Designing Sensor Networks to Protect Primary Users in Spectrum Access Systems

Matthew Clark*†, Konstantinos Psounis*
*University of Southern California, Los Angeles, CA
{clarkma,kpsounis}@usc.edu
†The Aerospace Corporation, El Segundo, CA

***Abstract*—In response to increasing demand on the radio frequency spectrum, regulators are implementing new spectrum access systems to share spectrum among diverse users. These systems leverage cognitive radio concepts to automatically identify suitable spectrum for users, avoiding harmful interference with higher priority users. An infrastructure of spectrum sensors is a key component of the spectrum access system, providing the means for the system to identify necessary protections. The geolocation precision of the sensing system will be limited in practice due to privacy concerns of the priority users as well as cost limitations. In this work, we examine the design options for the sensing component of a spectrum access system in terms of user performance and privacy. We apply machine learning techniques to treat the problem of estimating the priority users' state from sensor measurements, finding that the geolocation precision is closely tied to the density of the sensor deployment, and that the secondary user performance degrades rapidly if the density is too low, highlighting a tradeoff between priority user privacy and secondary user performance.**

I. Introduction

In response to increasing demand on the radio frequency spectrum from a growing variety of applications, spectrum regulators are implementing new spectrum access systems (SAS) to share spectrum among diverse users and radio technologies. In the U.S., the Federal Communications Commission has issued rulemakings to create a Citizens Broadband Radio Service (CBRS), opening the 3550-3700 MHz band for access to commercial mobile wireless systems, e.g., cellular and Wi-Fi [1], [2], where new entrants are expected to share the band with incumbent systems, such as military radars, which will retain priority access. The spectrum is anticipated to be managed by dynamic SAS that will leverage cognitive radio concepts to automatically identify suitable spectrum for users.

An infrastructure of spectrum sensors, called the Environmental Sensing Capability (ESC), is a key component of the SAS. Based on the measurements of the ESC, the SAS is expected to identify suitable protections to prevent harmful interference to priority/primary users (PUs). These protections must be enforced by the SAS when granting spectrum access to secondary users (SUs).

To make efficient use of the spectrum, the ESC inputs should enable the SAS to differentiate between multiple PU states at any given time in order to best allocate spectrum to other users. Specifically, the SAS will need to estimate the number of PUs operating, the location of the PUs, and the frequencies used. The precision of the sensing system will be limited in practice due to privacy concerns of the PUs. Because the detailed operations of the incumbent military radars is considered sensitive, prospective SAS operators and other industry stakeholders have proposed that sensing systems should not enable precise geolocation of PUs [3]. Instead, sensors in the system should be limited to detecting received signal strength, and should not include the use of directional antennas to support angle of arrival estimation, nor the use of precise timing information that would support time and frequency difference geolocation techniques.

Coarse geolocation of PUs will require more conservative access to the spectrum by SUs in order to avoid harmful interference. As a result, there is a potential tradeoff between the geolocation precision, corresponding to the privacy of the PUs, and the utility of the shared spectrum that can be achieved by the SUs accessing the spectrum via the SAS. The performance of the SAS in both these respects will depend on the physical deployment of the sensor network as well as the interpretation of the measurements from the sensors.

In this work, we formulate a model for the spectrum access system with a spectrum sensing infrastructure. We then examine the design options for the spectrum sensing component, exploring the feasible performance with respect to competing SU utility and PU privacy objectives in the 3550-3700 MHz CBRS band. We exploit machine learning techniques to estimate the PU operational state based on inputs from a network of energy detectors. We find that the performance of the secondary users drops dramatically if the sensor network is not deployed with sufficiently high density. Because the sensor density is also closely tied to the geolocation precision of the ESC, we characterize SU performance as a function of sensor density to evaluate the tradeoff between PU privacy and SU utility. With this method, we can quantify how a combination of strict PU privacy requirements and interference protection criteria limits SU access to the spectrum, helping to identify the SU applications that can operate effectively in CBRS.

The remainder of the paper is organized as follows. Related work is reviewed in Section II. A system model for the SAS, ESC, and user interface is provided in Section III. Several potential implementation approaches for the ESC are described in Section IV, and the performance in the SAS setting is quantified in Section V. A brief discussion of practical considerations for the ESC implementation is offered in Section VI followed by our final conclusions in Section VII.

Throughout this paper, we will use uppercase to denote vectors, arrays, and their elements, and subscripts to index the elements. Lowercase will denote scalar variables, where subscripts are used to distinguish variables that are similar in nature. Similarly, we use superscripts to distinguish related arrays. Calligraphic font will be used to denote sets, and bold face to denote random variables.

II. Related Work

Spectrum sensing has received much attention in the literature. Many formulations on the subject of cognitive radio assume sensing is conducted directly by SUs, which make local decisions on how to access the spectrum [4], [5]. In the SAS setting, the ESC is required to be part of the permanent infrastructure, leveraging a network of spectrum sensors to enable centralized SAS decision-making on SU access to the spectrum. The restriction to coarse geolocation for PU privacy precludes the application of related work that leverages directional antennas or antenna arrays to perform angle of arrival estimation as proposed in [6]. Further, the restriction to received signal strength measurements precludes the application of techniques such as cyclostationary detection and matched filter detection as proposed in [7], [8].

