Adaptation in a Channel Access Game with Private Monitoring

Zaheer Khan [*], Janne Lehtomäki[*], Luiz A. DaSilva [†], Matti Latva-aho [*] and Markku Juntti [*]
[*]Centre for Wireless Communications (CWC) University of Oulu, Oulu, Finland
[†]Virginia Tech, USA, and Trinity College Dublin, Ireland

Abstract—**Under the opportunistic spectrum access paradigm, the shared pool of spectrum bands that the multiple autonomous cognitive radios (CRs) need to compete for is not necessarily homogeneous. The non-homogeneity in channels may lead to payoff distribution conflict among autonomous CRs, as each CR would prefer the outcome in which it selects the more desirable channels. To address this challenge, we have designed an adaptive strategy that (without explicit coordination) enables the CRs to autonomously reach an outcome that maximizes the total CR network throughput and minimizes the payoff distribution conflict among the CRs. We utilize the framework of repeated games with private monitoring to: 1) study the dynamic channel selection problem; 2) analyze the stability of the proposed strategy; and 3) investigate the impact of deviations by a selfish CR on the performance of the proposed strategy. In our model, multiple autonomous CRs are not able to observe the channel selections of other competing CRs. Rather, they get a signal from which the selections must be inferred.**

Index Terms—**Autonomous cognitive radios, game theory, non-homogenous channels, adaptation.**

I. Introduction

To help address the critical stress on scarce spectrum resources spurred by ever more powerful and more capable smart devices, a recent presidential advisory committee report and the FCC recommend the use of spectrum sharing technologies [1], [2]. One technology recommended in these reports is cognitive radio (CR), in which a network entity is able to adapt intelligently to the environment through observation, exploration and learning. A CR utilizes spectrum opportunistically by monitoring the licensed frequency spectrum to reliably detect primary user (PU) signals and operating whenever the PU is absent.

Multiple autonomous CRs often have to search a shared pool of potentially available spectrum bands for transmission opportunities, and they face competition from one another to access these bands. For instance, if a particular channel is simultaneously sensed free by two or more autonomous CRs and more than one of them decide to transmit on the channel, then a collision occurs. In this context their probability of successful access will be affected by their channel sensing order $\mathbb{P}$, i.e., the order in which radios competing for the channels visit those channels.

In this research, we consider a heterogeneous environment in which some spectrum bands may be more desirable because primary users are less likely to be active there. When multiple autonomous CRs compete for a shared pool of non-homogeneous spectrum resources, the problem of fair allocation of these resources is a major challenge. The question we seek to answer is how CRs can autonomously arrive at an outcome that maximizes the average CR network reward (the total average number of successful transmissions) in the distributed CR network in a way that also minimizes the payoff distribution conflict among autonomous CRs.

The main contributions of this paper are: 1) We propose and evaluate an adaptive Win-Shift, Lose-Randomize (WSLR) strategy that enables the CRs to maximize the total average number of successful transmissions in the network and also leads the autonomous CRs to engage in intertemporal sharing of the rewards from cooperation. The concept of fairness we focus on is envy-freeness [3], as explained in Section IV-B; 2) We formulate a repeated dynamic channel access game with private monitoring to analyze the problem of dynamic channel selection among autonomous CRs. We prove that the proposed adaptive WSLR strategy leads to a Nash Equilibrium. Much of the recent research assumes that a node operating in a network is able to perfectly perceive the actions of all other nodes [4], [5]. To overcome these limitations, our model considers the case where CRs are not able to observe the actions of other CRs; and 3) Using analytical and simulation results we compare the performance of our proposed strategy against other existing strategies.

The rest of this paper is organized as follows. Related work to our research is summarized in Section II. In Section III the system setup is presented, while the dynamic channel selection game with private monitoring is introduced in Section IV. In Section V we present, analyze and compare our proposed strategy to related strategies proposed in other works. Finally, Section VI summarizes our main conclusions.

II. Related Work

The works in [6]–[8] proposed distributed learning and allocation strategies for CRs employing a single channel sensing policy. Under a single sensing policy, CRs can explore a single channel in a given time. Our work in [9] proposed adaptive allocation strategies for both single and sequential sensing policies. Under a sequential channel sensing policy, CRs can explore more than one channel sequentially in a given time slot. However, all these works considered the problem of maximizing the total CR system throughput and ignored the payoff distribution conflict that may arise among multiple CRs due to the non-homogeneity in potentially available spectrum resources. Moreover, the works in [6]–[9] also assumed that each CR cooperatively follows the same strategy (protocol). Unfortunately, in the presence of non-homogeneous spectrum

This work was pursued under the Wireless Innovation between Finland and the US (WiFiUS) program, funded by the National Science Foundation (NSF), under grant no. 1265280, and by the Academy of Finland. It was also partially funded by the Science Foundation Ireland, grant no. 10/IN.1/I3007.

resources this assumption is not valid, as non-homogeneity of spectrum resources may induce some CRs to deviate from the protocol to maximize their own usage at the expense of the aggregate CR system throughput. It is useful to model these scenarios as a repeated game in which competition and conflict among multiple autonomous CRs searching multiple channels for spectrum opportunities is analyzed [10], [11].

Much of the recent research has utilized the framework of repeated games with perfect monitoring to study the problem of dynamic spectrum access [4], [5]. In this class of repeated games, it is assumed that players can observe the other players' actions directly. However, the assumption that a CR has perfect observation of the actions of their opponents lacks consideration of practical constraints imposed by autonomous CRs operating in a wireless network. To address this challenge, we utilize the framework of the repeated games with private monitoring. In this class of repeated games, players do not have perfect observation of other players' actions. In our model, each autonomous CR is required to infer the actions of other CRs based on feedback (signals) from its receiver (as explained in Sections III and IV).

