

Decentralized Dynamic Spectrum Access for Cognitive Radios: Cooperative Design of a Non-Cooperative Game

Michael Maskery, Vikram Krishnamurthy, and Qing Zhao

Abstract—We consider dynamic spectrum access among cognitive radios from an adaptive, game theoretic learning perspective. Spectrum-agile cognitive radios compete for channels temporarily vacated by licensed primary users in order to satisfy their own demands while minimizing interference. For both slowly varying primary user activity and slowly varying statistics of “fast” primary user activity, we apply an adaptive regret based learning procedure which tracks the set of correlated equilibria of the game, treated as a distributed stochastic approximation. This procedure is shown to perform very well compared with other similar adaptive algorithms. We also estimate channel contention for a simple CSMA channel sharing scheme.

Index Terms—Cognitive radio, dynamic spectrum access, game theory, stochastic approximation, correlated equilibrium.

I. INTRODUCTION

TECHNOLOGIES such as mobile computing and cellular telephony are increasingly striving to deliver an “always connected” user experience. As these technologies become more ubiquitous, it becomes critical to make efficient use of limited radio resources to reliably deliver this experience to as wide a market as possible. This in turn requires active management of spectral resources, a challenge considering the decentralized structure of the radio system. To address this, researchers in cognitive radio [1], [2] propose RF devices that actively monitor and adjust to their radio environment to efficiently communicate in a crowded spectrum. This *dynamic spectrum access* functionality, in which cognitive radios compete for resources while respecting legacy (licensed) users, is the subject of this paper.

We explore the spectrum overlay approach (also referred to as opportunistic spectrum access) to dynamic spectrum access, using a game theoretic framework to highlight issues of cooperation and competition among multiple radios. In this model, cognitive radios share portions of the RF spectrum

(channels) that are temporarily unoccupied by licensed users. Each radio dynamically selects several available channels so as to balance its own demand (competition) against system-imposed sharing incentives (cooperation). Selections are made independently by each radio, based only on its own performance history. We focus on applications where primary users’ spectrum access activities either vary slowly with time (see [3], [4]), or where their spectrum access activities vary quickly, but average behaviour varies slowly. Example applications include the reuse of certain TC-bands that are not used for TC broadcast in a particular region.

Since optimal resource allocation in a decentralized, competitive environment is not straightforward, we propose to operate radios according to a game theoretic algorithm which slowly adapts resource allocation over time. We show that this algorithm tracks the time-varying set of *correlated equilibrium* actions of the game, so that each radio learns to respond optimally to its environment. For appropriate radio utilities, this equilibrium leads to globally efficient use of resources.

There are several reasons for using a game theoretic approach. First, since game theory explicitly recognizes the interdependence across radios, it can be used as a *synthesis tool* to provide decentralized algorithms for adaptive resource allocation. Second, the game theoretic concept of equilibrium provides a useful *analysis tool*; if we specify a simple algorithm that converges to an equilibrium, then we can characterize the long-run behaviour of the system, which may be measured against a global, system-wide objective.

In this paper, we assume that detecting the activities of primary users is sufficiently accurate that interference to primary users is below the required level. To resolve contention among cognitive radios, we use CSMA (carrier sense multiple access) to randomly allocate channel times among competing cognitive radios based on a reservation system. This simple mechanism allows us to capture fundamental issues such as uncertainty of the activity of others and nonlinear channel degradation due to crowding effects.

Related Work: Dynamic spectrum access presents technical challenges across the entire networking protocol stack. An overview of challenges and recent developments in dynamic spectrum access can be found in [5]. In the context of spectrum overlay, basic design components include spectrum opportunity identification and spectrum opportunity exploitation. The opportunity identification module is responsible for accurately identifying and intelligently tracking idle frequency bands that

Paper approved by F. Santucci, the Editor for Wireless System Performance of the IEEE Communications Society. Manuscript received April 4, 2007; revised December 18, 2007.

M. Maskery and V. Krishnamurthy are with the University of British Columbia, Department of Electrical and Computer Engineering, 2356 Main Mall, Vancouver, Canada, V6T 1Z4 (e-mail: {mikem, vikramk}@ece.ubc.ca).

Q. Zhao is with the University of California Davis, Department of Electrical and Computer Engineering 3165 Kemper Hall, CA 95616 (e-mail: qzhao@ece.ucdavis.edu).

This work was supported in part by a NSERC Strategic Grant, Canada, the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011 and by the National Science Foundation under Grant CNS-0627090.

Digital Object Identifier 10.1109/TCOMM.2009.02.070158

may be dynamic in both time and space. The opportunity exploitation module takes input from the opportunity identification module and decides whether and how a transmission should take place.

Spectrum opportunity detection can be reduced to a classic signal processing problem: detecting the presence of primary users' signals. Based on the cognitive radios' knowledge of the signal characteristics of primary users, three traditional signal detection techniques can be employed: matched filter, energy detector (radiometer), and cyclostationary feature detector [6]. While classic signal detection techniques exist in the literature, detecting primary transmitters in a dynamic wireless environment with noise uncertainty, shadowing, and fading is a challenging problem that has attracted much research attention [7]–[9]. When the activities of primary users are fast varying, spectrum opportunity tracking becomes a critical issue. This problem is addressed within the framework of Partially Observable Markov Decision Processes (POMDP) in [10].

Once spectrum opportunities are detected, cognitive radios need to decide whether and how to exploit them. In the design of the spectrum exploitation module, specific issues include whether to transmit given that opportunity detectors may make mistakes, what modulation and transmission power to use, and how to share opportunities among secondary users to achieve a network-level objective. The optimal design of spectrum access strategies in the presence of spectrum sensing errors has been addressed in [11]. Specifically, the interaction between the spectrum access protocols at the MAC layer and the operating characteristics of the spectrum opportunity detector at the physical layer is quantitatively characterized, and the optimal joint design of opportunity detectors, access strategies, and opportunity tracking strategies is obtained. Orthogonal frequency division multiplexing (OFDM) has been considered as an attractive candidate for modulation in spectrum overlay networks as discussed in [12], [13]. Power control for cognitive radios needs to take into account the detection range of the opportunity detector, the maximum allowable interference level, and the transmission power of primary users [14].

Spectrum opportunity sharing among cognitive radios, which is the focus of this paper, has been addressed in the literature. The problem of noncooperative radio resource allocation is considered in [3], [4] and related work from a non-game theoretic perspective, and in [15] from a game theoretic one, using a similar approach to the one presented here.

In related areas of wireless communications, game theoretic approaches have been used with considerable success. For example, efficient decentralized power control algorithms in CDMA networks have been devised using non-cooperative game theory in [16]–[20]. Each node in the CDMA network chooses a transmission power level to maximize its own signal-to-noise ratio while conserving power. However, there are two key differences between our approach and that of the above references. First, our application domain is a set of collision channels since channels here are shared according to a CSMA instead of a CDMA scheme. Second, the game considered in these papers is highly structured, with a strategic complementarity present between players: if one

node increases its transmission power, other nodes will find it optimal to increase their own power in turn. The game we analyze does not have this structure, and hence a more complex algorithm (regret tracking) is required for achieving a game theoretic correlated equilibrium instead of a Nash equilibrium. Correlated equilibria [21] are a generalization of Nash equilibria. The set of correlated equilibria is more natural in decentralized adaptive learning environments than Nash equilibria since it allows for individual players to coordinate their actions. This coordination can lead to higher performance than if each player picked actions independently as required by a Nash equilibrium. Furthermore, as pointed out in [22], it is typically unreasonable to expect in a learning environment that players act independently since the common history observed by all players act as a natural coordination device.

a) Organization of Paper: There are several underlying themes in this paper. Based on the model of Section II, we investigate how radios can estimate information about their environment in Section III. The issue of performance evaluation based on this information is dealt with in Section IV, which includes consideration of whether radios are designed for selfish or cooperative behaviour. Section V discusses our main decentralized adaptive algorithm for spectrum access, along with some related variants. These are compared through simulation in Section VI, and the main procedure is shown to be superior.