With these limitations, we can consider the ESC as a network of energy detectors, where such networks have been well studied and can be analytically designed to achieve target missed detection and false alarm rates with respect to a given signal to noise ratio at the sensor [9]. The performance of an individual sensor is limited by any fading along the interference path, potentially producing a "hidden node problem," while uncertainty in the thermal noise floor of the sensors can also degrade performance of an individual sensor [10]. As a result, cooperative sensing techniques have been proposed where multiple networked sensor measurements are used to achieve more accurate detection. Optimal determination from the sensor measurements can provide increased robustness [11]–[14], but at the cost of communication overhead and complexity. As a compromise, approximate approaches are employed to fuse measurements, including hard decision voting methods and linear fusion [15], [16]. One limitation of these approaches is that they consider a binary PU state where either the PU is present or it is not. This presents a challenge in extending these approaches to the SAS setting, where multiple PUs may be operating, and coarse estimation of location, frequency and time of use is needed to enable resource allocation to SUs, even though precise geolocation is prohibited.

Multiple PU states are considered in [17] where an integer linear programming problem is formulated and solved to determine the underlying PU state from hard decisions across multiple sensors. The proposed solution requires knowledge of the missed detection and false alarm performance of the sensors a priori, complicating the application of this approach to the SAS setting where estimating these probabilities for each potential PU state will be challenging. Further, the approximate OR mixture model is only suggested for high SNR regimes, e.g., greater than 10 dB, where it is not clear that this assumption is realistic in the SAS setting.

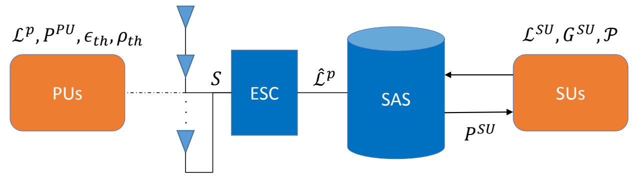


Fig. 1. System Model.

Machine learning techniques, where solutions are learned from a set of training data have been considered in limited cognitive radio and spectrum sharing settings [18], [19]. Machine learning has specifically been applied to the sensing problem in [20] where several classification methods were considered for determining whether a particular frequency channel was free of PU activity based on inputs from a number of energy detectors. This approach has potential applicability to the SAS problem, but the binary classification formulation must be generalized to the SAS setting before the impact of the achieved missed detection and false alarm probabilities on SU utility and PU geolocation precision can be assessed.

Where some of the above works can potentially provide insight into implementation of an ESC, they do not treat how the ESC inputs should be handled by the SAS and translated into spectrum access decisions for SUs. The problem of granting SU access was treated by [21] with a reinforcement learning solution, but this requires deployment of sensors at PU boundary locations to provide reliable feedback to support the learning process. Since PUs are mobile and cannot be confined to a fixed area, it is not clear that such a reinforcement learning approach can be applied.

In this work, we implement and assess an ESC with a network of energy detectors. We evaluate the dependency of the SAS performance on both the physical parameters of the deployed sensor network as well as on the approach for interpreting and translating the detector measurements to SU spectrum access. We specifically apply several machine learning approaches, demonstrating how design of the ESC can trade between SU utility and the effect on PU privacy due to the precision of geolocation.

III. System Model

Spectrum sharing with a SAS consists of a direct interface with SUs, a network of sensors to detect PUs, an ESC to process the sensor measurements, and a central SAS that takes ESC and SU inputs and determines spectrum assignments for SUs as outputs. This overall system is illustrated in Figure 1 and we offer a detailed model of each component in the following sub-sections.

A. SU Information Requirements

An SU requests an assignment from the SAS by sending information on its present location and parameters via a connection that does not rely on spectrum from the SAS. SU devices, indexed from 1 to n_{SU} and referred to as SUs for brevity from this point on, will each provide a range of useful transmission powers, a set of frequency tuning ranges, and a range of useful bandwidths, with each reflecting the capabilities of the device and the requirements of the application(s) running on that device. To simplify the notation, we will assume these SU restrictions are captured in a set of feasible assignments $\mathcal{P}$. We also treat the SU location as a scalar index for a cell in a discretized region, although in practice, it may be given as a vector of coordinates. The SAS stores all SU locations in a set we denote $\mathcal{L}^{SU} \subseteq \mathcal{L}$, where $\mathcal{L}$ is the set of all discretized locations in the considered region. To allow the access system to take advantage of frequency dependent scheduling, an SU may also send channel state information for the links in the SU network. Assuming n_c discrete frequency channels, the SAS will store all SU reported channel gains in an array $G^{SU} = [G^{SU}_{k,i}]$, where k is the frequency channel index and i is the index for the ith SU.

To ensure concepts are clear, we will focus on a specific example where the PUs consist of military radars and the SUs consist of cellular network operators, which send information about the transmissions they would like to schedule. Suppose the transmissions are specifically from the cellular user equipment (UE) to base station (BS) receivers. Frequency, power and bandwidth ranges identify device hardware limitations as well as application specific requirements, while G^{SU} specifies the estimated channel gains on the UE-to-BS transmission path.