III. System Model

We examine a multichannel CR network in which a set of $\mathcal{N} = \{1,2,...,N\}$ autonomous CRs have a set of $\mathcal{M} = \{1,2,...,M\}$ potentially available channels. Each CR can sense only one channel at a time and, due to hardware constraints, at any given time each CR can either sense or transmit, but not both. Our work in [9] investigates the impact of different PU channel occupancy models on various adaptive sensing order selection strategies adopted by CRs and finds that these strategies are not strongly affected by the stochastic model of the PU behavior. In this work, for simplicity, we assume that for each channel, the PU activity in a time slot is independent of the PU activity in other time slots and is also independent of the PU activity in other channels; this (i.i.d.) model of PU channel occupancy is also adopted by [12]. The probability of the PU being present in the ith channel is $\theta_i \in (0,1)$, and each θ_i is known to the autonomous CRs. In practice, the autonomous CRs may obtain the primary user duty cycle statistics through the use of geolocation databases [13].

Without loss of generality, we assume that the channels are ordered by increasing probability of the PU being present, i.e., $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_M$. The primary users and CRs are both assumed to use a time-slotted system, and each primary user is either present in a channel for the entire time slot, or absent for the entire time slot [6], [12], [14].

The selection of the channel for opportunistic transmission is determined as follows: The CRs use the beginning of each slot to sense the channels in some order $\mathbb{P}$ (based on their sensing order selection strategies, as explained in Sections IV and V) to find a channel that is free of PU (or other CR) activity. We refer to this as the sensing stage (see Fig. 1). The CR then accesses the first vacant channel it finds, if one exists. We refer to this as the data transmission stage. Let $\mathcal{S}$ denote the set of possible sensing orders. Note that the sensing order that a CR employs can either come from the space of all permutations of M channels, or from some subset thereof. Our work in [9] shows that CRs can increase their average number of successful transmissions by adaptively selecting sensing orders from a predefined Latin Square of M channel indices (as compared to when they select sensing orders from the space of all permutations of M channels). Hence, we consider the case of CRs selecting sensing orders from a common predefined Latin Square. A Latin Square is an M by M matrix whose entries consist of M symbols such that each symbol appears exactly once in each row and each column. Note that when CRs select sensing orders from a Latin Square, $|\mathcal{S}| = M$, and two or more CRs can collide only if they select the same sensing order.

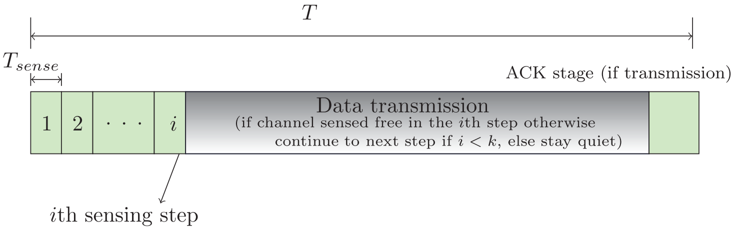
T
T_{sense}
ACK stage (if transmission)
1
2
$\cdots$
i
Data transmission
(if channel sensed free in the ith step otherwise continue to next step if $i < k$, else stay quiet)
ith sensing step

Fig. 1: Time slot structure with sensing, data transmission and acknowledgement stages.

The sensing stage in each slot is divided into a number of sensing steps. Each sensing step is used by a CR to sense a different channel. In practice, improving the sensing accuracy implies increasing the sensing duration, whereby CRs may not be able to sense all the channels within the duration of a slot. We evaluate our proposed strategy for the scenario where the number of sensing steps k that a CR can utilize in a given time slot varies from 1 to M. If a CR finds a channel free in its ith sensing step, it transmits in that channel. However, if in all sensing steps channels are found to be busy, then the CR stays silent for the remaining duration of that time slot (see Fig. 1). When a free channel is found in the ith sensing step, the durations of the sensing stage and data transmission plus acknowledgement stage are iT_{sense} and $T - iT_{sense}$, respectively, where T_{sense} is the time required to sense each channel, T is the total duration of each slot and $T >> T_{sense}$. When multiple autonomous CRs search multiple potentially available channels for spectrum opportunities, then from an individual CR perspective one of the following three events will happen in each sensing step: 1) The CR visits a given channel and is the only one to find it free and transmit; the CR then has the channel for itself for the remainder of the time slot; 2) The CR visits a given channel, finds it occupied by the PU or by another CR, then it continues looking in the next sensing step; 3) The CR visits a given channel, finds it free and transmits, but so does at least one other CR; a collision occurs. A CR infers that a collision has occurred whenever it fails to receive an acknowledgement (ACK) for a transmitted data frame.

IV. A Repeated Channel Access Game with Private Monitoring

We now formally define a dynamic channel selection game with private monitoring. CR i, where $i \in \mathcal{N}$, repeatedly plays

TABLE I: The reward table for the two-CR, two-channel stage game with private monitoring. $s_1 = (1,2)$ and $s_2 = (2,1)$ represent the two sensing orders.

