Retrospective Spectrum Access Protocol: A Payoff-based Learning Algorithm for Cognitive Radio Networks

Stefano Iellamo, Lin Chen, Marceau Coupechoux

Retrospective Spectrum Access Protocol: A Payoff-based Learning Algorithm for Cognitive Radio Networks

Stefano Iellamo* Lin Chen† Marceau Coupechoux*
* Telecom ParisTech - LTCI CNRS 5141
46, rue Barrault, Paris 75013, France
{iellamo,coupecho}@enst.fr
†Laboratoire de Recherche en Informatique (LRI)
University of Paris-Sud XI, 91405 Orsay, France
chen@lri.fr

***Abstract*—Decentralized cognitive radio networks (CRN) require efficient channel access protocols to enable cognitive secondary users (SUs) to access the primary channels in an opportunistic way without any coordination. In this paper, we develop a distributed retrospective spectrum access protocol that can orient the network towards a socially efficient and fair equilibrium state. With the developed protocol, each SU j chooses a channel to select based on the experienced payoff in past H_j periods. Each SU is thus supposed to be equipped with bounded memory and should make its decision based on only local observations. In that sense, the SUs behavioral rules are said to be payoff-based. The protocol also models a natural human decision making behavior of striking a balance between exploring a new choice and retrospectively exploiting past successful choices. With both analytical demonstration and numerical evaluation, we illustrate the two noteworthy features of our solution: (1) the entirely distributed implementation requiring only local observations and (2) the guaranteed statistical convergence to the equilibrium state within a bounded delay.**

I. Introduction

In decentralized cognitive radio networks (CRN), a fundamental while challenging task is the design of distributed spectrum (channel) access mechanisms enabling cognitive secondary users (SUs) to access the primary channels in an efficient way without any coordination. In this paper[1], we develop and analyze a framework of retrospective spectrum access protocols that can orient the network towards a socially efficient and fair equilibrium state. We further assume that each SU is equipped with bounded memory. Our developed retrospective spectrum access protocol has two noteworthy features: (1) the entirely distributed implementation requiring only local observations (a player doesn't even know the payoffs obtained by others) and (2) the guaranteed statistical convergence to the equilibrium state within a bounded delay.

To analyze the performance of the developed protocol, we apply the mistakes model introduced in [1], [2], [3], [4] and establish the statistical convergence of dynamics to the system equilibrium within a bounded latency $O(1/\epsilon)$. We would like to emphasize that despite our focus on CRNs, the developed learning protocol and the analysis methodology presented in this paper also provide valuable insights on the design of decentralized load balancing algorithms.

[1]This research is supported by the LIP6/LTCI project FERRARI and the ANR project NETLEARN.

The paper is organized as follows. Section II is a short literature study on distributed learning with particular focus on CRNs. Section III defines the considered system model and the related spectrum access game. Our protocol is presented in Section IV. Convergence and performance are analyzed in Sections V and VI.

II. Related Work

The problem of distributed spectrum access in CRNs has been widely addressed in the literature. A first set of papers assumes that the number of SUs is smaller than the number of channels. In this case, the problem is closely related to the classical Multi-Armed Bandit (MAB) problem [5]. Some recent work has investigated the issue of adapting traditional MAB approaches to the CRN context, among which Anandkumar *et al.* proposed two algorithms with logarithmic regret, where the number of SUs is known or estimated by each SU [6]. Complementary, other works assume large population of SUs and study the system dynamics under asymptotic assumptions. In [7], the authors propose a distributed learning procedure for spatial spectrum access which is proven to converge to a Nash Equilibrium (NE) in the asymptotic case. The analysis relies however on a random backoff mechanism, which requires the modification of the SUs packet structure for channel contention. In [8] the authors propose imitation rules that are used by a large population of SUs to converge to a Pure NE (PNE). In this paper, it is assumed that the SUs are able to capture packets transmitted by any other SU in the network. Contrary to this literature, the proposed protocol is proven to converge to a PNE regardless of the number of SUs in the system. Furthermore, it is completely distributed and does not require any additional packet fields: it can be in fact used with any decentralized random access MAC protocol such as CSMA/CA.