B. PU Information Requirements

The locations of PUs operating in the region are described by a set denoted by $\mathcal{L}^p \subseteq \mathcal{L}$, where there is a total of n_{PU} operational PUs. PUs will operate with transmission powers $P^{PU} \in \mathbb{R}_+^{n_{PU}}$, where the range of potential transmit powers should be provided a priori to the SAS operators to enable reliable detection by the ESC. In our military radar example, all PUs are both transmitters and receivers. In general, PU transmitters and receivers may be physically separated. In this case, a policy should be identified to the SAS operator to infer potential receiver locations based on detection of the transmitter locations. The PU operators will also provide interference protection criteria to the SAS a priori, captured by variables $\epsilon_{th} \in \mathbb{R}_+$ and $\rho_{th} \in \mathbb{R}_+$, where ϵ_{th} is a harmful received interference power threshold and $0 < \rho_{th} < 1$ is a reliability requirement for the PUs, i.e., the maximum probability that the threshold given by ϵ_{th} can be exceeded. This reliability parameter accounts for inherent uncertainty in both the ESC detection process and the ability of the SAS to predict aggregate interference from the SUs. In our example, ϵ_{th} could be the received interference power that would prevent a radar from detecting a target. Since a practical radar will have some baseline probability of missed detection even without SU interference, selecting ρ_{th} to be significantly less than this baseline probability will ensure that sharing spectrum with SUs will have a negligible effect on the overall performance of the radar.

C. ESC Interface

A total of n_{esc} sensors are deployed throughout the region to enable detection and coarse geolocation of the PUs. We assume quiet periods of duration ν are scheduled where no SUs are granted access to the spectrum such that measurement of the PU transmissions is only effected by the sensor bandwidth b and the sensor thermal noise η_{esc}. The measurements, corresponding to the energy detected by each sensor, are denoted by the random vector $\mathbf{S} \in \mathbb{R}_+^{n_{esc}}$ where the randomness follows from modeling the thermal noise on the sensor as additive white gaussian noise. The energy detected by the kth sensor is

$$\mathbf{S_k} = \frac{b}{\eta_{esc}} \int_\nu \mathbf{V_k}^2(t)dt,$$

where $\mathbf{V}(t) \in \mathbb{R}^{n_{esc}}$ is the vector of voltages on the sensors at time t, and b/η_{esc} is a normalization factor. If the received signal is just the thermal noise, then the output of the detector follows a chi-square distribution with $2b\nu$ degrees of freedom, i.e., $\mathbf{S}_k \sim \chi^2_{2b\nu}$. If n_{PU} PUs are transmitting during the sensing interval with transmit powers given by P^{PU} and gain between the PUs and the detectors denoted by $G^{esc} = [G^{esc}_{jk}] \in \mathbb{R}_+^{n_{PU} \times n_{sc}}$, then the detector output follows a non-central chi-square distribution with $2b\nu$ degrees of freedom, i.e., $\mathbf{S}_k \sim \chi^2_{2b\nu}(2\gamma_k)$, where $2\gamma_k$ is the non-centrality parameter and $\gamma_k = b\nu \sum_{j=1}^{n_p} P_j^{PU} G^{esc}_{jk}/\eta_{esc}$.

The sensor measurements in $\mathbf{S}$ may be sent to a centralized ESC for processing, or a hard binary decision may be sent back to the ESC, indicating whether the measured energy is above or below a specified threshold at each sensor. In either case, the ESC must translate the inputs from the sensors into a set of estimated PU locations, denoted $\hat{\mathcal{L}}^p$, that should be protected from harmful interference by the SAS. Optionally, the ESC may also provide the SAS with a corresponding confidence vector, indicating an estimate of the probability that each identified location corresponds to the location of an operational PU. Methods to interpret the sensor inputs by the ESC will be addressed in the next section.

D. SAS Assignments to SUs

The SAS will manage a set of frequency channels $\mathcal{F} = \{1, ..., n_c\}$. The SAS will also use discrete power levels and discrete time slots, where the duration of these slots are chosen as a trade between efficiency and complexity of the system. The access system will assume a propagation model with uncertainty to predict the channel gain between SUs and interference protected locations, i.e., $\mathbf{G}^{PU} = [\mathbf{G}^{PU}_{k,i,j}] \in \mathbb{R}^{n_c \times n_{SU} \times n_{PU}}$ is the random array for the channel gains between each PU (e.g., radar) and SU device (e.g., UE) with indices k, i and j corresponding to the frequency channel, the SUs and the n_p PUs identified to the SAS respectively.

For each upcoming time slot, the SAS will need to solve a scheduling problem to maximize some utility function subject to constraints protecting the PUs from harmful interference. A SAS assignment function $f()$ operating on n_{SU} SUs and n_{PU} PUs should return maximum transmit power assignments for each SU-channel pair as an array $P^{SU} = [P^{SU}_{k,i}] \in \mathbb{R}_+^{n_c \times n_{SU}}$ according to

$$\begin{aligned} P^{SU} =& f(\mathcal{L}^p, \mathcal{L}^s, G^{SU}, \epsilon_{th}, \rho_{th}, \mathcal{P}) \\ \approx& \arg\max_{P' \in \mathcal{P}} \quad U(P', G^{SU}) \\ & \text{subject to} \quad Pr\left(\sum_{k=1}^{n_c}\sum_{i=1}^{n_{SU}} P'_i \mathbf{G^{PU}_{k,i,j}} \geq \epsilon_{th}\right) \leq \rho_{th} \\ & \qquad 1 \leq j \leq n_{PU}, \end{aligned} \tag{1}$$

where $U()$ is some utility function to be defined for the SUs. Note that a power assignment of zero is possible and corresponds to excluding a particular SU from being served in the corresponding frequency channels during this time slot. In this way, $f()$ acts as an admission control, channel assignment, and power assignment function.