Payoff Matrix

CR 2

		s_1	s_2
CR 1	s_1	0 0	$(1-\theta_2)$ $(1-\theta_1)$
	s_2	$(1-\theta_1)$ $(1-\theta_2)$	0 0

the channel selection game over an infinite time horizon, $t = 0, 1, \cdots$. Autonomous CRs operating in a CR network are unsure about when precisely their interactions will end; the model of repeated games with an infinite time horizon can be used to represent such situations. In each stage (corresponding to a time slot), CR i chooses a sensing order $s_i \in \mathcal{S}$ to sense the channels sequentially for spectrum opportunities, where $\mathcal{S}$ is the set of sensing orders. In a given time slot, CRs searching for spectrum opportunities face one of the following outcomes: successful transmission, unsuccessful transmission, or no transmission (when all channels sensed by that CR were found busy). We denote the set of possible outcomes as Ξ, i.e., $\Xi = \{$Unsuccessful transmission (U), Successful transmission (T), Channels found busy $(B)\}$. At the end of each stage, a CR observes an outcome $\xi_i \in \Xi$. The action s_i and outcome ξ_i are CR i's private information. The private outcome observed by a CR in each stage depends on the current action profile $\mathbf{s}$ (the vector of current sensing order selections of N CRs). For instance, if CR 1 selects s_1, CR 2 selects s_2, and so on, then the current action profile is $\mathbf{s} = (s_1, s_2, \cdots, s_N)$.

CR i's expected reward in the stage game is given by

$$g_i\big((s_i, \mathbf{s}_{-i})\big) = \sum_{\xi_i \in \Xi} u_i(s_i, \xi_i) p(\xi_i \mid (s_i, \mathbf{s}_{-i})) \tag{1}$$

where s_i is the action of CR i, $\mathbf{s}_{-i}$ is the action profile of all other CRs, and $u_i(s_i, \xi_i)$ is the realized reward of CR i. $u_i(s_i, \xi_i)$ equals 1 if using sensing order s_i CR i transmits successfully, i.e., $\xi_i = T$, otherwise it is 0, and $p(\xi_i \mid (s_i, \mathbf{s}_{-i}))$ is the conditional probability of private outcome ξ_i. In a repeated game with private monitoring, the average reward of CR i is

$$G_i(\mathbf{S}_a, \mathbf{S}_\xi) = \lim_{\bar{T} \to \infty} \frac{1}{\bar{T}} \sum_{t=0}^{\bar{T}} u_i\big(s_i(t), \xi_i(t)\big), \tag{2}$$

where $s_i(t)$ is the action profile of CR i at time t, $\xi_i(t)$ is the private outcome at time t, $\mathbf{S}_a = \big((s_i(t), \mathbf{s}_{-i}(t))\big)_{t=0}^{\bar{T}}$, and $\mathbf{S}_\xi = \big(\xi_i(t)\big)_{t=0}^{\bar{T}}$ are the sequences of action profiles and private outcomes respectively.

A. Case Study: N=M=2

Consider a multichannel cognitive radio network in which $N = 2$ autonomous CRs have $M = 2$ potentially available channels.

This case reduces to the well-known battle of the sexes game [11], and it is simple to prove that the game admits two pure strategy and one mixed strategy Nash equilibria. The pure strategy vectors (s_1, s_2) and (s_2, s_1) are both pure strategy equilibria but, for $\theta_1 < \theta_2$, CR 1 prefers the first and CR 2 prefers the second (see Table I). The mixed strategy equilibrium is given by the *equalizing strategies* $\mathbf{p_1} = \big(\frac{(1-\theta_1)}{(2-\theta_1-\theta_2)}, \frac{(1-\theta_2)}{(2-\theta_1-\theta_2)}\big)$ and $\mathbf{p_2} = \big(\frac{(1-\theta_2)}{(2-\theta_1-\theta_2)}, \frac{(1-\theta_1)}{(2-\theta_1-\theta_2)}\big)$, where $\mathbf{p_1}$ and $\mathbf{p_2}$ are probability mass functions assigned by CRs 1 and 2 over their action spaces $\mathcal{S}$. The equalizing strategy is a strategy that produces the same average reward no matter what the opponent does.

In the stage game, an asymmetric action profile corresponds to orthogonal sensing orders, i.e., each CR picks a different action. When $N \le M$ there are M^N total possible outcomes and (out of these total outcomes) there are $\frac{M!}{(M-N)!}$ asymmetric outcomes.

B. Envy-ratio in the Proposed Game

We study the problem of efficient and fair utilization of potentially available channels that may offer different rewards due to their non-homogeneity. The concept of fairness we focus on is envy-freeness [3]. An outcome is envy-free if no CR prefers the expected reward of another CR to its own, i.e., an envy-free outcome equalizes everyone's rewards.

We next define the *envy-ratio* of CR i for CR j as follows.

Definition 1: In an action profile $\mathbf{s}$, the *envy ratio* of CR i for CR j is the ratio of the reward obtained by j to the reward obtained by i. It is given as

$$\varepsilon_{ij}(\mathbf{s}) = \frac{g_j\big(s_j, \mathbf{s}_{-j}\big)}{g_i\big(s_i, \mathbf{s}_{-i})\big)}, \text{ for } g_i\big((s_i, \mathbf{s}_{-i})\big) > 0 \tag{6}$$

In the repeated game, the *average envy ratio* of CR i for CR j is given by

$$\Upsilon_{ij}(\mathbf{S}_a, \mathbf{S}_\xi) = \frac{G_j(\mathbf{S}_a, \mathbf{S}_\xi)}{G_i(\mathbf{S}_a, \mathbf{S}_\xi)}, \tag{7}$$

The *highest average envy ratio* between any pair of CRs is given as

$$\Upsilon(\mathbf{S}_a, \mathbf{S}_\xi) = \max\{\Upsilon_{ij}(\mathbf{S}_a, \mathbf{S}_\xi),\ i, j \in \mathcal{N},\ i \ne j\} \tag{8}$$

Note that Υ in some sense indicates the worst-case fairness for $\mathbf{S}_a$ and $\mathbf{S}_\xi$. Note also that an outcome is envy-free if $\Upsilon(\mathbf{S}_a, \mathbf{S}_\xi) = 1$.