II. OPPORTUNISTIC SPECTRUM ACCESS MODEL

We consider a network of fully connected cognitive radios communicating by exploiting channels unused by primary users. We divide time into equal slots of length Λ , and label discrete time slots $n = 1, 2, \dots$ (we also refer to a slot as a decision period). At the beginning of the n^{th} time slot, each cognitive radio $l = 1, 2, \dots, L$ knows the following:

- 1) \mathcal{C} , the number of channels in the radio system.
- 2) $C_n \in \mathbb{R}^{\mathcal{C}}$, the channel quality vector (bits per time slot for each channel) at time n .
- 3) $\mathbf{Y}_n \in \{0, 1\}^{\mathcal{C}}$, the current channel usage pattern of primary users; channel $i \in \mathcal{C}$ is in use if $\mathbf{Y}_n(i) = 1$.
- 4) \mathbf{X}_n^l , the channel allocation decision of cognitive radio l at time n .
- 5) $d_n^l \in \mathbb{R}$, the current demand level of cognitive radio l (in bits per time slot).
- 6) $m^l \in \mathbb{Z}^+$, the maximum number of channels that l may use simultaneously.

All these quantities are static or vary slowly in time. An important characteristic of this model is that radio-specific quantities d_n^l and m^l need only be known to radio l , thus allowing for efficient decentralized resource allocation algorithms. The time-dependence of C_n allows us to consider sharing channels with fast primary users, whose *statistics* are slowly varying. A fast primary user on channel i periodically preempts cognitive radio activity, reducing the effective quality to $C_n(i)$. We will refer to these specifically as fast primary users throughout the paper, while the term “primary user” will be reserved for the slow varying type. Next, each radio l chooses channel

allocation \mathbf{X}_n^l from the state space:

$$\mathbf{S}_n^l = \left\{ x \in \{0,1\}^C : x \cdot \mathbf{Y}_n = 0, \sum_{i \in C} x(i) \leq m^l \right\} \quad (1)$$

That is, each radio can select up to m^l unused (by primary users) channels. Denote

$$\mathbf{X}_n \in \mathcal{S}_n = \mathbf{S}_n^1 \times \dots \times \mathbf{S}_n^L \quad (\text{joint action space of all radios}) \quad (2)$$

Cognitive radios share channels using a simple carrier sensing multiple access (CSMA) scheme, as follows: Divide each decision period n into K equal subslots, labeled n_1, \dots, n_K . (So each subslot has length Λ/K .) For every subslot n_k and channel i such that $\mathbf{X}_n^l(i) = 1$, radio l executes the following:

- 1) Generate a backoff time $\tau_{n_k}^l(i)$ according to a uniform distribution on the interval $(0, \tau_{\max})$ for some fixed parameter τ_{\max} .
- 2) Upon expiry of the backoff timer, monitor channel i and transmit data only if the channel is sensed clear.

Exactly one radio will transmit successfully on Channel i in subslot n_k , provided that its backoff time is sufficiently less than the next smallest time (allowing time to sense the channel clear and switch from receive to transmit mode). Otherwise there will be a collision on Channel i for that subslot.

For each $n, i : \mathbf{X}_n^l(i) = 1$, and $k = 1, 2, \dots, K$, denote success in subslot n_k by:

$$\gamma_{n_k}^l(i) = I\{\text{channel } i \text{ captured by } l \text{ in subslot } n_k\}, \quad i \in \mathcal{C} \quad (3)$$

where $I\{\cdot\}$ is the usual indicator function. At the end of decision period n (of length Λ), each radio l will have collected the following information:

$$(\gamma_n^l(i), \tau_n^l(i)) = \{(\gamma_{n_k}^l(i), \tau_{n_k}^l(i)) : \mathbf{X}_n^l(i) = 1, k = 1, \dots, K\} \quad (4)$$

This information will be used for adaptive decision making in subsequent sections.

A block diagram of our proposed system is given in Figure 1. There are three time scales in our problem formulation. The slowest time scale corresponds to the variation of primary user activity and demand levels. Second, and much faster, is the decision time scale (intervals of length Λ) of the cognitive radios themselves, and third, the fastest time scale (intervals of length Λ/K), are the CSMA channel access attempts.

III. ESTIMATING CHANNEL CHARACTERISTICS

It is clearly undesirable that cognitive radios compete for the same channel while other channels lie idle. It is thus crucial that a cognitive radio recognize and avoid crowded channels. To address this, we discuss how to use the CSMA feedback (γ_n^l, τ_n^l) to estimate competition for channels, and calculate the expected throughput and number of collisions experienced by a radio, which will be used to adjust channel allocation decisions. In the remainder of this section, we analyze channel characteristic estimates for any arbitrarily chosen cognitive radio user l .

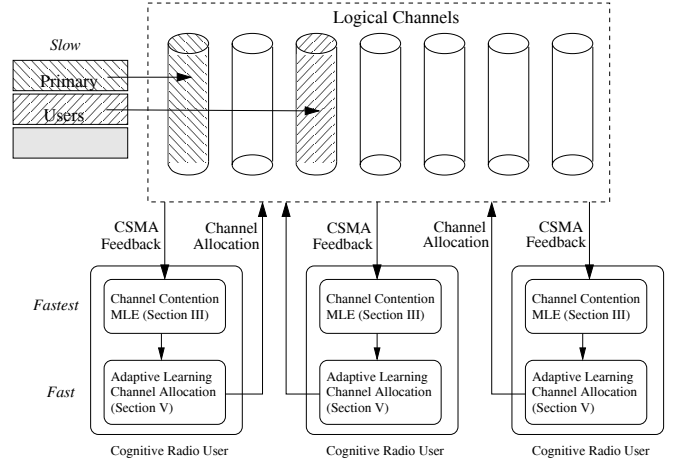


Fig. 1. Block diagram of decentralized learning system for cognitive radio opportunistic spectrum access. Cognitive radios select channels that are unoccupied by primary users so as to maximize their utility (formulated in Section IV). Feedback depends on the channel quality and the contention for resources due other users.

A. Channel Contention Estimate

To estimate $N_n^l(i)$, the number of users competing with l for channel i during decision slot n , consider a single CSMA channel access attempt on Channel i . There are $N_n^l(i) - 1$ other users (not including l), each choosing a random backoff time $\tau^m(i)$ uniformly on $(0, \tau_{\max})$. If l chooses backoff time $\tau^l(i)$, it captures the channel if $\tau^l(i) < \tau^m(i) - \delta$ for all $m \neq l$, where δ is the time required to sense the channel clear and switch its receiver from receive to transmit mode. Let $\tau_{N_n^l(i)-1}^{(1)}$ denote the smallest of the $N_n^l(i) - 1$ other backoff times (the first order statistic). It is well known that:

$$\begin{aligned} \mathbf{P}(l \text{ captures channel}) &= \mathbf{P}\left(\tau_{N_n^l(i)-1}^{(1)} > \tau^l(i) + \delta\right) \\ &= \begin{cases} \left(1 - \frac{\tau^l(i) + \delta}{\tau_{\max}}\right)^{N_n^l(i)-1}, & \tau^l(i) \leq \tau_{\max} - \delta. \\ 0, & \tau^l(i) > \tau_{\max} - \delta. \end{cases} \end{aligned} \quad (5)$$

Since $N_n^l(i)$, $i \in \mathcal{C}$ is fixed for the duration of the K CSMA subslots, and CSMA attempts are independent between subslots, radio l can compute the likelihood of contention level $N_n^l(i) - 1$ as:

$$\begin{aligned} L(N_n^l(i) - 1) &= \prod_{k: \gamma_{n_k}^l(i)=1} \mathbf{P}\left(\tau_{N_n^l(i)-1}^{(1)} > \tau_{n_k}^l(i) + \delta\right) \cdot \\ &\quad \prod_{k: \gamma_{n_k}^l(i)=0} \mathbf{P}\left(\tau_{N_n^l(i)-1}^{(1)} < \tau_{n_k}^l(i) + \delta\right), \\ &= \prod_{k: \gamma_{n_k}^l(i)=1} \left(1 - \frac{\tau_{n_k}^l(i) + \delta}{\tau_{\max}}\right)^{N_n^l(i)-1} \cdot \\ &\quad \prod_{k: \gamma_{n_k}^l(i)=0} \left(1 - \left(1 - \frac{\tau_{n_k}^l(i) + \delta}{\tau_{\max}}\right)^{N_n^l(i)-1}\right). \end{aligned} \quad (6)$$

The likelihood of contention level $N_n^l(i) - 1$ is defined as the probability that l is competing with $N_n^l(i) - 1$ other users on channel i , given l 's current observations of $(\gamma_n^l(i), \tau_n^l(i))$.