Identifying solutions to (1) is non-trivial. For the purpose of this paper, we will offer a general methodology, but when a specific form for $f()$ is called for in the following results and ESC evaluations, we will limit the analysis to a single frequency channel for tractability, and make use of the algorithm in [22]. This approach accounts for the network topology of the SUs, the aggregate interference due to multiple SUs interfering with a single PU, the need for the algorithm to be low in complexity for practical implemenation, and the uncertainty in estimating the interference that will result for a particular scheduling decision. We can view this particular algorithm as a method to determine with a function $I()$, the maximum mean equal interference power level that can be caused by each of the SUs and still satisfy the interference constraints. $I()$ is specific to the uncertainty model assumed and is negatively correlated with the number of SUs that will receive nonzero power assignments, which we denote by a, $0 \leq a \leq n_{SU}$, and where a is selected to maximize $U()$. The power assignment P_i^{SU} for the ith SU is given by

$$P_i^{SU} = \frac{I(a, \epsilon_{th}, \rho_{th})}{\max_j \bar{G}^{PU}_{ij}}, \tag{2}$$

where we drop the channel index subscript to simplify the notation in the single channel case. In order to handle multiple PU locations, the SAS computes the mean SU-to-PU channel gain, $\bar{G}^{PU}_{ij}$, i.e., the mean of $\mathbf{G}^{PU}_{ij}$, from the inputs $\hat{\mathcal{L}}^p$ and $\mathcal{L}^s$. The maximization clusters SUs with the nearest estimated PU, i.e., for a location $l \in \hat{\mathcal{L}}^p$, we select all SUs that are closer to l than any other element in $\hat{\mathcal{L}}^p$ and compute equal interference assignments for each cluster of SUs separately. a therefore depends on the clustering, and we define a vector $A = [A_i] \in \mathbb{R}_+^{n_s}$ to denote, for each SU, the number of SUs that receive nonzero power assignments in the same PU cluster. We also include a small margin in the output of $I()$ to account for interference sources other than those in the specific PU cluster.

The utility $U()$ may be left general, addressing considerations including throughput, fairness and multiple access among SUs. In the remainder of this work, we assume a single SU network where fairness between SUs is assumed to be handled external to the SAS, e.g., by the cellular network operator, allowing us to focus on design of the ESC. Specifically, when a specific form for $U()$ is required, we will use the sum-rate of the SUs as our metric.

E. *Adversary Threat to PU Privacy*

Considering that the PUs will elect to rely on the ESC for protection from harmful interference rather than provide their operational information to the SAS directly, it is the connection to the SAS that appears to be untrusted, and we assume a threat model where an adversary has direct access to the information provided to the SAS by the ESC. Threat models considered in other SAS related works, e.g., inference attacks based on the assignments to the SUs as considered in [23], could be applied, but this would be a less direct measure of the effect of the ESC design on PU privacy. Instead, by focusing on the compromised ESC-SAS interface threat model, the privacy of the PUs can be characterized by the geolocation precision of the ESC output, as well as the missed detection and false alarm probabilities achieved.

IV. ESC Reporting and Detection

There are two design variations on sensor reporting. One approach is for each sensor to report their measurement directly to the ESC, subject to some quantization error, for further processing. Alternatively, each sensor, or group of sensors, will make a local decision regarding the presence of a PU, and report that decision to the ESC. The latter approach requires less communication overhead between the sensors and the ESC, but may offer reduced performance. However, the reduced fidelity of sensor information may provide a privacy advantage.

The detection problem for the ESC consists of translating the collection of sensor inputs (measurements or decisions) into estimated PU locations that require harmful interference protection from the SAS. In this section, we describe several approaches to the ESC detection problem, and consider a quantitative comparison in the next section.

A. *Optimal Assignments from Sensor Measurements*

Given a vector of sensor inputs S, the ESC can treat the PU state as a random vector $\mathbf{X} \in \mathbb{X}^{n_t}$, where $\mathbb{X}$ is the sample space, and n_t is the maximum number of PUs to be tracked by the ESC. The ESC can compute the conditional probability mass function

$$p_{\mathbf{X}}(X|S) = \frac{p_{\mathbf{S}}(S|X)p_{\mathbf{X}}(X)}{\sum_{X' \subset \mathbb{X}} p_{\mathbf{S}}(S|X')p_{\mathbf{X}}(X')},$$

where the term on the right follows from Baye's theorem and $\mathbf{S}$ is also treated as random. $p_{\mathbf{S}}(S|X)$ can be computed by the ESC based on an assumed noise model for the sensors, such as AWGN, and a propagation model for the gain between the PUs and the sensors. For simplicity and to make key concepts clear, suppose the SAS can compute the gain between any two locations exactly, i.e., we ignore fading, shadowing and position uncertainty. Let $g : \mathbb{X} \to \mathbb{R}_+^{n_{SU}}$ be the mapping from a single PU state to the gain between the SUs and that PU. For the jth PU described by $\tilde{X}_j$, $G_j^{PU} = g(\tilde{X}_j)$, where the subscript on the gain matrix refers to a specific column. Since the SAS does not know the true PU state $\tilde{X}$, it cannot compute G^{PU} directly, but it can treat this gain as a random variable such that $p_{\mathbf{G}^{PU}}(\hat{G}|S)$ can be computed from $p_{\mathbf{X}}(X|S)$. With this formulation, we can write the SAS problem of generating SU assignments P^{SU} as

$$\begin{aligned} \text{maximize} \quad & U(P^{SU}, G^{SU}, \mathcal{P}) && (3)\\ \text{subject to} \quad & P^{SU} = f(S, \mathcal{L}^s, G^{SU}, \mathcal{P}) && (4)\\ & Pr\{(P^{SU})^T g(\mathbf{X}_j) \geq \epsilon_{th} | S\} \leq \rho_{th} && (5)\\ & \forall j \in \{1, ..., n_t\}. && (6) \end{aligned}$$