The envy ratio between a pair of autonomous CRs does not depend only on the selection of sensing orders by the given pair, but also on the selection of sensing orders by other autonomous CRs in the network. To illustrate this situation, we can construct an example, for $N = 3$ CRs and $M = 5$ potentially available channels.

Example 4.1: Let $\Theta = (0.2, 0.3, 0.5, 0.5, 0.5)$ represent the primary user duty cycle statistics vector for channels 1 to 5 respectively, and $\mathcal{S} = \{s_1, s_2, s_3, s_4, s_5\} = \{(1,2,3,4,5), (2,3,4,5,1), (3,4,5,1,2), (4,5,1,2,3), (5,1,2,3,4)\}$ represent the set of available sensing orders. Autonomous CRs are able to sense two channels, i.e., $k = 2$, in a given time slot and CRs independently select sensing orders. The expected reward values of the stage game for CRs 1 and 2 when CR 1 selects s_1, CR 2 selects s_3, and CR

Adaptive WSLR Strategy

1) Initialize, $\mathbf{p}=[\frac{1}{N},\frac{1}{N},\cdot,\frac{1}{N}]$, an N-element probability vector $\mathbf{p}$ (all components are nonnegative and add to 1), i.e., the CR utilizes independent and random (with an equal probability) selection.

2) Toss a weighted coin to select a sensing order, with p_i the probability of selecting sensing order i. Sense the channels sequentially in the order given in the selected sensing order.

3) One of three possibilities occurs:

a) *Successful transmission:* On a successful transmission using the current sensing order i, the CR updates $\mathbf{p}$ as $p_i = 1$,

where $i = (i \bmod N) + 1$ and $p_j = 0$, $\forall i \neq j$, i.e., it shifts to the next sensing order to visit the channels in the next slot. The CR then returns to 2.

b) *CR finds all channels busy:* On finding all channels busy using the current sensing order i, the CR updates $\mathbf{p}$ as $p_i = 1$, where $i = (i \bmod N) + 1$ and $p_j = 0$, $\forall i \neq j$, i.e., it shifts to the next sensing order to visit the channels in the next slot. The CR then returns to 2.

c) *CR transmits but no ACK is received:* When the CR transmits but it receives no ACK in the current slot using sensing order i then the CR returns to 1.

Fig. 2: The Win-shift, lose-randomize (WSLR) strategy.

3 selects s_5, i.e., $\mathbf{s} = (s_1, s_3, s_5)$, are given as

$$\begin{aligned} g_1\big((s_1,\mathbf{s}_{-1})\big) &= (1-\theta_1)+\theta_1(1-\theta_2) = 0.94, \\ g_2\big((s_3,\mathbf{s}_{-3})\big) &= (1-\theta_3)+\theta_3(1-\theta_4) = 0.75. \end{aligned} \tag{9}$$

Using (6) and (9), $\varepsilon_{21}(\mathbf{s}) = 0.94/0.75$. However, if CR 3 selects s_2, i.e., $\mathbf{s}'' = (s_1, s_3, s_2)$, then $\varepsilon_{21}(\mathbf{s}'') = 0.8/0.75$, i.e., the envy ratio of CR 2 for CR 1 decreases for $\mathbf{s}''$. This is due to the reason that when CR 3 selects s_2 the probability of success of CR 1 is decreased as CR 3 can now find channel 2 free before CR 1, if it is free. On the other hand, if CR 3 selects s_4, i.e., $\mathbf{s}''' = (s_1, s_3, s_4)$, then $\varepsilon_{21}(\mathbf{s}'') = 0.94/0.5$, i.e., the envy ratio of CR 2 for CR 1 increases for $\mathbf{s}'''$. This is due to the reason that when CR 3 selects s_4 the probability of success of CR 2 is decreased as CR 3 can now find channel 4 free before CR 2, if it is free. Hence the envy ratio of CR 2 for CR 1 does not depend only on the selection of sensing orders by 2 and 1, but on the selection of sensing order by the other autonomous CR 3.

We can then state the following result.

Proposition 4.1: In the proposed sensing order selection game, the highest envy ratio is $\frac{(1-\theta_1)}{(1-\theta_M)}$ for the scenarios where $N = M$.

Proof: When $N = M$ then in any asymmetric action profile CRs can only find a free channel in the first sensing step. The envy ratio is highest between the pair of CRs, one of which selects the sensing order with the best channel in its first step and the other CR selects the worst channel in its first step, i.e., $\frac{(1-\theta_1)}{(1-\theta_M)}$. ■

V. An Adaptive WSLR Strategy

In this section, we propose an adaptive Win-shift, lose-randomize (WSLR) strategy for the autonomous channel selection, where adaptations are in the autonomous choice, by CRs, of the channel sensing order. In the WSLR strategy, each CR employs a common pre-defined sequence matrix (a Latin Square) Φ to select a sensing order in which k potential channels are to be visited in a given time slot, where k takes integer values between 1 to M. For a given number of channels M there can be many Latin Squares [15]. To select a sensing order from a common predefined Latin Square, CRs can employ any of the many Latin Squares. However, to make the analysis tractable, we assume that each CR employs a circulant matrix (which is an example of a Latin Square). A circulant matrix associated to M is the $M \times M$ matrix whose rows are given by the iterations of the shift operator acting on M. Such a matrix will be denoted by $\Phi = \mathrm{circ}\{1, 2, \cdots, M\}$, where $1, 2, \ldots, M$ are the channel indices which are ordered by increasing probability of the PU being present, i.e., $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_M$. For example, with $M = 4$, the matrix Φ is given as:

$$\Phi = \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{pmatrix}$$

For efficient channel utilization, we consider the scenarios where the N CRs utilize the N top rows of Φ for the selection of sensing orders. This is reasonable as the channel indices $1, 2, \ldots, M$ are ordered by increasing probability of the PU being present, hence the top N rows of Φ dominate in terms of having channels (in their initial columns) where PU's are less likely to be present. Note that for $N = M$, the entire matrix Φ of sensing orders is utilized by a CR for the selection of sensing orders. Let $\mathbf{S}_N$ represent the matrix of the top N rows of Φ.

