaining unknown parameters are the reference power P_0 and the path loss exponent γ. In general, it is hard to obtain their accurate values, so we use linear least squares to estimate them.

Let $\mathbf{P} = (P_1, \ldots, P_N)^T$ be the vector of received power at all the sensors, and

$$\mathbf{A} = \begin{bmatrix} 1 & -10\log_{10}(d_1/d_0) \\ \vdots & \vdots \\ 1 & -10\log_{10}(d_N/d_0) \end{bmatrix}. \quad (9)$$

Then,

$$\begin{aligned} \begin{bmatrix} \hat{P}_0 \\ \hat{\gamma} \end{bmatrix} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P} \\ &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{A}\begin{bmatrix} P_0 \\ \gamma \end{bmatrix} + \mathbf{S}) \\ &= \begin{bmatrix} P_0 \\ \gamma \end{bmatrix} + (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}, \end{aligned} \quad (10)$$

where $\mathbf{S} = (S_1, S_2, \ldots, S_N)^T$ as defined in Section III-C. Further, we define the vector of the estimation error (residuals) $\hat{e}$ as

$$\begin{aligned} \hat{\mathbf{e}} = \mathbf{P} - \hat{\mathbf{P}} &= \mathbf{A}\begin{bmatrix} P_0 \\ \gamma \end{bmatrix} + \mathbf{S} - \mathbf{A}\begin{bmatrix} \hat{P}_0 \\ \hat{\gamma} \end{bmatrix} \\ &= \mathbf{S} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S} \\ &= (\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{S}. \end{aligned} \quad (11)$$

Note that the residuals are independent of the transmission power. Based on the linearity of the propagation model, we postulate that the distribution of the residuals in the normal usage case should differ from that of the anomalous case. Therefore, the residuals $\hat{\mathbf{e}}$ are a viable measure of linearity.

The distribution of the residuals in (11) depends on the distance matrix $\mathbf{A}$ and thus relies on the location of the authorized transmitter. Since the distance d can be either known or estimated by localization, we next tailor our study to these two cases, respectively.

In this part, we assume that the location of the authorized transmitter is *known* under certain conditions. For example, the location is fixed or specified in the policy, or the transmitter periodically reports its location to the spectrum management unit. Further, by assuming $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ (see Section III-C), the residuals $\hat{\mathbf{e}}$ are also Gaussian. Let $\mathbf{M} = (\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)$ in (11). We obtain

$$\hat{\mathbf{e}} \sim \mathcal{N}(0, \mathbf{\Sigma}_e), \quad (12)$$

a multivariate Gaussian with

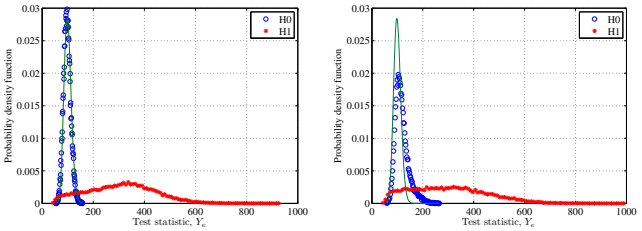
$$\mathbf{\Sigma}_e = E[\hat{\mathbf{e}}\hat{\mathbf{e}}^T] = \mathbf{M}\mathbf{\Sigma}\mathbf{M}. \quad (13)$$

Similar to the significance test in [21], we can define a likelihood-based acceptance region for $\hat{\mathbf{e}}$ as

$$\Omega = \{\hat{\mathbf{e}} : f(\hat{\mathbf{e}}) \geq f_T\}, \quad (14)$$

where

$$f(\hat{\mathbf{e}}) = (2\pi)^{-N/2}|\mathbf{\Sigma}_e|^{-1/2}\exp\left(-\frac{1}{2}\hat{\mathbf{e}}^T\mathbf{\Sigma}_e^{-1}\hat{\mathbf{e}}\right). \quad (15)$$



(a) (b)

Fig. 6. Probability density of the test statistic Y_e by simulation. The solid curve is the PDF of $\mathcal{X}^2(N)$. Path loss exponent $\gamma = 3.5$ and $\sigma_{dB} = 4$ dB. See Section VI-D for other parameters. (a) The location of the authorized transmitter is exactly known; (b) it is estimated by the weighted centroid.