Recognizing the relationship between the PU state and the gain, we can also write the constraint in (5) as

$$\sum_{\{X \in \mathbb{X}; (P^{SU})^T g(X_j) \geq \epsilon_{th}\}} p_{\mathbf{X}}(X|S) \leq \rho_{th}. \tag{7}$$

The sensors are assumed not to have the capability to distinguish between PUs with the same characteristics, e.g., PUs with the same location and transmit power. If the a priori information regarding each tracked PU is identical, then ensuring that (7) is satisfied for the jth PU also ensures satisfaction of the constraint for the other $n_t - 1$ tracked PUs. This potentially offers a means to determine power assignments with a probabilistic guarantee on PU harmful interference. Suppose the power assignment ensures the interference threshold ϵ_{th} is satisfied at a subset of locations in the region. Then only candidate PU states X containing locations not in this subset will impact the lefthand side of the inequality in (7). By ensuring enough locations from the most probable X are protected from harmful interference, we can identify a feasible solution to the SAS power assignment problem.

Note that computing (7) is intractable if we can't enumerate all possible $X \in \mathbb{X}$ which is likely to be the case in practical scenarios. We are forced to consider approximate approaches to treating sensor inputs, though this formulation may be useful to gain insight into the limitations made in such approximations.

B. Hard Local Decisions in Sensing Cells

Instead of guaranteeing a probabilistic interference constraint for a fixed number of PUs, the SAS can instead partition the region into sensing cells and guarantee interference protection in each cell separately. Specifically, suppose the region is divided into n_{sc} sensing cells, and in each cell, a subset of the sensors in the ESC are used to compute the probability that a PU is operating within the cell. Rather than reporting the output of the energy detector directly to the ESC, we suppose that the sensor reports a hard decision, where the output of the detector is compared to a threshold θ, i.e.,

$$S_k \underset{D_k=0}{\overset{D_k=1}{\gtrless}} \theta,$$

where D_k denotes the hard decision reported to the SAS based on the comparison. Let H_k denote the hypothesis that a PU is operating in the kth sensing cell. The SAS can provide the following interference protection constraint to such a PU

$$Pr\{(P^{SU})^T \mathbf{G}_k^{PU} \geq \epsilon_{th} | H_k\} \leq \rho_{th}, \tag{8}$$

i.e., given that there is a PU in the kth cell, we can design the ESC and the SAS assignment process to ensure that the probability of harmful interference for that PU is held below the reliability threshold. Denote the correct detection and missed detection probabilities for this sensing cell as $p_d = Pr\{D_k = 1 | H_k\}$ and $p_{md} = Pr\{D_k = 0 | H_k\}$ respectively. We can expand (8) as

$$\begin{aligned} &Pr\{(P^{SU})^T \mathbf{G}_k^{PU} \geq \epsilon_{th} | D_k = 1\} p_d + \\ &\qquad Pr\{(P^{SU})^T \mathbf{G}_k^{PU} \geq \epsilon_{th} | D_k = 0\} p_{md} \leq \rho_{th}. \end{aligned} \tag{9}$$

Since $p_d = 1 - p_{md}$, with a little algebra, we find

$$\begin{aligned} &Pr\{(P^{SU})^T \mathbf{G}_k^{PU} \geq \epsilon_{th} | D_k = 1\} \leq \\ &\qquad \frac{\rho_{th} - Pr\{(P^{SU})^T \mathbf{G}_k^{PU} \geq \epsilon_{th} | D_k = 0\} p_{md}}{1 - p_{md}}. \end{aligned} \tag{10}$$

A conservative assumption by the SAS would be to suppose $Pr\{(P^{SU})^T \mathbf{G}_k^{PU} \geq \epsilon_{th} | D_k = 0\} \approx 1$, i.e., if the PU is not detected by the ESC, the resulting SU power assignment is very likely to cause harmful interference. With this assumption, the harmful interfence constraints can be guaranteed if the SAS assignments satisfy

$$Pr\{(P^{SU})^T \mathbf{G}_k^{PU} \geq \epsilon_{th} | D_k = 1\} \leq \frac{\rho_{th} - p_{md}}{1 - p_{md}}. \tag{11}$$

This constraint can be applied to a grid of potential PU locations within any sensing cell providing a positive hard local detection. Given γ and θ, we can analytically compute p_d and p_{md} for the case of a single PU, where the probability of a particular decision has been shown in [13] to be

$$Pr\{D_k = 1\} = Pr\{S_k > \theta\} = Q_{b\nu}(\sqrt{2\gamma_k}, \sqrt{\theta}),$$

where Q denotes the generalized Marcum Q-function. Although this could be used to set θ that upper bounds p_{md}, in practice, the potential operation of PUs in nearby sensing cells introduces a large false alarm rate with such a threshold. Instead, we will identify suitable thresholds and study the performance of this approach via simulation in the following section.

C. Machine Learning Methods

To approximate the potential performance of an ESC where sensor measurements are reported for centralized processing rather than decided on locally, we consider supervised machine learning classification and regression. For these methods, we will generate a set of labeled training data, $\{S, \mathcal{L}^p\}$, i.e., samples of sensor measurements and the underlying PU operational state that generated them. We then fit a model to this training data that can be used to predict $\hat{\mathcal{L}}^p$ from new observations S. In particular, we will consider the operation of support vector machines (SVM), logistic regression, and decision trees. We will provide a brief review of each of these methods here, referring the interested reader to the literature for an in-depth understanding [24].