The WSLR strategy is described in Fig. 2. The WSLR strategy is meant to address three aims:

1) Convergence: Utilizing randomization based on observed private outcomes, the WSLR strategy leads the autonomous CRs to eventually converge to sensing orders that minimize the likelihood of collisions among CRs. When N CRs independently and randomly (with equal probability) select a sensing order (among N sensing orders) in each time slot, then the probability of arriving at orthogonal sensing orders in a time slot is $\left(1/N\right)^N \left(N!\right)$, and consequently the expected time required to arrive at orthogonal sensing orders is $N^N \left(1/N!\right)$. Clearly, this random strategy is inefficient as even when a CR

The transition probability matrix for the scenario when the two CRs (with private monitoring) utilize the WSLR strategy:

$$
P = \begin{array}{c|cccccccccccc}
 & (\mathbf{s}',(U,U)) & (\mathbf{s}',(B,B)) & (\mathbf{s}'',(U,U)) & (\mathbf{s}'',(B,B)) & (\mathbf{s}''',(B,B)) & (\mathbf{s}''',(B,T)) & (\mathbf{s}''',(T,B)) & (\mathbf{s}''',(T,T)) & (\mathbf{s}'''',(B,B)) & (\mathbf{s}'''',(B,T)) & (\mathbf{s}'''',(T,B)) & (\mathbf{s}'''',(T,T)) \\
\hline
(\mathbf{s}',(U,U)) & 1/4\phi_a & 1/4\theta_1\theta_2 & 1/4\phi_b & 1/4\theta_1\theta_2 & 1/4\theta_1\theta_2 & 1/4(1-\theta_2)\theta_1 & 1/4(1-\theta_1)\theta_2 & 1/4\phi_c & 1/4\theta_1\theta_2 & 1/4(1-\theta_1)\theta_2 & 1/4(1-\theta_2)\theta_1 & 1/4\phi_c \\
(\mathbf{s}',(B,B)) & 0 & 0 & \phi_a & \theta_1\theta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
(\mathbf{s}'',(U,U)) & 1/4\phi_a & 1/4\theta_1\theta_2 & 1/4\phi_b & 1/4\theta_1\theta_2 & 1/4\theta_1\theta_2 & 1/4(1-\theta_2)\theta_1 & 1/4(1-\theta_1)\theta_2 & 1/4\phi_c & 1/4\theta_1\theta_2 & 1/4(1-\theta_1)\theta_2 & 1/4(1-\theta_2)\theta_1 & 1/4\phi_c \\
(\mathbf{s}'',(B,B)) & \phi_a & \theta_1\theta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
(\mathbf{s}''',(B,B)) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_1)\theta_2 & (1-\theta_2)\theta_1 & \phi_c \\
(\mathbf{s}''',(B,T)) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_1)\theta_2 & (1-\theta_2)\theta_1 & \phi_c \\
(\mathbf{s}''',(T,B)) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_1)\theta_2 & (1-\theta_2)\theta_1 & \phi_c \\
(\mathbf{s}''',(T,T)) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_1)\theta_2 & (1-\theta_2)\theta_1 & \phi_c \\
(\mathbf{s}'''',(B,B)) & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_2)\theta_1 & (1-\theta_1)\theta_2 & \phi_c & 0 & 0 & 0 & 0 \\
(\mathbf{s}'''',(B,T)) & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_2)\theta_1 & (1-\theta_1)\theta_2 & \phi_c & 0 & 0 & 0 & 0 \\
(\mathbf{s}'''',(T,B)) & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_2)\theta_1 & (1-\theta_1)\theta_2 & \phi_c & 0 & 0 & 0 & 0 \\
(\mathbf{s}'''',(T,T)) & 0 & 0 & 0 & 0 & \theta_1\theta_2 & (1-\theta_2)\theta_1 & (1-\theta_1)\theta_2 & \phi_c & 0 & 0 & 0 & 0
\end{array} \qquad (10)
$$

where $\mathbf{s}' = (s_1,s_1)$, $\mathbf{s}'' = (s_2,s_2)$, $\mathbf{s}''' = (s_1,s_2)$, $\mathbf{s}'''' = (s_2,s_1)$, $\phi_a = (1-\theta_1)+\theta_1(1-\theta_2)$, $\phi_b = (1-\theta_2)+\theta_2(1-\theta_1)$ and $\phi_c = (1-\theta_1)(1-\theta_2)$.