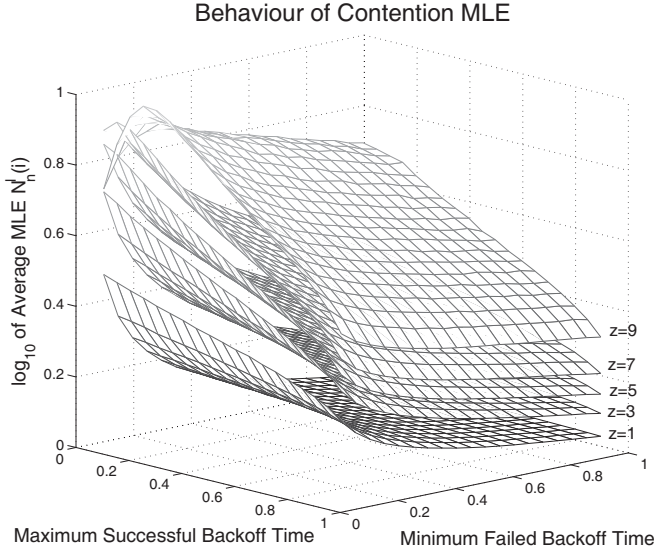


Fig. 2. Average result of the contention estimate $\widehat{N}_n^l(i)$. z denotes the number of failed CSMA attempts out of $K = 10$, and the maximum backoff time is $\tau_{\max} = 1$. Each data point represents an average of 5000 randomly generated observations with the specified maximum successful backoff time and minimum failed backoff time.

The maximum likelihood estimate $\widehat{N}_n^l(i)$ is the estimate of the number of competing users obtained by maximizing (6), i.e. solving:

$$\sum_{k:\gamma_{n_k}^l(i)=0} \frac{a_k^{\widehat{N}_n^l(i)-1}(i) \log(a_k(i))}{1 - a_k^{\widehat{N}_n^l(i)-1}(i)} = \sum_{k:\gamma_{n_k}^l(i)=1} \log(a_k(i)), \quad (7)$$

where $a_k(i) = 1 - (\tau_k^l(i) + \delta)/\tau_{\max}$.

Eq. (7) is difficult to solve, but we characterize the average behaviour numerically in Figure 2, which shows that $\widehat{N}_n^l(i)$ increases with the number of channel access failures, decreases when the maximum successful backoff time $\max_{k:\gamma_{n_k}^l(i)=1}(\tau_{n_k}^l(i))$ among the K subslots increases, and increases when the minimum unsuccessful backoff time $\min_{k:\gamma_{n_k}^l(i)=0}(\tau_{n_k}^l(i))$ among the K subslots decreases.

By replacing the terms on the left-hand side of (7) with their average \bar{a}_0 , one can approximate $\widehat{N}_n^l(i)$ by:

$$\widehat{N}_n^l(i) \approx 1 - \log \left(1 + \frac{|I_0(i)| \log(\bar{a}_0(i))}{\sum_{k:\gamma_{n_k}^l(i)=1} \log(a_k(i))} \right) / \log(\bar{a}_0(i)), \quad (8)$$

for $i \in \mathcal{C}$, where $|I_0(i)| = \sum_{k=1}^K (1 - \gamma_{n_k}^l(i))$, and $\bar{a}_0(i) = (\sum_{k:\gamma_{n_k}^l(i)=0} a_k(i)) / |I_0(i)|$. Numerical studies show that (8) is quite accurate on average, but may have large error when either $|I_0|$ or the successful backoff times are large. In this case, we recommend using (8) to generate an initial guess, which may be refined by the Newton-Raphson method.

B. Other Channel Characteristics

Using contention estimates $\widehat{N}_n^l(i)$, we can estimate the throughput and number of collisions on a channel (which are more relevant measures of a radio's performance), as follows:

Using (5) and the uniform distribution of backoff times, we may compute the unconditional probability (with respect to

the backoff time) of channel capture in a subslot as:

$$\begin{aligned} R_n^l(i) &= \mathbf{X}_n^l(i) \int_0^{\tau_{\max} - \delta} \left(1 - \frac{t + \delta}{\tau_{\max}} \right)^{\widehat{N}_n^l(i)} \frac{1}{\tau_{\max}} dt \\ &= \frac{\mathbf{X}_n^l(i)}{1 + \widehat{N}_n^l(i)} \left(1 - \frac{\delta}{\tau_{\max}} \right)^{1 + \widehat{N}_n^l(i)}. \end{aligned} \quad (9)$$

(This formula holds for $\widehat{N}_n^l(i) > 0$, otherwise $R_n^l(i) = 1$.)

Next, observe that l is involved in a channel collision in a subslot if either (a.) it has the lowest backoff time, but by a margin less than δ , or (b.) it does not have the lowest backoff time, but is within δ of the lowest. It follows that, for $\widehat{N}_n^l(i) > 0$, the probability of l being involved in a collision on Channel i in subslot n_k , using backoff time $\tau_{n_k}^l = t$ is given by:

$$\begin{aligned} Q_n^l(i, t) &= \mathbf{P}(t < \tau_{N_n^l(i)-1}^{(1)} < t + \delta \text{ or } t - \delta < \tau_N^{(1)} < t) \\ &= \mathbf{P}(\tau_{N_n^l(i)-1}^{(1)} < t + \delta) - \mathbf{P}(\tau_N^{(1)} < t - \delta) \\ &= \mathbf{X}_n^l(i) \left[\left(1 - \frac{\max\{t - \delta, 0\}}{\tau_{\max}} \right)^{\widehat{N}_n^l(i)} \right. \\ &\quad \left. - \left(1 - \frac{\min\{t + \delta, \tau_{\max}\}}{\tau_{\max}} \right)^{\widehat{N}_n^l(i)} \right]. \end{aligned} \quad (10)$$

Again, integrating out the backoff time gives the unconditional probability of collision:

$$\begin{aligned} Q_n^l(i) &= \frac{\mathbf{X}_n^l(i)\delta}{\tau_{\max}} + \frac{\mathbf{X}_n^l(i)}{1 + \widehat{N}_n^l(i)} \left[1 - \left(\frac{\delta}{\tau_{\max}} \right)^{1 + \widehat{N}_n^l(i)} \right. \\ &\quad \left. - \left(1 - \frac{\delta}{\tau_{\max}} \right)^{1 + \widehat{N}_n^l(i)} \right], \quad i \in \mathcal{C}. \end{aligned} \quad (11)$$

Since channel access attempts are i.i.d., $R_n^l(i)$ is also the expected proportion of successful CSMA attempts in a given decision period n , and $Q_n^l(i)$ is the expected proportion of CSMA attempts that result in collisions during period n . These quantities will be very useful for measuring radio performance in the next section.

IV. SYSTEM PERFORMANCE AND RADIO UTILITY

The goal of this paper is to achieve a *global* objective (efficient allocation of radio resources) using a *decentralized* scheme (local adaptation by individual radios). Consequently, we must demonstrate a connection between a global utility and the local utility function that will guide the allocation decisions of each radio in Section V. This connection is presented below through the derivation of global (Section IV-A) and local (Section IV-B) performance measures.

A. Global System Utility

When each cognitive radio has roughly the same priority and bandwidth requirements, a reasonable global objective is for channel resources to be allocated fairly, such that the proportion of resources captured by the worst-off cognitive radio (relative to its demand) is maximized (max-min fairness). That is, the global system utility in decision period n is:

$$U(\mathbf{X}_n) = \min_{l=1, \dots, L} \left(\min \left(\frac{C_n^T R_n^l}{d^l}, 1 \right) \right). \quad (12)$$

Here $R_n^l = [R_n^l(1), \dots, R_n^l(C)]^T$ is the vector of channel capture probabilities, where $R_n^l(i)$ is defined in (9). Each term $C_n^T R_n^l / d^l$ represents the ‘‘satisfaction level’’ of Radio l . This is the total amount of resource captured ($C_n^T R_n^l$) divided by the demand level d^l .

Substituting (9) into (12) allows evaluation of the global utility for given CSMA parameters (δ, τ_{\max}).

The system objective is to maximize $U(x)$ over all $x \in \mathbb{S}$, thus maximizing the satisfaction level of the worst-off user. However, for a decentralized implementation, we cannot choose an action from \mathbb{S} in a coordinated fashion, but each radio l must choose its own action $\mathbf{X}_n^l \in \mathbf{S}^l$ at time n . Moreover, as we will see in the next section, the global utility cannot be easily evaluated by any one user, so an appropriate substitute must be found.