SVM: A support vector machine identifies a separating hyperplane for the training samples, maximizing the separation between the samples of each class. Selection of a kernel specifies the dimension of the hyper plane. A linear kernel will construct a linear hyperplane in the dimensional space of the samples while, e.g., a gaussian kernel will allow for non-linear separation of the samples. We will consider both a linear and gaussian kernel in our analysis of the ESC.

Logistic Regression: In logistic regression, a vector of weights is identified such that a logistic function of the inner product of the weight vector with the observation vector provides the classification prediction. The weight vector is selected to maximize the log-likelihood of the classification on the training set. This weight vector can be viewed as the normal of a hyperplane, providing an interpretation similar to SVM as identifying a separating boundary between the classes in the training set.

Decision Trees: Decision trees branch on features of the observation such that each leaf of the tree corresponds to a decision about the corresponding class label of the feature vector. Learning is accomplished by recursively partitioning the training data based on an attribute value. An ensemble of trees can be grown by sampling from the training data with replacement. Several methods can be applied to make predictions with ensembles. We will make use of bootstrap aggregation, i.e., "bagging," the random subspace method, and Adaptive Boosting [25].

V. Simulation Results

We evaluate the performance of ESC implementations for CBRS via simulation, assuming a 20 km by 20 km region of operation. A number of PUs and SU transmitters are deployed randomly throughout the region. SU receivers are deployed on a grid with a 2.5 km inter-receiver spacing, a 5 MHz receive bandwidth and thermal noise of -101.5 dBm. A nearest neighbor approach is used to identify SU transmit and receive pairs. Sensors are also deployed on a grid with variable inter-sensor spacing. The sensors are assumed to have a thermal noise floor of -100 dBm, a measurement bandwidth of 5 MHz, and an integration time of $5e{-4}$ resulting in a time bandwidth product of 100. PUs transmit with +30 dBm, while SUs can utilize transmit power assignments from the SAS in the range -40 dBm to +24 dBm corresponding to typical UE transmit powers. A two-ray model for pathloss is assumed, with a path loss exponent of 2 out to a break point, and a path loss exponent of 4 beyond the breakpoint. The PUs and SU receivers are assumed to be at a height of 15 meters, while the SU transmitters are at a height of 2 meters, and the sensors at a height of 3 meters. A log-normal shadowing model is assumed such that any realization of the path loss will be normally distributed in the dB domain with a mean corresponding to the two-ray model, and a standard deviation of 10 dB. PU receivers are assumed to have a harmful interference power threshold $\epsilon_{th} = -114$ dBm and we will consider various ρ_{th} reliability parameters.

A. Tuning and Training

To identify suitable parameters for our models and train our machine learning classifiers, we generate a set of training data with 12,000 observation-label pairs, $\{S, \mathcal{L}^p\}$, in the training set. The first 10,000 pairs are used to train the models while the remaining 2,000 are held out for validation, i.e., to verify that the models generalize beyond the data they are trained on. Each training sample is generated with an independent identically distributed topology and realization of random shadowing on all channel gains. The number of PUs included in the topology is uniform over $\{0, 1, 2, 3, 4\}$.

Because the number of potential classes described by $\mathcal{L}^p$ is very large, we partition the region into sensing cells, where the kth sensing cell consists of the discrete locations nearer to the kth sensor than any other sensor. The training data is then used to separately train a classifier for each sensing cell, such that the kth classifier makes a binary prediction on whether a PU is or is not present in the kth cell. The i.i.d. generation of the training data results in an unbalanced training set for such a classifier in the sense that there are many more training samples where no PU is present in the kth cell than samples where a PU is present. To avoid fitting a model that reflects biased a priori estimates for the presence of a PU, we enforce balance in the training samples by placing an additional PU in the kth sensing cell in enough training samples that half correspond to the case that a PU is present in the cell. This modified training set is only used for training the kth classifier, and we return to the original training set for, e.g., the $k+1$ classifier, where the training set is again modified to enforce balance on the $k+1$ sensing cell.

For the hard local decision implementation, the training data is used to asses the effect of the decision threshold on detection reliability. The receiver operating characteristic (ROC) is provided in Figure 2 for hard local decision sensing, where the ROC shows the operating points with respect to true detections $(1-p_{md})$, and false alarms, i.e., when a sensor returns a positive detection even though no PU is present in the vicinity of the sensor. Each curve in Figure 2 corresponds to a different inter-sensor spacing in the deployment of the sensor grid. This inter-sensor spacing can be viewed as the geolocation precision of the system since this is the resolution with which ESC estimated PU locations are reported to the SAS.

In Figure 3 the ROC of the center-most sensing cell with the hard local decision approach is compared with machine learning approaches for a 2 km inter-sensor spacing. The machine learning approaches include two SVMs, one with a

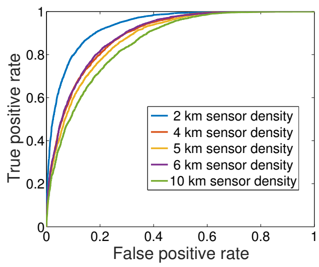


Fig. 2. Receiver Operating Characteristic of hard local decision sensing for various sensor densities.

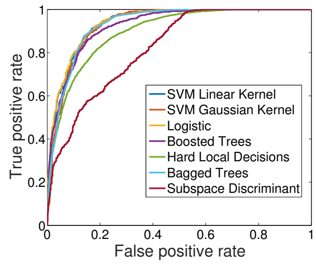


Fig. 3. Receiver Operating Characteristic of ESC methods with 2 km inter-sensor spacing.