Reward vectors $\hat{\mathbf{g_1}}$ and $\hat{\mathbf{g_2}}$ (associated with the states of the Markov chain) for CRs 1 and 2 respectively are given as

$$
\hat{\mathbf{g_1}} = \begin{array}{cccccccccccc}
(\mathbf{s}',(U,U)) & (\mathbf{s}',(B,B)) & (\mathbf{s}'',(U,U)) & (\mathbf{s}'',(B,B)) & ((s_1,s_2),(B,B)) & ((s_1,s_2),(B,T)) & ((s_1,s_2),(T,B)) & ((s_1,s_2),(T,T)) & ((s_2,s_1),(B,B)) & ((s_2,s_1),(B,T)) & ((s_2,s_1),(T,B)) & ((s_2,s_1),(T,T)) \\
(\; 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \;)
\end{array} \qquad (11)
$$

$$
\hat{\mathbf{g_2}} = \begin{array}{cccccccccccc}
(\mathbf{s}',(U,U)) & (\mathbf{s}',(B,B)) & (\mathbf{s}'',(U,U)) & (\mathbf{s}'',(B,B)) & ((s_1,s_2),(B,B)) & ((s_1,s_2),(B,T)) & ((s_1,s_2),(T,B)) & ((s_1,s_2),(T,T)) & ((s_2,s_1),(B,B)) & ((s_2,s_1),(B,T)) & ((s_2,s_1),(T,B)) & ((s_2,s_1),(T,T)) \\
(\; 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \;)
\end{array} \qquad (12)
$$

attains a singleton status, i.e., the sensing order it has selected was not selected by another CR, it randomizes and with high probability it may lose the singleton status in the next time slot. In contrast to that, the WSLR strategy requires that singleton CRs should shift and non-singleton CRs should randomize. This reduces the number of CRs that randomly select a sensing order in the next time slot and hence increases the probability of arriving at orthogonal sensing orders.

2) Intertemporal sharing of rewards: Since different sensing orders may result in different rewards, intertemporal sharing of the sensing orders among autonomous CRs is achieved by allowing a CR to shift to the next sensing order if it has not observed an unsuccessful transmission, i.e., private outcome (U), in the previous time slot.

3) Discourage deviations: To discourage deviations, i.e., the CRs that select the sensing orders with higher rewards may prefer to again select those sensing orders in the next rounds, some punishment mechanism must be devised. This is achieved by triggering a switch to the randomization phase when an unsuccessful transmission is observed. Section V-A will further describe how the proposed mechanism discourages deviations by any of the autonomous CRs.

A. Analysis of the Adaptive WSLR Strategy

The state of the dynamic game at each time slot is characterized by the tuple $\omega(t) = (\mathbf{s}(t),\xi(t))$, where $\mathbf{s}(t)$ is the action profile at time t and $\xi(t)$ is the associated vector of private outcomes at time t. Let the state space of the game be represented by $\Omega = \{\omega = (\mathbf{s},\xi) \mid \mathbf{s} \in \mathcal{S}_p, \xi \in \Psi\}$, where $\mathcal{S}_p$ is the set of possible action profiles and Ψ is the set of possible vectors of private outcomes.

Two-CR Two-channel Scenario: The proposed WSLR strategy for the two-CR, two-channel scenario naturally lends itself to analysis using Markov chains with rewards [16]. We represent the mechanism of transitions between states ω by a Markov chain, with transition probabilities denoted by $P_{\omega\omega'}$.

We show, in Eq. 10, that the Markov chain above is ergodic unichain. The steady state reward per step for a CR i is, then, independent of the starting state and is given by

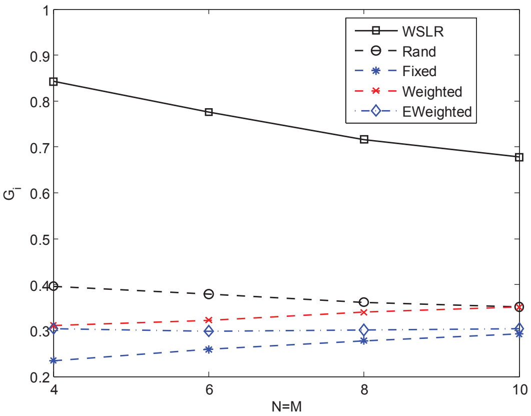

Fig. 3: Expected reward per time slot of the CR i as a function of $N = M$ CRs for different scenarios. $\Theta = (0.1,0.1,0.2,0.2,0.3,0.3,0.5,0.5,0.5,0.5)$ represents the primary user duty cycle statistics vector for channels 1 to M respectively.

$$G_i^{(\nu_i,\nu_{-i})} = \sum_{j\in\Omega} \delta_j \hat{g}_{j,i} \qquad (13)$$

where ν_i is the strategy of CR i and ν_{-i} is the strategy of the other CR, δ_j is the steady state probability of the jth state and $\hat{g}_{j,i}$ is the reward associated with the jth state for an individual CR i. When the two CRs utilize the adaptive WSLR strategy there are $\mid \Omega \mid= 12$ states of the game.

Proposition 5.1: The WSLR strategy for the two-CR two-channel scenario (when adopted by both CRs) is a Nash Equilibrium (please see a remark below).

Proof: By constructing the transition probability matrix of the Markov chain (see Eq. 10) one can see that the first four states form a transient class and the remaining eight states form a recurrent class. In a given time slot if two CRs select the same sensing order the network is in one of the transient states and the reward associated with these states is zero. This is due to the reason that, when two CRs select the same sensing order then in their first or second sensing step either both CRs will

TABLE II: Total average reward per time slot in the CR network and highest average envy ratio between a pair of CRs in the network as a function of $N = M$ for different strategies. $\Theta = (0.1, 0.1, 0.2, 0.2, 0.3, 0.3, 0.5, 0.5, 0.5, 0.5)$ represents the primary user duty cycle statistics vector for channels 1 to M respectively.