While our model is presented in the specific context of CRNs, we intend it more generically as a contribution to the literature on bounded rationality and learning in games with mutations, which thus far have been mostly explored in biology and economics. Relying on the by now classical [1], [2], [3], [4] and on [9], Dieckmann analyzed the evolution of conventions in a society with local interactions and mobile players [10]. Friedman and Mezzetti investigated

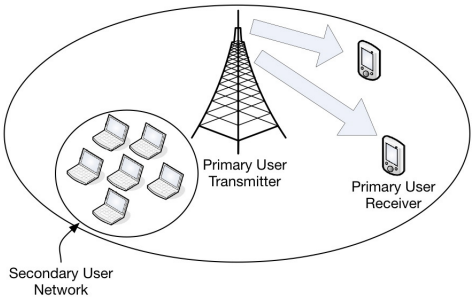

Fig. 1. Network model.

mistakes models that induce better and best reply dynamics [11]. H.P. Young *et al.* proposed a series of payoff-based rules[2] (among them, [12], [13], [14], [15], [16]) possessing several appealing properties. The version of Trial and Error presented in [15], for instance, is able to converge to the PNE maximizing the social welfare. Nevertheless, complexity is high and convergence speed is very slow, as it has been shown in [8]. The retrospective learning protocol that we propose has a similar architecture to the learning procedures developed in [12] and [17]. Our main contributions with respect to this literature is the introduction of nontrivial memories and inertia, as well as the results on convergence time.

III. System Model and Game Formulation

In this paper, we consider the downlink of primary network and SUs trying to opportunistically accessing the free spectrum (Fig. 1). The primary spectrum consists of a set $\mathcal{C}$ of C frequency channels, each with bandwidth B. The users in the primary network are operated in a synchronous time-slotted fashion. A set $\mathcal{N}$ of N SUs tries to opportunistically access the channels when they are left free by PUs.

Each SU j has a finite memory containing the history (strategies and payoffs) relative to the H_j past iterations. Let $\mathcal{H}_j$ be the history recalled by SU j. Let $\xi_i(k)$ be the random variable equal to 1 when channel i is unoccupied by the PU at slot k and 0 otherwise. We assume that the process $\{\xi_i(k)\}$ is stationary and independent for each i and k. We also assume that at each time slot, channel i is free with probability μ_i, i.e., $\mathbb{E}[\xi_i(k)] = \mu_i$. We define an iteration t as a block of PU-slots of fixed duration T during which the SUs don't change their strategy (see Fig. 2). At the end of each iteration, SUs obtain a payoff which corresponds to the achieved throughput.

In our work, each SU j is modeled as a rational decision maker, aiming at load-balancing the total system throughput. The *instantaneous* throughput it can achieve in terms of packets per second, denoted as T_j, can be expressed as a function of μ_{s_j} and n_{s_j}, where s_j denotes the channel which j chooses, and n_{s_j} denotes the number of SUs choosing channel s_j. The expected value of T_j, which has to be intended as the *long-term* throughput when T is very large, can be written as: $\mathbb{E}[T_j] = f(\mu_{s_j}, n_{s_j})$. In this paper, SUs implement a generic random access protocol to avoid collisions. This yields: $f(\mu_{s_j}, n_{s_j}) = B\mu_{s_j} p_j(n_{s_j})$, where $p_j(n_{s_j})$ is

[2]An individual's learning rule is *payoff-based* or *completely uncoupled* if it does not depend directly on the actions or payoffs of anyone else.

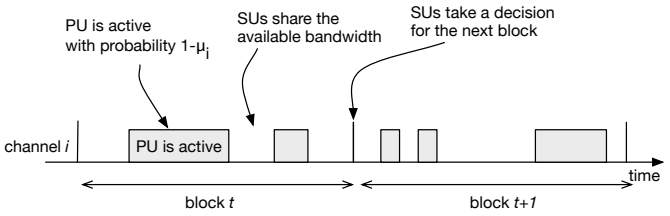

Fig. 2. SU operation on a channel $i \in \mathcal{C}$.

a *decreasing* function denoting the successful transmission probability when n_{s_j} SUs are interfering with SU j on channel s_j. B is a constant standing for the available bandwidth per channel. Without loss of generality, we will now assume that $B = 1$.