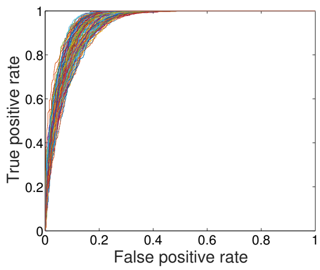


Fig. 4. Receiver Operating Characteristic of different gaussian SVM sensors over the region.

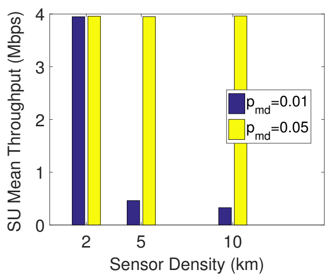
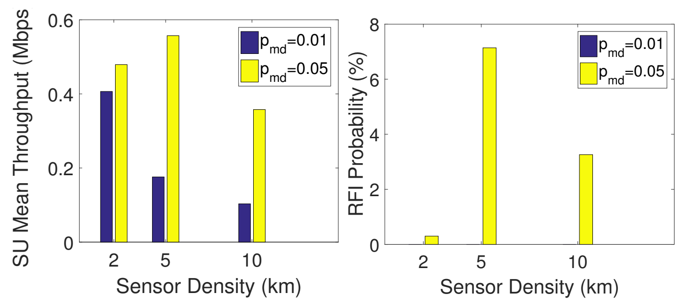


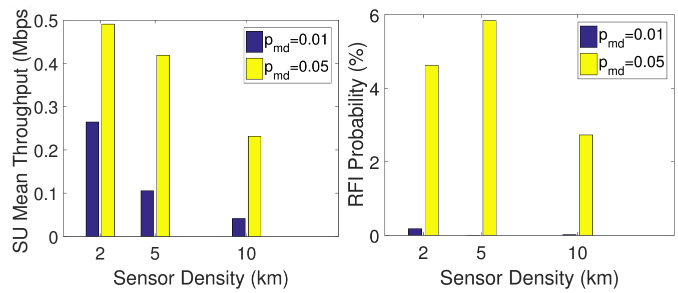
Fig. 5. Throughput per SU with no PUs present.



(a) Throughput per SU

(b) RFI to PU

Fig. 6. SAS and ESC Performance with one PU



(a) Throughput per SU

(b) RFI to PUs

Fig. 7. SAS and ESC performance with two PUs

linear kernel and one with a gaussian kernel, the three decision tree ensembles, and logistic regression. Clearly, the decision tree and hard local decision methods are outperformed by the SVM and logistic regression implementations. This holds for inter-sensor spacings we examined from 2 km to 10 km.

The center-most sensing cell is considered in Figure 3 because boundary effects lead to variations in performance between sensing cells. Figure 4, for example, overlays ROCs for each sensing cell using the gaussian SVM classifier, demonstrating the variability.

B. SAS Performance

The prior results allow us to compare the effectiveness of the different ESC approaches, but we want to quantify performance with respect to the SAS setting. To that end, we select the gaussian SVM classifier as the highest performing approach considered, and implement it as an ESC in the SAS setting. We deploy 40 SU transmitters in the region which make requests of the SAS and receive updated assignments every 30 seconds. All results are provided as the average over 200 sample topologies, where 20 minutes of user operations and SAS assignments are simulated with each topology.

Figure 5 plots the SU network achieved mean user throughput when there are no PUs present in the region with respect to the density of the sensor deployment and the target missed detection probability, which is set by selecting the appropriate threshold from the ROC for the classifiers. Because there are no PUs present, the SU throughput is only limited by any false alarms that occur. With a target missed detection probability of 5%, deployments with 2 km, 5 km, and 10 km inter-sensor spacing all perform similarly, achieving very low false alarm rates. However, with a target missed detection probability of 1%, the 5 km and 10 km density deployments experience much higher incidences of false alarms, reducing the achieved throughput significantly. Meanwhile the 2 km network is still able to operate with a low false alarm rate and offer nearly the same utility to the SU network as in the 5% missed detection probability case. This suggests that even if PU operations are relatively rare and sparse, a high density sensor network may be necessary to meet strict PU interference constraints without severely degrading SU utility due to frequent false alarms.

In Figure 6 we plot the case of a single PU operating in the region, with the SU utility in 6a and the percentage of time that the PU experienced harmful interference in 6b. In the 5% missed detection probability case, we do not observe a clear improvement in SU utility with increasing sensor density as we might have anticipated. Referring to Figure 6b, this can be explained by the higher incidence of harmful interference in the 5 km and 10 km case than in the 2 km case. This demonstrates potential for significant variability between performance observed with respect to training data, versus performance in an operational scenario. In the 1% missed detection probability case though, all three densities avoid causing harmful interference, and we again see that higher sensor densities are able to achieve fewer false alarms, and a higher SU utility as a result. In Figure 7 we provide similar plots for the case of two PUs, again observing a trend of increased SU utility with more dense sensor deployments, and significant variability between the target missed detection rate and observed rate of harmful interference.

VI. Practical Considerations

From Section V, we find a clear correlation between the SU utility and the sensor deployment density. However, sensor density may be limited to reduce the precision of geolocation information reported by the ESC to the SAS to alleviate privacy concerns. Further, obtaining property rights and cost factors are likely to limit the density of such deployments. Another challenge with a very dense deployment of sensors is the required backhaul infrastructure to provide communication between the ESC and the remote sensors. The ESC cannot rely on availability of the shared spectrum to support communication of this information. This may place additional limitations on potential sensor deployments.