B. Local Radio Utility

If each cognitive radio had a reliable estimate of $U(\mathbf{X}_n)$, then decentralized operation would be straightforward; each radio would simply act to maximize the global utility directly. Unfortunately, this would require l knowing d^m and \mathbf{X}_n^m for all $m \neq l$, which is unrealistic. We therefore construct a locally computable alternative utility function $u^l(\mathbf{X}_n^l)$ which mimics the behaviour of $U(\mathbf{X}_n)$ in the following sense. Channel allocation activities of radio l which are known to directly lead to a higher global utility $U(\mathbf{X}_n)$ by increasing the l^{th} term in the minimization (12) lead to a corresponding increase in $u^l(\mathbf{X}_n^l)$. Channel allocation activities of radio l which are likely to lead to a lower global utility $U(\mathbf{X}_n)$ by decreasing a different ($m^{\text{th}} \neq l^{\text{th}}$) term in the minimization (12) lead to a corresponding decrease in $u^l(\mathbf{X}_n^l)$. Since the actual effect of l 's channel allocation actions on the global utility $U(\mathbf{X}_n)$ are unknown, the weighting to each of these effects is adjustable by changing weighting (pricing) parameters inherent in the local utility function $u^l(\mathbf{X}_n^l)$.

The first portion of l 's local utility reflects the self-interested component of (12):

$$u^l[0](\mathbf{X}_n^l) = \min \left(\sum_i \frac{C_n(i)}{d^l} \frac{\mathbf{X}_n^l(i)}{1 + N_n^l(i)} \left(1 - \frac{\delta}{\tau_{\max}} \right)^{1+N_n^l(i)}, 1 \right). \quad (13)$$

Maximizing (13) directly maximizes l 's part of the global utility $U(\mathbf{X}_n)$

A game with (13) as the only component of the utility function would resemble a classic congestion game, which might be readily solved in closed form. However, (13) neglects a good portion of the global objective; l should maximize others' satisfaction of demand as well as its own. Since we assume that each radio knows only its own demand and actions, (i.e. d^m and \mathbf{X}_n^m are unknown to radio l for $m \neq l$), we induce such cooperation through the following two principles:

- Radio l 's realized rate should not exceed its demand d^l , as this leaves fewer resources for other users.
- Radio l should minimize the number of CSMA collisions it causes, as this impacts the performance of other users.

These principles can be justified by noting that the components of (12) belonging to $m \neq l$ are decreasing in $\mathbf{X}_n^l(i)$.

To satisfy the first principle, we introduce a penalty for achieving excess rate:

$$u^l[1](\mathbf{X}_n^l) = -\frac{1}{d^l} (C_n^T R_n^l - (d^l + \beta))^+, \quad (14)$$

where $(y)^+$ denotes the operation $\max\{y, 0\}$, and parameter β represents the size of a ‘‘grace’’ region, where excess rate is not penalized. (Since l 's realized rate is observed in noise, it may ‘‘accidentally’’ satisfy more than its demand, so small excesses are not penalized.)

The second principle, to minimize collisions, is satisfied by considering $Q_n^l(i)$ in (11). Neglecting collisions involving three or more users, $Q_n^l(i)$ can be interpreted as the degradation of Channel i (proportional to $C_n(i)$) caused by l . If this degradation is spread evenly among all users, then the average performance degradation seen by any other radio caused by activity of Radio l on Channel $i \in \mathcal{C}$ is $D^l(i) = Q_n^l(i) / \hat{N}^l(i)$. Our penalty is a weighted sum of the channel degradations:

$$u^l[2](\mathbf{X}_n^l) = -\frac{1}{\sum_k C_n(k)} \sum_{i: \hat{N}^l(i) > 0} C_n(i) D^l(i). \quad (15)$$

The final utility function for Cognitive radio l is given by:

$$u^l(\mathbf{X}_n^l) = \max \left\{ u^l[0](\mathbf{X}_n^l) + \alpha_1 u^l[1](\mathbf{X}_n^l) + \alpha_2 u^l[2](\mathbf{X}_n^l), 0 \right\}, \quad (16)$$

Eq. (16), with user-defined parameters (α_1, α_2) , is used to guide channel allocations by allowing each radio to take action X_n^l , observe utility u_n^l , and generate a new action that increases its expected utility. The method for choosing actions is the subject of Section V.

V. DECENTRALIZED ADAPTIVE CHANNEL ACCESS

In this section we describe our decentralized learning approach to opportunistic spectrum access. Our approach is based on the regret matching procedure of [22], formulated as a distributed stochastic approximation algorithm. However, this procedure bases the behaviour of each cognitive radio user on the average history of all past performance. This is not desirable since the underlying conditions change as primary users and cognitive radio demand levels change over time. Instead, we have developed and investigated the use of an adaptive procedure, called ‘‘Regret Tracking.’’ The procedure is game theoretic in nature; it converges even when multiple cognitive radio users are simultaneously adapting their behaviour. This is critical observation; since naive, single-agent procedures do not account for the presence of multiple users, and hence may not converge.

A. Regret Tracking based Channel Access Algorithm

In the regret tracking procedure, each radio takes a sequence of actions

$$\{\mathbf{X}_n^l \in \mathbf{S}_n^l : n = 0, 1, 2, \dots\} \quad (17)$$

and define \mathbf{X}_n^{-l} as the vector of actions of the other radios and observes a sequence of rewards $\{u_n^l \in \mathbb{R} : n = 0, 1, 2, \dots\}$. The action at time $n+1$ is a random function of this history of actions and rewards.

Before proceeding, we clarify a point of notation. In (16), a radio's utility u^l is written only as a function of its own action. However, it is also implicitly a function of the channel contention $\tilde{N}^l(i)$, which depends on the actions of other players. In game theory, this is made explicit by writing $u^l(\mathbf{X}_n^l, \mathbf{X}_n^{-l})$, and we follow this convention below.

The exact channel allocation algorithm is summarized in Algorithm 5.1. This is a regret tracking procedure, executed independently by each radio.

Algorithm 5.1: Adaptive Learning for Channel Allocation: Define parameters $(u^l, \mu, \{\varepsilon_n : n = 1, 2, \dots\}, \theta_0^l, \mathbf{X}_0^l)$, where u^l are the radio utilities, μ satisfies

$$\mu > (S^l - 1)(u_{\max}^l - u_{\min}^l), l = 1, 2, \dots, L, \quad (18)$$

((u_{\max}^l, u_{\min}^l) are obtained from (16)), $\{\varepsilon_n\}$ are small step-sizes, and $\theta_0^l, \mathbf{X}_0^l$ are arbitrary initial regrets and actions. Also define the $\mathbf{S}^l \times \mathbf{S}^l$ instantaneous regret matrix with entries:

$$\mathbf{H}_{jk}^l(\mathbf{X}_n) = I\{\mathbf{X}_n^l = j\} (u^l(k, \mathbf{X}_n^{-l}) - u^l(j, \mathbf{X}_n^{-l})). \quad (19)$$

where \mathbf{X}_n^l and \mathbf{X}_n^{-l} are defined in (17). Each radio executes the following steps:

- 1) **Initialization:** Set $n = 0$, take action \mathbf{X}_0^l , and initialize $\theta_0^l = \mathbf{H}^l(\mathbf{X}_0)$. Set $n = 1$, take action $\mathbf{X}_1^l = \arg \max_k H_{jk}^l(\mathbf{X}_0)$, where $j = \mathbf{X}_0^l$, and set $\theta_1^l = \mathbf{H}^l(\mathbf{X}_1)$.
- 2) Repeat for $n = 2, 3, \dots$:
Action Update: Choose $\mathbf{X}_{n+1}^l = k$ with probability

$$\mathbf{P}(\mathbf{X}_{n+1}^l = k | \mathbf{X}_n^l = j, \theta_n^l = \theta^l) = \begin{cases} \max\{\theta_{jk}^l, 0\} / \mu, & k \neq j, \\ 1 - \sum_{i \neq j} \max\{\theta_{ji}^l, 0\} / \mu, & k = j. \end{cases} \quad (20)$$

Average Regret Update: Given $\mathbf{H}^l(\mathbf{X}_{n+1})$, update θ_{n+1} according to the following stochastic approximation (SA) algorithm with step size $\varepsilon_n > 0$

$$\theta_{n+1}^l = \theta_n^l + \varepsilon_n (\mathbf{H}^l(\mathbf{X}_{n+1}) - \theta_n^l). \quad (21)$$

Discussion of Algorithm 5.1: There are two possible choices for the step size. (i) Decreasing step size $\varepsilon_n = 1/(n+1)$ (ii) Constant step size $\varepsilon_n = \varepsilon$ where $0 < \varepsilon \ll 1$. The decreasing step size algorithm is used if the parameters of the cognitive radio system do not evolve; the algorithm converges with probability one to the set of correlated equilibria; see Sec.V-B below. The constant step size version of Algorithm 5.1 can track a slowly time-varying set of correlated equilibria or a set of correlated equilibria that jump change at infrequent intervals (e.g., if the mean time between jump changes is $O(1/\varepsilon)$, see [23]). Sec.VI-C below presents numerical examples of tracking correlated equilibria in dynamic environments.