We now formulate the channel selection problem as a spectrum access game where the players are the SUs. The game is defined formally as follows:

Definition 1 (Spectrum access game)**.** *The spectrum access game $\mathcal{G}$ is a 3-tuple ($\mathcal{N}$, $\mathcal{C}$, $\{U_j(\mathbf{s})\}$), where $\mathcal{N}$ is the player set, $\mathcal{C}$ is the strategy set of each player. Let $\mathbf{s}_{-j} = \{s_1, .., s_{j-1}, s_{j+1}, ..., s_C\}$ be the channels chosen by all users except user j. When a player j chooses strategy $s_j \in \mathcal{C}$, its player-specific utility function $U_j(s_j, s_{-j})$ is defined as*

$$U_j(s_j, \mathbf{s}_{-j}) = \mathbb{E}[T_j] = \mu_{s_j} p_j(n_{s_j}).$$

The users struggle for maximizing their utility function and a commonly accepted solution for the game is a PNE:

Definition 2 (Pure Nash Equilibrium)**.** *A PNE is a point $\mathbf{s}^*$ in the action profiles space, from which no user has incentive to deviate unilaterally. Thus*

$$s_j^* \in \underset{s_j \in \mathcal{C}}{\operatorname{argmax}} U_j(s_j, \mathbf{s}_{-j}^*), \forall j \in \mathcal{N}. \quad (1)$$

We can recognize that $\mathcal{G}$ is a congestion game with player-specific payoff functions. It then follows from [18] and [8] that $\mathcal{G}$ possesses at least one PNE in the general case.

IV. Retrospective Spectrum Access Protocol

In this section, we propose a distributed retrospective spectrum access protocol (RSAP) that achieves a PNE of the spectrum access game. We firstly provide some definitions we shall need in the sequel analysis.

Define the state $z(t)$ of the system at iteration t by $z(t) \triangleq \{s_j(t-h), U_j(t-h)\}_{j \in \mathcal{N}, h \in \mathcal{H}_j}$. Let $\lambda_j = \operatorname{argmax}_{h \in \mathcal{H}_j} U_j(t-h)$ be the number of iterations passed from the SU j highest remembered payoff. Furthermore, let ρ_j denote the *inertia*, which is defined as a positive probability that SU j is unable to adjust its strategy at each iteration. Note that the concept of inertia has already been included in models of evolution with noise (see, e.g. [19]). In those cases however, inertia was defined as an exogenous parameter, while we see it as an endogenous parameter.

Notations are summarized in Table I.

We now introduce the RSAP, as detailed in Algorithm 1. At each iteration t, each user j applies the following revision scheme. With probability $(1-\epsilon(t)) \cdot (1-\rho_j)$, SU j switches to

TABLE I
NOTATIONS

Set of channels	$\mathcal{C} = \{1, .., i, .., C\}$
History of SU j	$\mathcal{H}_j = \{0, .., h, .., H_j\}$
Strategy profile at t	$\mathbf{s}(t)$
SU j strategy at t	$s_j(t)$
SU j payoff	U_j
System state at t	$z(t)$
Migration Stable State (MSS)	ω
Limit Set (LS)	L
Union of LSs	Ω
Union of all LSs in PNEs	Ω^*

Algorithm 1 RSAP: executed at each SU j

Initialization: Set $\epsilon(t)$ and ρ_j.
At $t = 0$, randomly choose a channel to stay, store the payoff $U_j(0)$ and set $U_j(t-h)$ randomly $\forall h \in \{1, .., H_j\}$.
while at each iteration $t \geqslant 1$ **do**
  With probability $1 - \epsilon(t)$ **do**
    if $U_j(t - \lambda_j) > U_j(t)$
      Migrate to channel $s_j(t - \lambda_j)$ w. p. $1 - \rho_j$
      Stay on the same channel w. p. ρ_j
    else
      Stay on the same channel
    end if
  With probability $\epsilon(t)$ switch to a random channel.
end while

channel $s_j(t-\lambda_j)$ if $U_j(t-\lambda_j) > U_j(t)$, and with probability $\epsilon(t)$ selects for the next iteration a channel with uniform distribution. To characterize the equilibrium state of RSAP, we define a Migration Stable State (MSS) as follows.