Then, Ω in (14) can also be expressed as

$$\Omega = \{\hat{\mathbf{e}} : \hat{\mathbf{e}}^T\mathbf{\Sigma}_e^{-1}\hat{\mathbf{e}} < T\}. \quad (16)$$

That is, the test statistic $Y_e = \hat{\mathbf{e}}^T\mathbf{\Sigma}_e^{-1}\hat{\mathbf{e}}$. It can be shown that Y_e follows a chi-square distribution with N degrees (for example, see Section V-C in [21]). Thus, the false alarm probability is given by

$$Q_F = \frac{\Gamma(N/2, T/2)}{\Gamma(N/2)}. \quad (17)$$

B. Support Vector Machines (SVM)

In this part, we assume that the location of the authorized transmitter is *unknown*. Based on the measured signal power at sensors, the localization methods that employ received signal strength (RSS) can be used to estimate the source location. A detailed survey of localization approaches can be found in [22]. For the purpose of demonstration, we apply a simple scheme called "weighted centroid" in this study, where the transmitter's location $[x, y]$ is estimated by

$$[x, y] = \frac{\sum_{i=1}^{N} P_i[x_i, y_i]}{\sum_{i=1}^{N} P_i}, \quad (18)$$

and P_i is the linear measured power at the i-th sensor. Due to the environmental noise and the system bias during the measurement, the estimated location and thus the residuals $\hat{e}$ calculated from matrix $\mathbf{A}$ in (11) are affected by measurement errors. The distribution of $\mathbf{A}$ is hard to obtained in practice, which will result in an unknown distribution for the test statistic Y_e, as depicted in Fig. 6. In the more general case, if the statistic of the measurements follows an unknown distribution, it is also prohibitive to quantify the test statistic. Therefore, we propose the application of a machine learning technique, One-class SVM [23], to find an acceptance region.

Support vector machines (SVM) are a set of kernel based learning methods for data classification, which involves a training phase and a testing phase. In one type of problem named supervised learning, each data instance in the training set consists of a target value (*class label*) and several attributes (*features*). The goal of SVM is to produce a model that can predict the target value of data instances in the testing set that only have attributes [24]. In most cases of anomalous or outlier detection, we only have training data from a single class (i.e.,

a normal case). The anomalous class is either not available or severely undersampled. In such a case, unsupervised learning techniques are used to make a description of the training data and to identify new data which resemble the former ones [25]. Estimating the density of the training data can be viewed as an extreme example of this effort [23].

The unsupervised learning method, proposed in [23], estimates a subset Ω of the training data that are sampled from an underlying probability distribution. Asymptotically (i.e., well-sampled training data), with the probability equal to a specified value $\nu \in (0,1)$, the testing data that is drawn from the same distribution lies outside of Ω. This is achieved by the following optimization problem [23],

$$\begin{aligned} \min_{R\in\mathbb{R},\xi\in\mathbb{R}^l,c} \quad & R^2 + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & \|\Phi(\mathbf{v}_i) - c\|^2 \le R^2 + \xi_i, \\ & \xi_i \ge 0, \quad i = 1,\ldots,l, \end{aligned} \tag{19}$$

with the kernel function defined as $K(\mathbf{v}_i,\mathbf{v}_j) = \Phi(\mathbf{v}_i)^T\Phi(\mathbf{v}_j)$. We use the radial basis function (RBF) as the kernel function in our study:

$$K(\mathbf{v}_i,\mathbf{v}_j) = \exp(-\frac{1}{k}\|\mathbf{v}_i - \mathbf{v}_j\|^2), \quad \gamma > 0. \tag{20}$$

Here, $\mathbf{v}_i$ is the vector of features, l is the size of the training set, $\xi = (\xi_1,\ldots,\xi_l)$, and k is the number of attributes in each feature vector. The anomalous detection problem can be viewed as minimizing the radius R of a hypersphere that encloses a subset of the training data. The decision function is given by

$$\mathcal{H}_0: \quad \|\Phi(\mathbf{v}_i) - c\|^2 \le R^2. \tag{21}$$

Given the training data are all from the normal case $\mathcal{H}_0$, the fraction of the excluded data asymptotically equals the false alarm probability.

Therefore, given a set of training data that is well-sampled in the normal usage case, we can use the One-class SVM to find an empirically based acceptance region for a specified false alarm probability. Here, we use the residuals $\hat{\mathbf{e}}$ as the test statistic (i.e., the feature vector $\mathbf{v}$).