	$N = M = 6$		$N = M = 8$		$N = M = 10$	
	$\sum_{i=1}^{N} G_i$	Υ	$\sum_{i=1}^{N} G_i$	Υ	$\sum_{i=1}^{N} G_i$	Υ
rand − C	4.79	$\frac{0.9}{0.69} = 1.3$	5.8	$\frac{0.9}{0.49} = 1.8$	6.8	$\frac{0.9}{0.5} = 1.8$
WSLR	4.77	$\frac{0.79}{0.79} = 1$	5.8	$\frac{0.725}{0.725} = 1$	6.8	$\frac{0.68}{0.68} = 1$
Rand	2.27	1	2.896	1	3.51	1
EWD	2.384	$\frac{0.425}{0.26} = 1.63$	3.13	$\frac{0.41}{0.29} = 1.52$	3.83	$\frac{0.39}{0.29} = 1.34$

find the same channel free of PU activity, will start transmitting in that channel and collide, or they will find both channels busy. In the case of collision, each CR will again select sensing orders randomly (with uniform probability). However, if they find both channels busy they will shift to the next sensing order with probability 1, as an autonomous CR cannot determine on its own that the sensing order it has selected was not also selected by any other CR. In a given time slot if the two CRs select orthogonal sensing orders the network will enter one of the states in the recurrent class and will remain in the recurrent class with probability 1 (since both CRs will keep switching between the two sensing orders). The expected one time slot reward associated with the states in the recurrent class is $(2-\theta_1-\theta_2)/2$. Given that the game starts in the state where both CRs randomly select the sensing orders, the expected time for the two-CRs to reach orthogonal sensing orders, which we call the time-to-orthogonalize (TTO), is simply

$$E[\text{TTO}] = 2 - \frac{(\theta_1\theta_2)}{(\theta_1\theta_2 - 1)} \tag{14}$$

Using the WSLR strategy, on average each CR will obtain zero reward for $E[\text{TTO}] - 1$ time slots and an expected reward (per time slot) of $(2-\theta_1-\theta_2)/2$ thereafter (when the CRs arrive at orthogonal sensing orders). Hence using Eqs. (10), (11), (12) and (13) the steady state reward (per time slot) for an individual CR is $G^{(WSLR,WSLR)} = (2-\theta_1-\theta_2)/2$.

Now assume that CR 1 maintains the WSLR strategy and CR 2 considers deviating. Clearly it is inefficient if the CR 2 deviates by selecting s_2 (non-preferred sensing order) for $t > 0$ time slots. Essentially, the deviations that we need to consider are those where CR 2 will attempt to use sensing order s_1 with higher probability than sensing order s_2. The possibilities for deviations by CR 2 include a) Always select s_1 (fixed deviation, FD strategy), and b) Select s_1 (initially or when an unsuccessful transmission is observed) with $p_1 > 1/2$, s_2 with $p_2 < 1/2$; otherwise, move to switching phase if transmission is successful (Non-fixed deviation, NFD strategy). In case a), CR 1 alternates between randomization and switching phases and the steady state reward (per time slot) for CR 2 is $G_2^{(FD,WSLR)} < (1-\theta_1)/2$, which is clearly less than the expected reward obtained using the WSLR strategy. In case b), when CR 1 selects sensing orders randomly (with uniform probability) then $(E[\text{TTO}] - 1)$ is independent of the selection probabilities of CR 2. Hence, on average each CR will obtain zero reward for $(E[\text{TTO}] - 1)$ time slots and an expected reward (per time slot) of $(2-\theta_1-\theta_2)/2$ (when the CRs arrive at orthogonal sensing orders), which is the same as if CR 2 would have used the WSLR strategy.

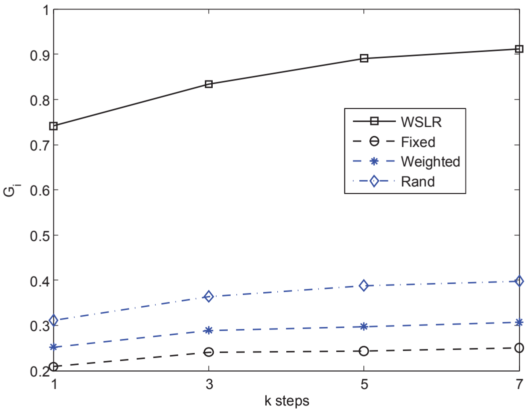

Fig. 4: Expected reward per time slot of the CR i as a function of the number of sensing steps k, with $N = 5$ CRs, $M = 8$ channels. $\Theta = (0.1, 0.1, 0.2, 0.3, 0.5, 0.5, 0.5, 0.5)$ represents the primary user duty cycle statistics vector for channels 1 to M respectively.

This checks all the possible deviations, so the WSLR strategy is a Nash equilibrium. ■

Remark Please note that we model repeated interactions of CRs with bounded rationality where strategies are represented by finite automata. Moreover, the CRs are restricted to utilizing automata with no more than two states. With such restriction the number of outcomes in equilibrium is small (see Theorem 4.3, [17]). The proposed WSLR strategy can be represented by a two-state automata in which one state is *randomization* and the other is *shifting*. Once both players have selected automata then the pair of automata forms a system which can be represented and analyzed by a finite Markov chain (as in [17]).

Deriving the proof for $N \leq M$, where $N > 2$, is challenging due to the combinatorial explosion in the number of ways that N CRs can find channels free or busy from PUs and other CRs, and also the number of ways the CRs can collide with one another. In the next section through extensive simulations we analyze the performance of the WSLR strategy for $N \leq M$ CRs with private monitoring.