Definition 3 (Migration Stable State)**.** *A migration stable state ω is a state where no more migration is possible, i.e., $U_j(t) \geqslant U_j(t-h)$ $\forall h \in \mathcal{H}_j$ $\forall j \in \mathcal{N}$.*

V. Convergence Analysis

Foster and Young, with their pioneering work dated 1990 [1], were the first to argue that the Evolutionary Stable Strategy (ESS) does not capture the notion of long-run stability when the system is subjected to *continual* (rather than isolated) stochastic perturbations. In this new context, it is possible to identify a set of stochastically stable equilibria which consists of the states attained almost surely by a dynamical system when the noise level approaches zero. The identification of such system states is particularly useful in games with multiple equilibria as it permits to find out whether some outcomes are much more likely than others when the noise vanishes. Our protocol is characterized by stochastic perturbations and we study the small noise limit by making use of the tools provided in [9]. For the sake of a self-contained exposition, we include here some definitions and results we shall need.

A. Model of Evolution with Noise

Definition 4 (Model of evolution with noise [9])**.** *A model of evolution with noise or mistakes model is a triple $(Z, P, P(\epsilon))$ where:*

1) *Z is the state space of a stochastic process X and is supposed to be finite;*
2) *$P = (p_{zz'})_{(z,z') \in Z^2}$ is a Markov transition matrix defined on Z;*
3) *$P(\epsilon) = (p_{zz'}(\epsilon))_{(z,z') \in Z^2}$ is a family of Markov transition matrices on Z indexed by $\epsilon \in [0, \bar{\epsilon})$ s.t.:*
   a) *$P(\epsilon)$ is ergodic for $\epsilon > 0$;*
   b) *$P(\epsilon)$ is continuous in ϵ and $P(0) = P$;*
   c) *there is a cost function $c : Z^2 \to \mathcal{R}^+ \cup \{\infty\}$ s.t. for any pair of states (z, z'), $\lim_{\epsilon \to 0} \frac{p_{zz'}(\epsilon)}{\epsilon^{c_{zz'}}}$ exists and is strictly positive for $c_{zz'} < \infty$ and $p_{zz'}(\epsilon) = 0$ for small ϵ if $c_{zz'} = \infty$.*

Definition 5 (Unperturbed and perturbed Markov chain)**.** *In a model of evolution with noise $(Z, P, P(\epsilon))$, (Z, P) is called the unperturbed Markov chain and, for any ϵ, $(Z, P(\epsilon))$ is a perturbed Markov chain. The family of perturbed Markov chains indexed by ϵ is called a regular perturbation.*

Remark. The fact that $P(\epsilon)$ is ergodic ensures that from any state $z \in Z$, we can reach any state $z' \in Z$ in a finite number of steps with positive probability. The unperturbed Markov chain is however not necessarily ergodic. If not, the Markov chain (Z, P) has one or more limit sets.

Definition 6 (Limit set)**.** *A limit set (or recurrent class) L of a Markov chain $X = (Z, P)$ is a set of states of X such that $\forall z \in L$, $P[X_{t+1} \in L | X_t = z] = 1$ and $\forall z, z' \in L$, there exists $\tau > 0$ s.t. $P[X_{t+\tau} = z' | X_t = z] > 0$.*

The unperturbed Markov chain can be interpreted as the evolution of the system when players follow a predefined rule of evolution like *Best Response*. Noise ϵ can be interpreted as a probability that players do not follow the rule of the dynamics. For example, if the rule is Best Response, players choose the best response strategy at the next iteration step with probability $1-\epsilon$ and choose any other strategy at random with probability ϵ. When a player does not follow the predefined rule, we say that there is a *mutation* by analogy with what happens in species evolution.

Definition 7 (State transition cost)**.** *The cost or resistance $c_{zz'}$ of the transition $z \to z'$ is the rate at which the transition probability $p_{zz'}(\epsilon)$ tends to zero as ϵ vanishes:*

$$c_{zz'} = \begin{cases} 0 & \text{if } P_{zz'}(0) > 0 \\ k & \text{if } P_{zz'}(\epsilon) = (a_{zz'} + o(1))\epsilon^k \\ \infty & \text{if } P_{zz'}(\epsilon) = 0 \quad \forall \epsilon \in [0, \bar{\epsilon}] \end{cases}$$

for some $\bar{\epsilon} > 0$ and constants $a_{zz'}$.