C. Calibrating Power (CAL)

Suppose the authorized transmitter periodically sends an "identity" signal that is decodable at the sensors. This signal power can be used as a calibration for the detection. We denote it by $\mathbf{P}^c$. To cope with the time-variant transmission power, we first scale both the calibrating and measured power to the same level. One scaling option is to normalize the power by

$$\tilde{P}_i = P_i - P_N \quad (\text{dB}), \quad i = 1,\ldots,N-1. \tag{22}$$

That is, without loss of generality, we normalize both the calibrating and measured power to 0 dB at the N-th sensor. Using (1) and denoting the normalized power by $\tilde{\mathbf{P}}^c$ and $\tilde{\mathbf{P}}$, respectively, we have

$$\tilde{P}_i^c = -10\gamma\log_{10}(d_i/d_N) + S_i^c - S_N^c \tag{23a}$$

$$\tilde{P}_i = -10\gamma\log_{10}(d_i/d_N) + S_i - S_N. \tag{23b}$$

Then, we obtain new residuals as

$$\hat{e}_i^c = \tilde{P}_i - \tilde{P}_i^c = S_i - S_N - (S_i^c - S_N^c). \tag{24}$$

Since $\mathbf{P}$ and $\mathbf{P}^c$ are measured at different channels, it is safe to assume $(S_i - S_N)$ and $(S_i^c - S_N^c)$ are independent. $\hat{e}_i^c$ is then Gaussian distributed for any $i \neq N$. Let

$$\hat{\mathbf{e}}^c = (\hat{e}_1^c, \hat{e}_2^c, \ldots, \hat{e}_{N-1}^c)^T, \tag{25}$$

so

$$\hat{\mathbf{e}}^c \sim \mathcal{N}(0, \mathbf{\Sigma}_e^c), \tag{26}$$

where (see the Appendix)

$$\mathbf{\Sigma}_e^c = 2\mathbf{B}\mathbf{\Sigma}\mathbf{B}^T \tag{27}$$

and

$$\mathbf{B} = [\mathbf{I},\ -\mathbf{e}] = \begin{bmatrix} 1 & & 0 & -1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & & 1 & -1 \end{bmatrix}_{(N-1)\times N}. \tag{28}$$

Similar to the formulation in Section VI-A, we define the acceptance region $\Omega = \{\hat{\mathbf{e}}^c : Y_c = (\hat{\mathbf{e}}^c)^T(\mathbf{\Sigma}_e^c)^{-1}(\hat{\mathbf{e}}^c) < T\}$ and the false alarm probability is thus

$$Q_F = \frac{\Gamma((N-1)/2, T/2)}{\Gamma((N-1)/2)}. \tag{29}$$

D. Simulation Evaluation

In this section we evaluate the performance of our derived methods, which we call them as LCL (Linearity-Check-given-Location in Section VI-A), SVM (One-class SVM in Section VI-A), and CAL (Calibrating power in Section VI-C), respectively. Their performance were tested in a 100-meter×100-meter square area, where N sensors are uniformly distributed. Both the authorized transmitter and unauthorized transmitters are randomly located in the area. Unless otherwise noted, we assume there is only one unauthorized transmitter and it uses the same transmission power as the authorized user. Also, in these numerical studies, we assume that the variation of measurements across all sensors is i.i.d.

Fig. 7 presents the receiver operating characteristic (ROC) curves for LCL, SVM, and CAL respectively. In this result, we set the path loss $\gamma = 3.5$ and $\sigma_{dB} = 4$ dB that is a typical value for the shadowing in an urban microcell environment [26]. For a false alarm probability 0.1, we observed that all three methods achieve a detection rate above 0.8. With the assisting information from the authorized transmitter, both LCL and CAL can detect about 90% of unauthorized usages. Further, Fig. 8 shows the detection probability under different interference-to-signal ratio (ISR), which is defined by the ratio of transmission power from unauthorized and authorized transmitter. Under the general case that does not have any knowledge about the authorized user, SVM can achieve a highest detection rate of 81% when the transmission ISR is 0 dB. Using additional information from the authorized user, both LCL and CAL can detect more than 90% anomalous usages with a false alarm rate $Q_F = 0.01$ when the unauthorized transmission power is 5 dB higher than the authorized