Recall the channel allocation of radio l at decision period n is denoted $\mathbf{X}_n^l \in \mathbf{S}^l$, with the joint allocation of all radios denoted $\mathbf{X}_n \in \mathbf{S}$. Based on the history of allocations and utilities, radio l computes its average regret values $\theta_{n,jk}^l : j, k \in \mathbf{S}^l$, according to:

$$\theta_{n,jk}^l = \sum_{\tau \leq n: \mathbf{X}_\tau^l = j} \varepsilon_{\tau-1} \left(\prod_{\sigma=\tau}^{n-1} (1 - \varepsilon_\sigma) \right) \cdot (u^l(k, \mathbf{X}_\tau^{-l}) - u^l(j, \mathbf{X}_\tau^{-l})). \quad (22)$$

Depending on whether $\varepsilon_n = 1/(n+1)$ (decreasing step size) or $\varepsilon_n = \varepsilon$, (constant step size), $\theta_{n,jk}^l$ in (22) is either an arithmetic average or moving average. Eq.(21) is simply a recursive implementation of the these two averages.

The regret values $\theta_{n,jk}^l$ measure the average gain that l would have received had he chosen allocation k in the past (from time 0 to n) instead of j . If the gain is positive, then l is more likely to switch to k in the future. Note, however, that (22) requires that l knows what utility he would have received for each action, *even if that action was not taken*. We address this difficulty either by assuming that extra resources are available to monitor unused channels to reconstruct potential utilities, or by using the modified regret tracking algorithm of Sec. V-C.

B. Convergence of Channel Access Algorithm

Our goal is to prove that if each cognitive radio deploys Algorithm 5.1 in a decentralized fashion, then the network performance converges to the set of correlated equilibria of a non-cooperative game. We first define the terms correlated equilibrium and network performance.

Definition. Correlated Equilibrium: For a non-cooperative game comprising of L cognitive radios, define the joint strategy $\pi = (\pi^l, \pi^{-l})$ as the probability distribution on joint action space \mathbb{S}_n (2). Here π^l is the marginal distribution (strategy) of radio l , and π^{-l} is the marginal distribution of the remaining players. Then the set of correlated equilibria is the convex polytope $\mathcal{C}_e = \{\pi : \sum_{x^{-l} \in \mathbf{S}^{-l}} \pi(j, x^{-l}) [u^l(k, x^{-l}) - u^l(j, x^{-l})] \leq 0, \forall l, j, k \in \{1, \dots, L\}\}$.

Next, we introduce the empirical distribution of play, which can be viewed as a diagnostic that monitors the performance of the entire cognitive radio network.

Definition. Network Performance: Let $e_x = [0, 0, \dots, 1, 0, \dots, 0]$ with a one in the x^{th} position. The empirical distribution of play to time n is (τ and σ below denote time indices):

$$\bar{z}_n = \sum_{\tau \leq n} \varepsilon_{\tau-1} \left(\prod_{\sigma=\tau}^{n-1} (1 - \varepsilon_\sigma) \right) e_{\mathbf{X}_\tau}, \quad (23)$$

where ε_n is step size in (21). Similarly to $\theta_{n,jk}^l$, \bar{z}_n can be viewed as an average or moving average frequency of play, with $\sum_i \bar{z}_n(i) = 1$. Moreover, as was the case for θ_n , \bar{z}_n may be represented by the recursion:

$$\bar{z}_{n+1} = \bar{z}_n + \varepsilon_n (e_{\mathbf{X}_{n+1}} - \bar{z}_n), \quad (24)$$

where \mathbf{X}_{n+1} given by Step (2a) of Algorithm 5.1.

We also need the notion of convergence to the *set* of correlated equilibria: \bar{z}_n converges to a set \mathcal{C}_e if, for any $\varepsilon > 0$, there exists $N_0(\varepsilon)$ such that for all $n > N_0$, we can find $\psi \in \mathcal{C}_e$ at a distance less than ε from \bar{z}_n .

The proof that \bar{z}_n generated by Algorithm 5.1 converges to the set of correlated equilibria \mathcal{C}_e follows using stochastic averaging theory as in [24] for the decreasing step size case, and [25] for the constant step size case. We only provide a brief sketch of the result. As is now standard in stochastic approximation proofs, the first step is to introduce the continuous

time interpolated process

$$\bar{z}^\varepsilon(t) = \bar{z}_n \text{ for } t \in [\varepsilon n, \varepsilon(n+1)), \text{ for any } \varepsilon > 0 \text{ and } n \geq 0. \quad (25)$$

The main result is that $\bar{z}^\varepsilon(t)$ converges to the trajectory of the following differential inclusion:

$$\begin{aligned} \frac{dz}{dt} &\in \nu(z) \times \Delta S^{-l} - z, \text{ where} \\ \sum_k \nu_k[\theta_{kj}^l(z)]^+ &= \nu_j \sum_k [\theta_{jk}^l(z)]^+ \text{ for all } j, k \in \mathbf{S}^l. \end{aligned} \quad (26)$$

Here $\nu(z)$ is the set of probability distributions over \mathbf{S}^l ; ΔS^{-l} is the set of all probability distributions over joint actions of the $L - 1$ competing radios, and $\theta_{jk}^l(z)$ is the average regret corresponding to play history z . Recall that a differential inclusion specifies a family of trajectories and is a generalization of a differential equation which comprises of a single trajectory. Typically for stochastic approximation algorithms used in physical layer wireless communications (such as least mean squares adaptive filtering), the limiting process is an ordinary differential equation. Here, we are interested in the *set* of correlated equilibria rather than a single optimal point – and the limiting process is a differential inclusion (26).

The main result is stated as follows:

Theorem 5.1: Consider the interpolated process (25) generated by Algorithm 5.1. If all radios operate according to this algorithm, the following results hold:

- 1) All solutions $z(t)$ to (26) converge to the set of correlated equilibria as $t \rightarrow \infty$.
- 2) For a decreasing stepsize $\varepsilon_n = 1/(n+1)$, the trajectory of the interpolated process $z^\varepsilon(t)$ converges almost surely to a trajectory $z(t)$ satisfying (26).
- 3) Under a constant stepsize $\varepsilon_n = \varepsilon$ in Algorithm 5.1, the trajectory of the interpolated process $z^\varepsilon(t)$ converges weakly as $\varepsilon \rightarrow 0$ to a trajectory $z(t)$ satisfying (26).
- 4) As $t \rightarrow \infty$, since $z(t)$ converges to the set of correlated equilibria, the trajectory $z^\varepsilon(t)$ also converges to this set.

The proof of Theorem 5.1, parts (1), (2) and (4) is given in [24]. Part (3) follows from standard arguments as in [25].