Let $\mu(\epsilon)$ be the stationary probability distribution of the perturbed Markov chain $(Z, P(\epsilon))$.

Lemma 1 (Existence of limit distribution [3])**.** *There exists a limit distribution $\mu^* = \lim_{\epsilon \to 0} \mu(\epsilon)$.*

Thus, μ^* is a stationary probability of the unperturbed Markov chain (Z, P):

Lemma 2 ([9])**.** *The set of stochastically stable states is included in the limit sets of the unperturbed Markov chain (Z, P).*

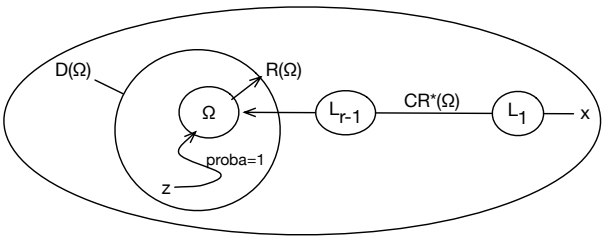

Fig. 3. Illustration of the main concepts: basin of attraction $D(\Omega)$, radius $R(\Omega)$, modified coradius $C^*(\Omega)$, L_1,...,L_{r-1} are limit sets, x is a state that maximizes the modified cost to Ω, z is a state in $D(\Omega)$.

Definition 8 (Long-run stochastically stable set)**.** *A state $z \in Z$ is said to be long-run stochastically stable if and only if $\mu_z^* > 0$.*

Let Ω be a union of one or more limit sets of (Z, P). We now want to study the conditions for Ω to be stochastically stable. We also want to know the speed at which Ω is reached. For this purpose, [9] defines $W(x, \Omega, \epsilon)$ to be the expected wait until set Ω is reached knowing that we start in state x and that the system follows the perturbed Markov chain $(Z, P(\epsilon))$. The goal is to characterize $\max_{x \in Z} W(x, \Omega, \epsilon)$.

B. Radius and Coradius Theorem

We start with some definitions of concepts illustrated in Fig. 3 before giving the main theorem. Define a path $(z_1, z_2, ..., z_\tau)$ as a sequence of states.

Definition 9 (Basin of attraction)**.** *Let Ω be a union of one or more limit sets of (Z, P) and let $(z_1, z_2, ..., z_\tau)$ be a sequence of states. The basin of attraction $D(\Omega)$ of Ω is the set of initial states from which the unperturbed Markov chain converges to Ω with probability 1, i.e.:*

$$D(\Omega) = \{z \in Z | Pr\left[\exists \beta \text{ s.t. } \forall \tau > \beta,\ z_\beta \in \Omega | z_0 = z\right] = 1\}$$

Definition 10 (Path cost)**.** *For two sets X and Y, a path in Z is a sequence of states $(z_1, z_2, ..., z_\tau)$ with $z_1, z_2, ... \in X$ and $z_\tau \in Y$. The cost of the path is the sum*

$$c(z_1, z_2, ..., z_\tau) = \sum_{i=1}^{\tau - 1} c_{z_i, z_{i+1}}.$$

Let $S(X, Y)$ be the set of all paths from X to Y and

$$C(X, Y) = \min_{(z_1, ..., z_\tau) \in S(X,Y)} c(z_1, ..., z_\tau)$$

be the set-to-set cost between X and Y. The radius of the basin of attraction of Ω is defined as the minimum number of mutations needed to leave $D(\Omega)$ given that we start in Ω.

Definition 11 (Radius)**.** *The radius $R(\Omega)$ of Ω is the minimum cost of any path from Ω out of $D(\Omega)$, i.e.:*

$$R(\Omega) = C(\Omega, Z - D(\Omega)).$$

Definition 12 (Coradius)**.** *The coradius $CR(\Omega)$ of Ω is defined by:*

$$CR(\Omega) = \max_{x \notin \Omega} \min_{(z_1, ..., z_\tau) \in S(x, \Omega)} c(z_1, ..., z_\tau).$$

In other words, the coradius is the minimum number of mutations needed to reach Ω from the most unfavorable state.