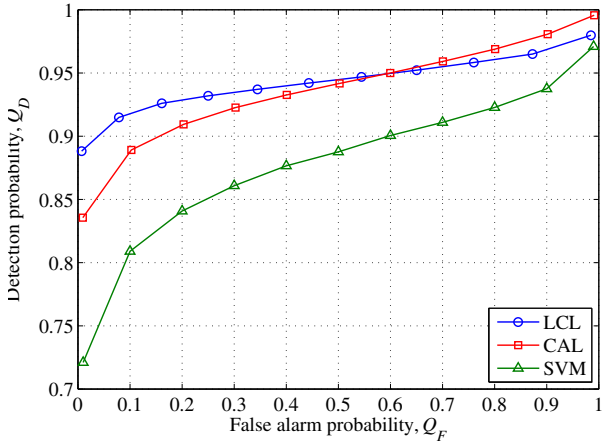


Fig. 7. Receiver operating characteristic (ROC) curves. Path loss exponent $\gamma = 3.5$. $\sigma_{dB} = 4$ dB. $N = 100$ sensors.

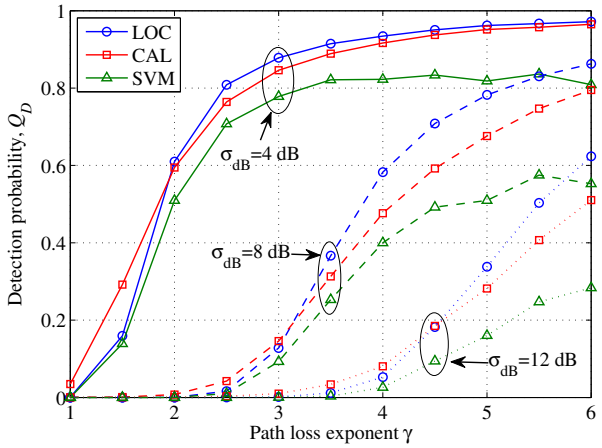


Fig. 9. Detection probability vs. path loss exponent γ. $Q_F = 0.1$. Transmission ISR = 0 dB. $N = 100$ sensors.

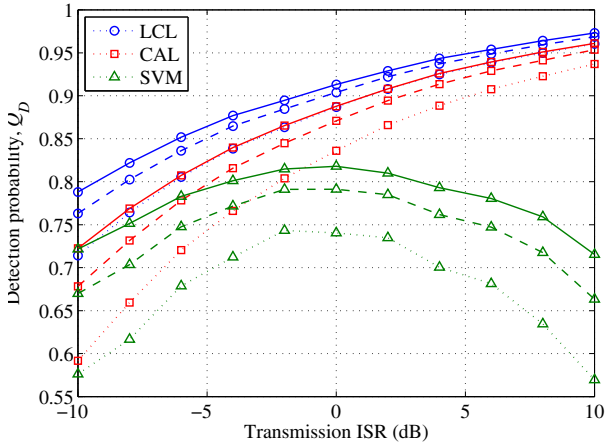


Fig. 8. Detection probability vs. interference-to-signal ratio (ISR). Path loss exponent $\gamma = 3.5$ and $\sigma_{dB} = 4$ dB. $N = 100$ sensors. Solid curves: $Q_F = 0.1$; dash curves: $Q_F = 0.05$; dotted curves: $Q_F = 0.01$.

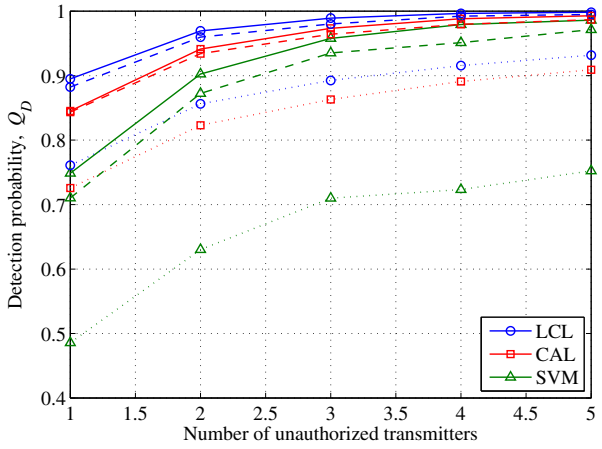


Fig. 10. Detection probability vs. the number of independent unauthorized transmitters, where the total transmission power of the unauthorized users equals the authorized transmission power (i.e., ISR = 0 dB). $\gamma = 3.5$ and $\sigma_{dB} = 4$ dB. $Q_F = 0.1$. Solid curves: $N = 100$; dashed curves: $N = 50$; dotted curves: $N = 20$ sensors.

one. The effects of propagation channels on the detection probability can be seen from Fig. 9. A higher γ results in a higher detection rate for all the methods; and a higher σ_{dB} severely deteriorates performance.