C. Other Adaptive Strategies

For reference, we briefly review three alternative methods for decentralized decision making that can be applied to the opportunistic spectrum access problem; best response, fictitious play and modified regret tracking. We relate these to our regret tracking approach, and compare the three approaches in numerical examples.

b) Best Response: In the simplest approach, each cognitive radio chooses the channel allocation that maximizes its utility, assuming that the actions of other cognitive radios will not change. In Algorithm 5.1, this corresponds to setting $\varepsilon_n = 1$ and replacing Step (2a) with $\mathbf{X}_{n+1}^l = \arg \max_k (H_{jk}^l(\mathbf{X}_n))$ where $j = \mathbf{X}_n^l$. In best response, the average history θ_n^l does not need to be tracked at all, since actions are based solely on feedback from the previous step. Nevertheless, it still provides a useful performance measure for the system since it indicates how close the system is to equilibrium.

c) Fictitious Play: Best response suffers from the defect that actions of other users are assumed to be constant between iterations. Fictitious play has since been widely studied and has been shown to converge in many, but not all, games. In [26] it was shown that fictitious play is a special case of regret-based algorithms. Specifically, (assuming a decreasing stepsize ε_n), it corresponds to replacing Step (2a) in Algorithm 5.1 with $\mathbf{X}_{n+1}^l = \arg \max_k (\theta_{n,jk}^l(\mathbf{X}_n))$ where $j = \mathbf{X}_n^l$. An adaptive version of fictitious play, with constant stepsize ε , can also be generated in this manner.

d) Modified Regret Tracking: The preceding strategies rely on each cognitive radio knowing the value of *all* actions, not just those actually taken. This means radios must monitor all possible channels in order to determine $\widehat{N}_n^l(i)$ for all i and hence $u^l(k, \cdot)$ for all possible k . This can be achieved straightforwardly if extra receivers are used to scan channels not currently in use by the cognitive radio. An alternative procedure is proposed in [27], which replaces the explicit utility of actions not taken with an estimate, proceeds as follows. Label the probability distribution used to choose action \mathbf{X}_n^l in Step (2) of Algorithm 5.1 by p_n^l . We first replace (19) by:

$$\begin{aligned} \mathbf{H}_{jk}^l(\mathbf{X}_n) &= I\{\mathbf{X}_n^l = k\} \frac{p_n^l(j)}{p_n^l(k)} u^l(k, \mathbf{X}_n^{-l}) \\ &\quad - I\{\mathbf{X}_n^l = j\} u^l(j, \mathbf{X}_n^{-l}). \end{aligned} \quad (27)$$

This avoids the need to evaluate u^l for actions not taken. However, we must compensate for this by allowing radios to explore alternative actions. We therefore replace Step (2) of Algorithm 5.1 by:

- 2') Generate a uniform random variable U . If $U < \delta$ (small), choose \mathbf{X}_{n+1}^l from a uniform distribution over \mathbf{S}^l . Otherwise, choose $\mathbf{X}_{n+1}^l = k$ with probability as in (20).

It is proven in [27] that the non-tracking version of this algorithm converges almost surely to the set of correlated equilibria. However, in our tracking simulations, convergence is much slower, due to the required periodic random exploration.

VI. NUMERICAL EXAMPLES

We now provide a numerical (Matlab) comparison of the methods of Section V. For demonstration purposes, we specify a relatively small number of cognitive radio users (six) and channels (ten). We focus on moderately congested systems, with total capacity of channels roughly equal to total user demand. We assume $K = 20$ CSMA attempts per decision period on selected channels, and assume unused channels are scanned randomly such that information is obtained as if $K \approx 10$ CSMA attempts were made.

A. Effect of Utility Parameters on Regret Tracking (Pricing)

In this section we investigate the impact of (α_1, α_2) on system performance. (Recall from (16) that (α_1, α_2) , parameterize the radio utility $u^l(\mathbf{X}_n^l)$ and affect l 's level of cooperation with system objectives.) The problem of selecting optimal parameters can be viewed as a pricing problem (in this case carried out offline).

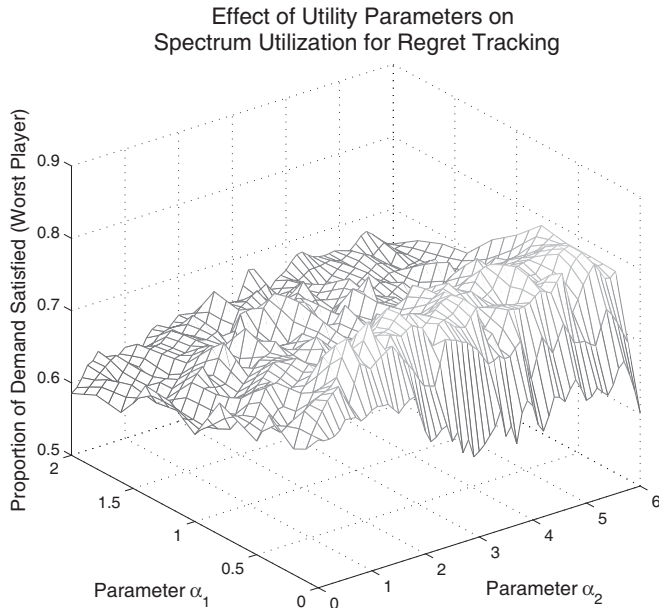


Fig. 3. The average system performance for a fixed scenario type (mild congestion) depends on the parameter choices (α_1, α_2) of the radio utility (16). The parameters can be thought of as unit prices, imposed by a system-wide authority to discourage different types of interference. The selfish (anarchy) case, $\alpha_1 = \alpha_2 = 0$, does not yield the best system performance, but left to their own devices radios may gravitate towards this case.

For each parameter choice (α_1, α_2) , 20 scenarios were generated with channel quality C_n and user demands d_n^l , selected from uniform distributions on $\{1, 2, 3\}$ and $\{1, 2, 3, 4\}$, respectively. Hence, on average there were 16 units of channel capacity available and 18 units of demand, indicating mild system congestion. We chose $m^l = 2$ for each radio, to limit the size of the decision space. For ideal operation, we initially assume that primary user activity is fixed to two specific channels. We will address varying channel occupancy in Section VI-C.

For each choice (α_1, α_2) , system performance (12) was averaged over 1000 iterations of Algorithm 5.1 and the 20 randomly selected scenarios. The result is shown in Figure 3. System performance was highest for $\alpha_1 \approx 0.3$ and $\alpha_2 > 1$. Hence completely selfish behaviour ($\alpha_1 = \alpha_2 = 0$) is not optimal. To simulate good cooperation, we chose $(\alpha_1, \alpha_2) = (0.2, 1.8)$ for subsequent analysis.

It is interesting to ask whether a player can benefit by deviating from these cooperative parameters (set by design or by a system authority). The answer appears to be “yes”; if one player takes $(\alpha_1, \alpha_2) = (0, 0)$, while the others use $(\alpha_1, \alpha_2) = (0.2, 1.8)$, the selfish player achieves much higher satisfaction of demand (96.5% after 3000 iterations), at the expense of the system (the worst off player satisfies 76% of his demand). This points to the ultimate benefit of cooperative design; even as radios seek to maximize their own utility, this should reflect system, not individual, performance. It also suggests that the parameter-tuning “meta-game” is of the prisoner’s dilemma type, with Nash equilibrium $(\alpha_1, \alpha_2) = (0, 0)$.

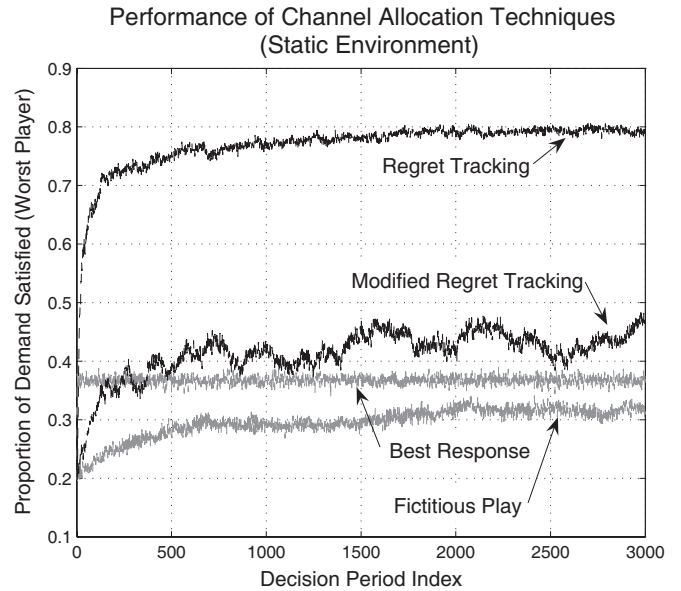


Fig. 4. Performance comparison of the channel allocation techniques of Section V, according to global design objective (12). The regret-based algorithms outperform the classical “greedy” algorithms due to their tolerance to noisy feedback. The optimal performance based on the selected scenarios is approximately 92.5%.