The methods we have considered for the ESC rely on generation of training data to fit the classification models. Embedded assumptions in the training data may bias the resulting classifiers such that they do not perform well in practice. Potential assumptions that could lead to poor performance include: propagation loss models that do not closely approximate the real-world environment, erroneous assumptions about PU system characteristics, and poor PU behavioral models. To address these concerns, the ESC and SAS will need to be thoroughly tested in the operational environment, and several iterations of retraining may be required. Further, even if accurate classifiers are established, future changes to PU systems or operational behavior could require periodic retraining of the classifiers.

VII. Conclusions

Motivated by the example of spectrum sharing in the 3550-3700 MHz band, we have found that the problem of designing an effective spectrum sensing infrastructure and spectrum access system presents a challenging problem. Tuning the geolocation precision of the sensing system trades between the utility of the spectrum and the privacy of the users. In this work, we stated the problem of the spectrum sensing and access system formally and found an optimal solution to be intractable. We examined approximate approaches, including machine learning techniques, and identified an effective, practical implementation method, demonstrating performance through simulation. In the simulated scenario, we found that SU utility can be increased with higher density sensor networks, and that relatively high densities may be needed to meet PU interference constraints while also avoiding frequent false alarms that may be prohibitive to any SU service.

While this work focused on the design of the sensing system, in the 3550-3700 MHz band, some primary users will request interference protection via direct communication with the SAS as opposed to relying on the ESC. In future work, we will need to generalize our formulation and analysis to consider the operation of both types of PUs.

References

[1] "Report and order and second further notice of proposed rulemaking," *Federal Communications Commission*, , no. 15-47, GN Docket No. 12-354, Apr. 2015.

[2] "Order on reconsideration and second report and order," *Federal Communications Commission*, , no. 16-55, GN Docket No. 12-354, May 2016.

[3] "CBRS operational security technical specification," *Wireless Innovation Forum*, , no. WINNF-15-S-0071, June 2016.

[4] Y. Gai and B. Krishnamachari, "Decentralized online learning algorithms for opportunistic spectrum access," in *IEEE GLOBECOM*, 2011.

[5] L. Luo and S. Roy, "Efficient spectrum sensing for cognitive radio networks via joint optimization of sensing threshold and duration," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2851–2860, Oct. 2012.

[6] J. Wang, J. Chen, and D. Cabric, "Cramer-rao bounds for joint rss/doa-based primary-user localization in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1363–1375, Mar. 2013.

[7] M. Derakhshani, T. Le-Ngoc, and M. Nasiri-Kenari, "Efficient Cooperative Cyclostationary Spectrum Sensing in Cognitive Radios at Low SNR Regimes," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3754–3764, Nov. 2011.

[8] Ashish Rauniyar and Soo Young Shin, "Multiple Antenna-Aided Cascaded Energy and Matched Filter Detector for Cognitive Radio Networks," *International Journal of Distributed Sensor Networks*, 2015.

[9] H. Urkowitz, "Energy detection of unknown deterministic signals," in *Proc. IEEE*, Apr. 2011, vol. 55, pp. 523–531.

[10] R. Tandra and A. Sahai, "Snr walls for signal detection," in *IEEE J. Select. Topics Signal Process.*, Feb. 2008, vol. 2, pp. 4–17.

[11] E. Visotsky, S. Kuffner, and R. Peterson, "On collaborative detection of TV transmissions in support of dynamic spectrum sharing," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov. 2005, pp. 338–345.

[12] H. Uchiyama et al., "Study on Cooperative Sensing in Cognitive Radio based AD-HOC Network," in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept. 2007, pp. 1–5.

[13] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov. 2005, pp. 131–136.

[14] Z. Quan et al., "Optimal Multiband Joint Detection for Spectrum Sensing in Cognitive Radio Networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.

[15] Z. Quan et al., "Optimal Linear Cooperation for Spectrum Sensing in Cognitive Radio Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, Feb. 2008.

[16] M. Zhou, H. Chen, L. Xie, and K. Wang, "A Reliable Collaborative Spectrum Sensing Scheme Based on the ROCQ Reputation Model for Cognitive Radio Networks," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, May 2012, pp. 1–5.

[17] H. Nguyen et al., "Binary Inference for Primary User Separation in Cognitive Radio Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1532–1542, Apr. 2013.

[18] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of Machine Learning to Cognitive Radio Networks," *IEEE Wireless Communications*, vol. 14, no. 4, pp. 47–52, Aug. 2007.

[19] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A Survey on Machine-Learning Techniques in Cognitive Radios," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1136–1159, 2013.

[20] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine Learning Techniques for Cooperative Spectrum Sensing in Cognitive Radio Networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2209–2221, Nov. 2013.

[21] A. Galindo-Serrano and L. Giupponi, "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks," *IEEE Trans. Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, May 2010.

[22] M. Clark and K. Psounis, "Efficient resource scheduling for a secondary network in shared spectrum," in *IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2015, pp. 1257–1265.

[23] M. Clark and K. Psounis, "Can the privacy of primary networks in shared spectrum be protected?," in *IEEE INFOCOM*, Apr. 2016, pp. 1–9.

[24] Kevin P. Murphy, *Machine learning: a probabilistic perspective*, MIT Press, Cambridge, Mass., 2012.

[25] Yoav Freund and Robert E Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997.