We now address the case of *multiple* unauthorized transmitters. The transmitters are taken here to be non-colluding and independent, so that their powers add noncoherently. Intuitively, more unauthorized users should lead to better detection because the total amount of transmitted power (and the resulting interference power at each sensor) increases. We thus address a more interesting question: If the sum of all transmit powers from all unauthorized users is fixed at some value, how does detection rate change with the number of such users? Fig. 10 shows that more unauthorized users leads to better detection for all three methods. In this example, the total transmit power for all the unauthorized users is equal to that of the authorized user.

VII. Conclusion

In this paper, we investigated the problem of detecting unauthorized spectrum usage in a dynamic spectrum access network. We first proposed a zone-based network structure and a spectrum access policy such that authorized users will not interfere with each other. We then formulated the detection of anomalous spectrum usage as several statistical significance testing problems. We focused our study on two representative anomalous usage scenarios. For the case where no authorized signal is present, we utilized the existing cooperative sensing schemes and analyzed the impact of signal path loss on their detection performance, which has not done before. For the case where an authorize signal is present, we proposed three methods, Linearity-Check-given-Location (LCL), One-class Support Vector Machine (SVM) and Calibrating power (CAL), to detect anomalous signals by making use of the propagation characteristics. When additional information is available about the authorized transmitter, we derived analytical models for the test statistics. In the general case where no information is available about the authorized user, we proposed a solution using the machine learning technique, One-class SVM.

For the case where no authorized signal is present, our

simulation has shown that the hard decision combining, OR rule, outperforms the soft decision combining, EGC, in a large path loss fading environment, while the latter is known to have a near-optimal performance without considering the path loss effect. For the case where an authorize signal is present, we have shown that, in a typical urban microcell environment (i.e., $\gamma = 3.5$ and $\sigma_{dB} = 4$ dB), the proposed methods can achieve detection rates of $80 \sim 90\%$ for a single unauthorized user while keeping the false alarm rate as low as 10%. The detection probabilities are higher when more unauthorized users are present, even if the total interference power is constant.

Future work will focus on improving the robustness of the proposed methods against the propagation effects, γ and σ_{dB}. Promising solutions include: (i) Using the maximum residual as the test statistic, instead of using all the residuals (such as in (16)). This resembles the OR rule in the case where we detect unauthorized signals in noise, which has been shown superior to the soft combining in a large path loss fading environment; and (ii) including measurement samples from the anomalous case in the machine learning process. Existing algorithms, as in [25], can make use of undersampled anomalous training data to refine the acceptance region of the detection. We will also carry out experiments to confirm our theoretical analysis.

APPENDIX
PROOF OF THE EQUATION (27)

Let $\tilde{\mathbf{S}} = (S_1, \ldots, S_{N-1})^T$. Given $(S_i - S_N)$ and $(S_i^c - S_N^c)$ have independent and identical distribution,

$$\begin{aligned}
\mathbf{\Sigma}_e^c &= E[\hat{\mathbf{e}}^c(\hat{\mathbf{e}}^c)^T] \\
&= E[(\tilde{\mathbf{S}} - \mathbf{e}S_N - \tilde{\mathbf{S}}^c + \mathbf{e}S_N^c)(\tilde{\mathbf{S}} - \mathbf{e}S_N - \tilde{\mathbf{S}}^c + \mathbf{e}S_N^c)^T] \\
&= E[(\tilde{\mathbf{S}} - \mathbf{e}S_N)(\tilde{\mathbf{S}} - \mathbf{e}S_N)^T] \\
&\quad + E[(\tilde{\mathbf{S}}^c - \mathbf{e}S_N^c)(\tilde{\mathbf{S}}^c - \mathbf{e}S_N^c)^T] \\
&= 2E[(\tilde{\mathbf{S}} - \mathbf{e}S_N)(\tilde{\mathbf{S}} - \mathbf{e}S_N)^T] \\
&= 2E[\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T - S_N\mathbf{e}\tilde{\mathbf{S}}^T - S_N\tilde{\mathbf{S}}\mathbf{e}^T + S_N^2\mathbf{e}\mathbf{e}^T] \\
&= 2E[\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T] - 2\mathbf{e}E[S_N\tilde{\mathbf{S}}^T] - 2E[\tilde{\mathbf{S}}S_N]\mathbf{e}^T + 2E[S_N^2]\mathbf{e}\mathbf{e}^T \\
&= 2[\mathbf{I},\ -\mathbf{e}] \begin{bmatrix} E[\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T] & E[\tilde{\mathbf{S}}S_N] \\ E[S_N\tilde{\mathbf{S}}^T] & E[S_N^2] \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{e}^T \end{bmatrix} \\
&= 2[\mathbf{I},\ -\mathbf{e}]\mathbf{\Sigma} \begin{bmatrix} \mathbf{I} \\ -\mathbf{e}^T \end{bmatrix}, \qquad (30)
\end{aligned}$$