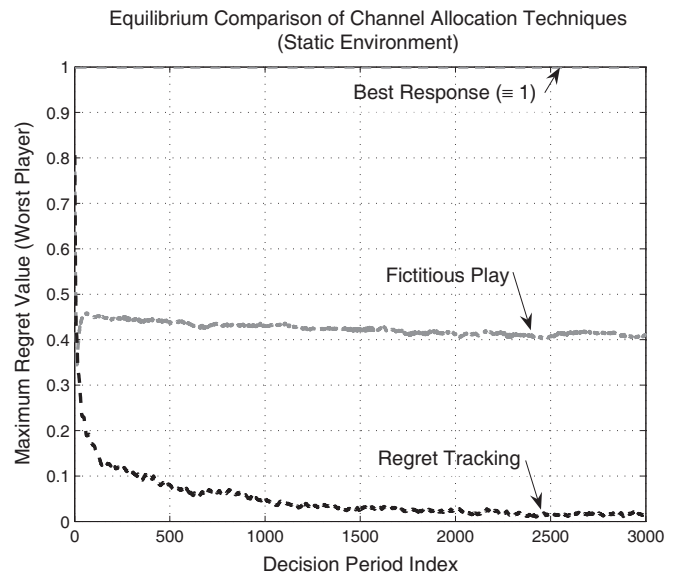


Fig. 5. Evolution of regret for the channel allocation techniques described in Section V. The regret indicates how close the system is to correlated equilibrium “competitively optimal” behaviour. The modified regret tracking algorithm has large, un-normalized regrets and is not shown.

B. Comparison of Algorithms in a Static Environment

We now compare the long-run performance of Algorithm 5.1 to other possible approaches assuming fixed cognitive radio demand d_n^l primary user activity \mathbf{Y}_n .

Simulation of Algorithm 5.1 and the three alternatives in Section V-C were performed as follows. 100 scenarios were randomly generated as in Section VI-A. Each cognitive radio ran the same algorithm over 3000 iterations, (CSMA access, channel contention estimation, and adaptive channel allocation). The results are shown in Figures 4 and 5.

As shown, the regret tracking algorithm achieves the best

performance, and most closely approaches the set of correlated equilibria of the spectrum access game (has the lowest regret). High utility was designed to correspond with good system performance, so it is not surprising that locally optimal (i.e. equilibrium) behaviour performs well.

Modified regret tracking performed significantly poorer, since radio awareness is severely impaired in this procedure (radios do not know the contention on a channel until they use it). Analysis reveals that the random exploration required by the modified procedure is chiefly to blame for its poor performance. (Exploration is done indifferently between good and bad actions, and may drastically impact the performance of other radios.)

The two remaining procedures, fictitious play and best response, performed surprisingly poorly, not even approaching 50% spectrum utilization. The problem here can be traced to the fact that fictitious play and regret tracking choose the *best* action, whereas regret tracking chooses randomly from among all *better* actions. This greedy behaviour performs poorly under noisy utility measurements, as is the case here, since users tend to overreact to bad information. This hypothesis was validated by running simulations with perfect utility information. In this case, all algorithms performed equally well.

C. Performance of Regret Tracking in a Dynamic Environment

Here we analyze the performance of Algorithm 5.1 with constant stepsize $\varepsilon = 0.1$ when the primary user activity and radio demands vary in time. We simulated 2 primary users, $C = 10$ channels and $L = 6$ cognitive radio users, with uniformly distributed channel qualities and demand levels. At each decision period, we assume that a total of w system parameters (e.g., demand levels d^l or primary user channel occupations) jump change independently with probability ρ . So the expected duration time between jump changes that the system parameters remain constant is: $T(\rho) = (1 - (1 - \rho)^w)^{-1}$. If a demand level changes, it is recomputed from the uniform distribution; if a primary user changed channels, the new channel was chosen uniformly from among the eight currently unoccupied channels. We allowed each channel quality C_n to be recomputed with probability ρ (slowly varying statistics), with $C_n(i)$ allowed to fluctuate randomly up to $\pm 10\%$ in each decision period (fast varying activity).

The performance of Algorithm 5.1 in this dynamic environment for various values of ρ is shown in Figure 6. We chose $w = 16$ when fast primary users are present, and $w = 8$ when only slow primary users are present. As expected, average utilization is higher for slower changes, since radios have more time to adapt to their environment. The regret tracking algorithm outperforms the three other algorithms (Best Response, Fictitious Play and Modified Regret Tracking) even in fast changing environments (small $T(\rho)$).

D. Imperfect Observations of Primary User Activity

So far we assumed that cognitive radio users observe the channel activity of primary users without error. This is a reasonable assumption, for example, if primary users broadcast with high power compared to cognitive radio users, as

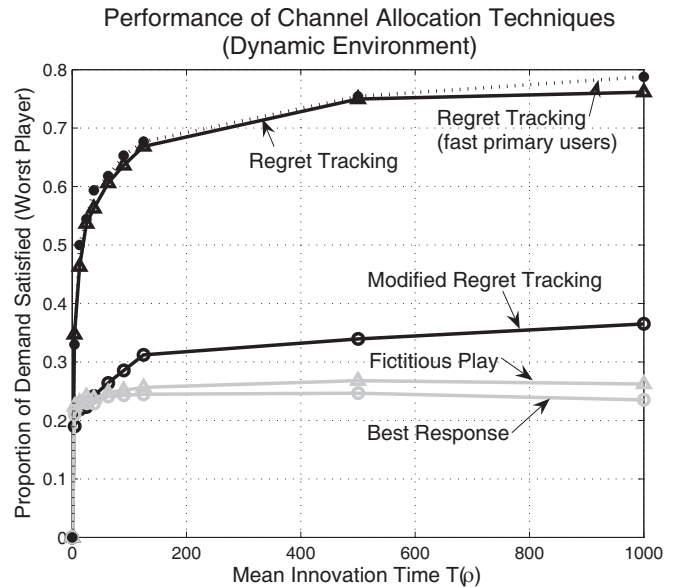


Fig. 6. Long-run average spectrum utilization in a dynamic environment for the channel allocation techniques of Section V. $T(\rho)$ (see Sec. VI-C) is the mean time between innovations in the system (changes in primary user activity, fast primary user statistics or cognitive radio demands).

such broadcasts are easily detectable in the listen-before-talk CSMA schemes. In this section, we investigate the performance of the proposed algorithm when primary user activity is *imperfectly* observed. We assume that the primary user activity on each radio channel evolves according to a Markov chain: $a_n \in \{\text{active}, \text{inactive}\}$ at time n with transition probability matrix $A = \begin{bmatrix} 0.995 & 0.005 \\ 0.005 & 0.995 \end{bmatrix}$. Assume that each cognitive radio assesses occupation of each channel $i = 1, 2, \dots, C$ by primary users based on noisy observations y_n of the primary user activity. Denote the error detection probability $P_e = \mathbf{P}(y_n = \text{inactive} | a_n = \text{active})$ and $Q_e = \mathbf{P}(y_n = \text{active} | a_n = \text{inactive})$. For convenience we set $P_e = Q_e$ in our simulations. The cognitive radio users then use a Hidden Markov Model (HMM) filter [28] together with A, P_e to compute a Bayesian estimate $\mathbf{P}(a_n = \text{active} | y_1, \dots, y_n)$, i.e., the probability that the channel is actually occupied. The utility assessed on each channel is modified to reflect the *expected* amount of throughput, which is the throughput as defined in (13) multiplied by the a posteriori probability of the channel being occupied from the HMM filter. Figure 7 shows the decrease in global network performance as error probability P_e increases. Thus Algorithm 5.1 requires at least reasonably accurate estimates of primary user activity to attain satisfactory system performance. In this scenario, game conditions change relatively quickly. Therefore it is understandable that global network performance does not attain the performance when perfect measurements of primary user activity are available.