B. Simulation Results

Using simulation our aim is to compare the performance (e.g., in terms of total average reward per time slot in the CR network $\sum_{i=1}^{N} G_i$, expected reward of a CR per time slot G_i, and the maximum envy ratio between a pair of CRs Υ) of the WSLR strategy against: 1) When all CRs utilize random selection of sensing orders, *Rand strategy*; 2) the randomize after every collision (rand-C) strategy. In the rand-C strategy [6], [9], initially each CR independently and randomly (with equal probability) selects a sensing order. In the next time slots, a CR randomly (with equal probability) selects a new sensing order only if it has experienced a collision in the previous slot; otherwise, it retains the previously selected sensing order; and 3) An autonomous CR i considers deviating from the WSLR strategy while all other CRs follow the strategy. The studied deviations by the CR i are: a) Always select the preferred sensing order $s_1 = (1, 2, ..., M)$, fixed deviation (FD); b) Always select s_1 with probability $q = 0.75$ and s_2 with probability $(1-q)$, weighted deviation (WD); and c) Always select s_1 with probability $q = 0.75$ and $s_2, s_3, ..., s_N$ with probabilities $[\frac{(1-q)}{(N-1)}, \frac{(1-q)}{(N-1)}, ..., \frac{(1-q)}{(N-1)}]$, extended weighted deviation (EWD). Moreover, we also evaluate the effect of varying the number of sensing steps on the performance of the proposed scheme. Note that calculations for G_i are performed using 15,000 Monte Carlo runs for dynamic channel selection game using different scenarios.

Fig. 3 evaluates the expected reward per time slot achieved by a CR i using the different strategies under different scenarios. From the figure we can see that the WSLR strategy achieves the highest expected reward per time slot for the CR i as compared to other strategies. Note that in Fig. 3 the loss in the expected reward is due to the non-homogeneity in channel availability statistics. The availability probabilities of the first 5 channels are at least 70%, and the availability probabilities of the last five channels are around 54%. Hence, as $N = M$ increases, the expected reward of the CR i decreases, as with the increasing number of CRs the number of potentially available channels also increases but with high probability of a PU being present. In Table II, we evaluate different strategies in terms of the total average reward per time slot in the CR network and the highest envy ratio between a pair of CRs. Table II shows that the WSLR strategy performs equally well as the rand-C strategy in terms of maximizing the total average reward per time slot and performs significantly better in terms of ensuring envy-freeness among the competing CRs. Fig. 4 evaluates the effect of varying the number of sensing steps on the performance of the different strategies in terms of expected reward of the CR i per time slot. It can be seen in Fig. 4 that with the increasing number of sensing steps when all the CRs utilize the WSLR strategy then the expected reward per time slot of a CR increases more as compared to the other strategies.

VI. Conclusions

We have studied the problem of coexistence among multiple autonomous CRs sharing a common pool of potentially available channels which may offer different rewards due to their non-homogeneity. In our model, autonomous CRs sense the channels sequentially (in some sensing order) for spectrum opportunities, where they are unable to observe the actions of other CRs. For efficient co-existence, we design an adaptive WSLR strategy that does not require coordination from a centralized entity and utilizes feedback (signals) to infer the actions of other CRs. We utilize the framework of the repeated games with private monitoring for the study of dynamic channel selection among autonomous CRs. We have shown that for the two-CR two-channel scenarios, the proposed strategy is a Nash equilibrium. For $N \leq M$, we have shown that the proposed strategy maximizes the total average number of successful transmissions in the network. It also ensures fairness by allowing the autonomous CRs to engage in intertemporal sharing of the non-homogenous rewards from cooperation as compared to other strategies.

References

[1] E. E. Schmidt and C. Mundie, “Realizing the full potential of government-held spectrum to spur economic growth,” Online, July 2012.

[2] J. Melvin, “US regulators ok T-mobile testing of shared use of airwaves,” Online, August 2012.

[3] R. J. Lipton, E. Markakis, E. Mossel, and A. Saberi, “On approximately fair allocations of indivisible goods,” in *Proceedings of the 5th ACM Conference on Electronic Commerce*, ser. EC ’04, 2004, pp. 125–131.

[4] R. Etkin, A. Parekh, and D. Tse, “Spectrum sharing for unlicensed bands,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 517–528, 2007.

[5] W. Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, “Repeated open spectrum sharing game with cheat-proof strategies,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1922–1933, 2009.

[6] A. Anandkumar, N. Michael, and A. Tang, “Opportunistic spectrum access with multiple users: Learning under competition,” in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, 2010, pp. 1–9.

[7] K. Liu and Q. Zhao, “Distributed learning in cognitive radio networks: Multi-armed bandit with distributed multiple players,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. 3010–3013.

[8] Y. Gai, B. Krishnamachari, and R. Jain, “Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation,” in *Proceedings of the IEEE International Dynamic Spectrum Access Networks (DySPAN)*, 2010, pp. 1–9.

[9] Z. Khan, J. Lehtomaki, L. DaSilva, and M. Latva-aho, “Autonomous sensing order selection strategies exploiting channel access information,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 2, 2013.

[10] M. Felegyhazi and J. P. Hubaux, “Game theory in wireless networks: a tutorial,” EPFL, LCA-REPORT-2006-002, Tech. Rep., 2006.

[11] M. J. Osbourne, *An Introduction to Game Theory*. Oxford University Press, 2004.

[12] R. Fan and H. Jiang, “Optimal multi-channel cooperative sensing in cognitive radio networks,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, pp. 1128–1138, Mar. 2010.

[13] A. Ghasemi and E. S. Sousa, “Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs,” *IEEE Communications Magazine*, vol. 46, no. 4, pp. 32–39, Apr. 2008.

[14] H. Li, “Multi-agent Q-learning for Aloha-like spectrum access in cognitive radio systems,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, pp. 1–15, Apr. 2010.

[15] C. F. Laywine and G. L. Mullen, *Discrete Mathematics using Latin Squares*, 1st ed., ser. Wiley-Interscience Series in Discrete Mathematics and Optimization. New York: John Wiley & Sons, 1998.

[16] R. G. Gallager, *Discrete Stochastic Processes*. Kluwer, Boston, 2001.

[17] J. Romero, “Finite automata in undiscounted repeated games with private monitoring,” Online, 2011.