where $\mathbf{I}$ is an $(N-1) \times (N-1)$ identity matrix and $\mathbf{e}$ is an all-1 vector.

REFERENCES

[1] W. Xu, P. Kamat, and W. Trappe, "TRIESTE: A trusted radio infrastructure for enforcing spectrum etiquettes," in *Proc. Networking Technologies for Software Defined Radio Networks, 2006. SDR '06.1st IEEE Workshop on*, Sep. 2006, pp. 101–109.

[2] S. Verdú, *Multiuser detectiony*. New York: Cambridge University Press, 1998.

[3] H. L. V. Trees, *Detection, estimation, and modulation theory*. New York: Wiley, 2001.

[4] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 201–220, Feb. 2005.

[5] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, pp. 523–531, Apr. 1967.

[6] F. F. Digham, M. S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," in *Proc. Communications, 2003. ICC '03. IEEE International Conference on*, May 2003, pp. 3575–3579.

[7] R. Chen and J.-M. Park, "Ensuring trustworthy spectrum sensing in cognitive radio networks," in *Proc. Networking Technologies for Software Defined Radio Networks, 2006. SDR '06.1st IEEE Workshop on*, Sep. 2006, pp. 110–119.

[8] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, Nov. 2004, pp. 772–776.

[9] I. F. Akyildiz *et al.*, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Comput. Netw.*, vol. 50, pp. 2127–2159, May 2006.

[10] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors I. fundamentals," *Proceedings of the IEEE*, vol. 85, pp. 54–63, Jan. 1997.

[11] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors II. advanced topics," *Proceedings of the IEEE*, vol. 85, pp. 64–79, Jan. 1997.

[12] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: the cooperation-processing tradeoff," *Wireless Communications and Mobile Computing*, vol. 7, pp. 1049–1060, May 2007.

[13] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Experimental study of spectrum sensing based on energy detection and network cooperation," in *Proc. the first international workshop on Technology and policy for accessing spectrum*. Boston, MA: ACM, Aug. 2006.

[14] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. IEEE ICC*. Istanbul, Turkey: IEEE, Jun. 2006.

[15] E. Visotsky, S. Kuffner, and R. Peterson, "On collaborative detection of tv transmissions in support of dynamic spectrum sharing," in *Proc. DySPAN 2005. 2005 First IEEE International Symposium on*, Nov. 2005, pp. 338–345.

[16] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio networks," in *Proc. DySPAN 2005. First IEEE International Symposium on*, Nov. 2005, pp. 137–143.

[17] R. Tandra and A. Sahai, "Fundamental limits on detection in low snr under noise uncertainty," in *Proc. Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, 2005, pp. 464–469.

[18] A. Ghasemi and E. S. Sousa, "Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing," *Communications Letters, IEEE*, vol. 11, pp. 34–36, Jan. 2007.

[19] V. Erceg *et al.*, "An empirically based path loss model for wireless channels in suburban environments," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1205–1211, Jul. 1999.

[20] A. Taherpour *et al.*, "Asymptotically optimum detection of primary user in cognitive radio networks," *Communications, IET*, vol. 1, pp. 1138–1145, Dec. 2007.

[21] Y. Chen, W. Trappe, and R. P. Martin, "Attack detection in wireless localization," in *Proc. INFOCOM 2007. 26th IEEE International Conference on Computer Communications*, Anchorage, AK, May 2007, pp. 1964–1972.

[22] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Computer Networks*, vol. 43, pp. 499–518, 2003.

[23] B. Schölkopf *et al.*, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, pp. 1443–1471, Jul. 2001.

[24] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[25] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, Jan. 2004.

[26] A. J. Goldsmith, L. J. Greenstein, and G. J. Foschini, "Error statistics of real-time power measurements in cellular channels with multipath and shadowing," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 439–446, Aug. 1994.