VII. CONCLUSIONS

We have presented a game-theoretic approach for cognitive radio dynamic spectrum access which seeks out and shares temporarily vacant radio spectrum between multiple users. Our approach is completely decentralized in terms of both radio awareness and activity; radios estimate spectral conditions

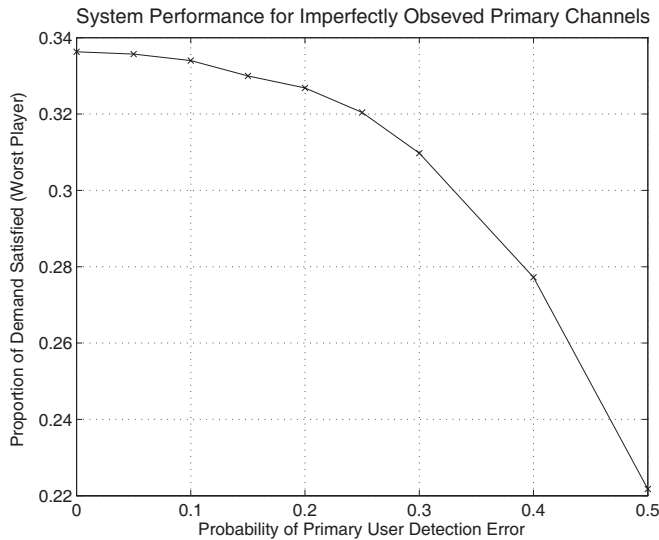


Fig. 7. Performance for imperfectly observed primary user activity. The probability of primary detection error P_e is defined in Sec.VI-D. Channel occupation by primary users evolves according to a 2-state Markov chain. Cognitive radio users discount their expected reward on each channel by the probability of a primary user occupying that channel, estimated using a hidden Markov Bayesian filter.

based on their own experience, and adapt by choosing spectral allocations which yield them the greatest utility. Iterated over time, this process converges so that each radio's performance is an optimal response to others' activity. Moreover, we are able to use this apparently selfish scheme to deliver system-wide performance by a judicious choice of utility function. For further details of learning algorithms for tracking correlated equilibria in terms of switched differential inclusions, see [29]. More recently, we have developed a Markovian dynamical games approach for cognitive radio systems in [30].

REFERENCES

- [1] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," *Mob. Netw. Appl.*, vol. 6, no. 5, pp. 435-441, 2001.
- [2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Select. Areas Commun.*, vol. 23, no. 2, pp. 201-220, 2005.
- [3] S. Sankaranarayanan, P. Papadimitratos, A. Mishra, and S. Hershey, "A bandwidth sharing approach to improve licensed spectrum utilization," in *Proc. IEEE DySPAN 2005*, pp. 279-288, 2005.
- [4] W. Wang and X. Liu, "List-coloring based channel allocation for open-spectrum wireless networks," in *Proc. IEEE VTC*, 2005.
- [5] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Mag.*, May 2007.
- [6] D. Cabric, S. Mishra, and R. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. 38th Asilomar Conf. Signals, Syst., Computers*, pp. 772-776, 2004.
- [7] K. Challapali, S. Mangold, and Z. Zhong, "Spectrum agile radio: detecting spectrum opportunities," in *Proc. International Symposium Advanced Radio Technol.*, 2004.
- [8] B. Wild and K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in *Proc. IEEE Symposium New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005.
- [9] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. IEEE Symposium New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005.
- [10] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive mac for opportunistic spectrum access in ad hoc networks: a pomdp framework," *IEEE J. Select. Areas Commun.: Special Issue Adaptive, Spectrum Agile Cognitive Wireless Networks*, Apr. 2007.
- [11] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 2053-2071, May 2008.

- [12] T. Weiss and F. Jondral, "Spectrum pooling: an innovative strategy for enhancement of spectrum efficiency," *IEEE Commun. Mag.*, pp. 8-14, Mar. 2004.
- [13] U. Berthold and F. Jondral, "Guidelines for designing OFDM overlay systems," in *Proc. First IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005.
- [14] W. Ren, Q. Zhao, and A. Swami, "Power control in spectrum overlay networks: how to cross a multi-lane highway?" in *Proc. IEEE International Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Mar. 2008.
- [15] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," in *Proc. IEEE DySPAN 2005*, pp. 269-278, 2005.
- [16] D. Famolari, N. Mandayam, D. Goodman, and V. Shah, "A new framework for power control in wireless data networks: games, utility, and pricing," *Wireless Multimedia Network Technologies*, R. Ganesh, et al., eds., Kluwer Academic Publishers, pp. 289-310, 2000.
- [17] D. Goodman and N. Mandayam, "Power control for wireless data," *IEEE Personal Commun.*, vol. 7, no. 2, pp. 48-54, Apr. 2000.
- [18] —, "Network assisted power control for wireless data," *Mobile Networks Applications*, vol. 6, no. 5, pp. 409-418, Sept. 2001.
- [19] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 1883-1892, Oct. 2001.
- [20] C. Saraydar, N. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, no. 2, 2002.
- [21] R. Aumann, "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, vol. 55, no. 1, pp. 1-18, 1987.
- [22] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127-1150, 2000.
- [23] G. Yin, V. Krishnamurthy, and C. Ion, "Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization," *SIAM J. Optimization*, vol. 14, no. 4, pp. 117-1215, 2004.
- [24] M. Benaim, J. Hofbauer, and S. Sorin, "Stochastic approximations and differential inclusions ii: Applications," UCLA Department of Economics, Levine's Bibliography, May 2005.
- [25] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. New York, NY: Springer-Verlag, 2003.
- [26] S. Hart and A. Mas-Colell, "Uncoupled dynamics do not lead to Nash equilibrium," *American Economic Rev.*, vol. 93, no. 5, pp. 1830-1836, Dec. 2003.
- [27] —, "A reinforcement procedure leading to correlated equilibrium," in *Economic Essays*. Springer, 2001, pp. 181-200.
- [28] R. Elliott, L. Aggroun, and J. Moore, *Hidden Markov Models: Estimation and Control*. New York, NY: Springer-Verlag, 1995.
- [29] V. Krishnamurthy, M. Maskery, and G. Yin, "Decentralized adaptive filtering algorithms for sensor activation in an unattended ground sensor network: a correlated equilibrium game theoretic analysis," *IEEE Trans. Signal Processing*, vol. 56, no.12, pp. 6086-6101, Dec. 2008.
- [30] J. Huang and V. Krishnamurthy, "Transmission control in cognitive radio systems with latency constraints as a switching control dynamic game," in *Proc. 47th IEEE Conf. Decision Control*, Cancun, Dec. 2008.



Michael Maskery received his B.Sc. degree in mathematics and electrical engineering from Queen's University, Kingston, Canada, in 2000 his M.Sc. degree in mathematics from the University of Ottawa, Canada in 2003, and his Ph.D. in electrical engineering from the University of British Columbia, Canada in 2008. He is currently working as a patent associate in Vancouver, Canada. His research interests include competition, cooperation and information propagation and processing among networked agents, wireless communications,

stochastic processes, network theory and game theory.



Vikram Krishnamurthy (S'90-M'91-SM'99-F'05) was born in 1966. He received his bachelor's degree from the University of Auckland, New Zealand in 1988, and Ph.D. from the Australian National University, Canberra, in 1992. He is currently a professor and holds the Canada Research Chair at the Department of Electrical Engineering, University of British Columbia, Vancouver, Canada. Prior to 2002, he was a chaired professor at the Department of Electrical and Electronic Engineering, University of Melbourne, Australia, where he also served as

deputy head of department. His current research interests include computational game theory, stochastic dynamical systems for modeling of biological ion channels and stochastic optimization and scheduling. Dr. Krishnamurthy has served as associate editor for several journals including IEEE TRANSACTIONS AUTOMATIC CONTROL, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS AEROSPACE AND ELECTRONIC SYSTEMS, IEEE TRANSACTIONS CIRCUITS AND SYSTEMS B, IEEE TRANSACTIONS NANOBIOSCIENCE, and SYSTEMS AND CONTROL LETTERS. He is co-editor of the book *Biological Membrane Ion Channels – Dynamics Structure and Applications*, published by Springer-Verlag in 2007.



Qing Zhao (S'97-M'02-SM'08) received the Ph.D. degree in Electrical Engineering in 2001 from Cornell University, Ithaca, NY. In August 2004, she joined the Department of Electrical and Computer Engineering at UC Davis where she is currently an associate professor. Prior to that, she was a communications system engineer with Aware, Inc., Bedford, MA. Her research interests are in the general area of signal processing, communications, and wireless networking.

Qing Zhao is an Associate editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING, and an elected member of IEEE Signal Processing Society SP-COM Technical Committee. She received the 2000 Young Author Best Paper Award from IEEE Signal Processing Society and the 2008 Outstanding Junior Faculty Award from UC Davis College of Engineering.