Definition 13 (Modified path cost)**.** *Let $(z_1, ..., z_\tau)$ be a path from x to Ω. Let $L_1, ..., L_r$ be a set of consecutive limit sets with $L_r \subset \Omega$ and $L_i \not\subset \Omega$ for all $i < r$, through which the path passes. The modified cost of the pathis obtained by substracting from the initial cost function the intermediate radii of the limit sets L_i:*

$$c^*(z_1, ..., z_\tau) = c(z_1, ..., z_\tau) - \sum_{i=2}^{r} R(L_i). \tag{2}$$

Definition 14 (modified coradius)**.** *The modified coradius of the basin of attraction of Ω is defined as:*

$$CR^*(\Omega) = \max_{x \notin \Omega} \min_{(z_1, ..., z_\tau) \in S(x, \Omega)} c^*(z_1, ..., z_\tau). \tag{3}$$

where $\min_{(z_1, ..., z_\tau) \in S(x, \Omega)} c^*(z_1, ..., z_\tau)$ is the modified cost between a state x and Ω.

The theorem proposed by Ellison in [9] is a sufficient condition to identify a long-run stochastically stable set of the system. It also gives an lower bound on convergence rate.

Theorem 1 (Convergence to long-run stochastically stable set with modified cost [9])**.** *Let $(Z, P, P(\epsilon))$ be a model of evolution with noise, and suppose that for some set Ω which is a union of limit sets $R(\Omega) > CR^*(\Omega)$. Then:*

1) *The long-run stochastically stable set of the model is contained in Ω.*
2) *For any $y \notin \Omega$, $W(y, \Omega, \epsilon) = O(\epsilon^{-CR^*(\Omega)})$ as $\epsilon \to 0$.*

In other words, if it is more difficult to leave Ω and its basin of attraction than to come back to it, the long-run stochastically stable set is contained in Ω.

C. Main Results

This section establishes the convergence of RSAP. To this end, we first state the following definitions required in the study of convergence.

Definition 15 (Single player improvement [11])**.** *A strategy profile $\mathbf{s}'$ is a **single player improvement** over the strategy profile $\mathbf{s}$ if it coincides with $\mathbf{s}$ in every coordinate except one, say coordinate j, and the payoff of player j is higher under $\mathbf{s}'$ than under $\mathbf{s}$.*

Definition 16 (Weak finite improvement property [11])**.** *A game $\mathcal{G}$ has the **weak finite improvement property** (weak-FIP) if from each strategy profile $\mathbf{s}$ there exists a finite sequence of single-player improvements that ends in a PNE.*

We then recall the following result from [18]:

Theorem 2 ([18])**.** *Given a congestion game $\mathcal{G}$ with player-specific decreasing payoff functions, the weak-FIP holds and $\mathcal{G}$ admits at least one Pure Nash Equilibrium.*

We now state the main results on the convergence of RSAP.

Proposition 1. *Under RSAP, LS≡MSS, i.e., all MSSs are limit sets and all limit sets are made of a single state, which is MSS, (a) in the general case with endogenous inertia $\rho_j > 0$, or (b) in the particular case $H_j = 1$ and $\rho_j = 0$, for all $j \in \mathcal{N}$.*

Proof: First note that every MSS is obviously a LS. Second, a LS is either a MSS or a set of states among which the unperturbed dynamics switches endlessly. We want to show that the latter is not possible under the proposed RSAP. Suppose by contradiction that the unperturbed dynamics is captive between two or more states. Case (a): As $\rho_j > 0$ for all $j \in \mathcal{N}$, there is a non-zero probability that all SUs does not modify their strategy during $\max_j H_j$ consecutive iterations. After such an event, the history of every SU j contains $H_j + 1$ times the same strategy. In the unperturbed dynamics, such a state is stable (MSS), i.e., the system is in an absorbing state. This contradicts our assumption and concludes the proof. Case (b): As $H_j = 1$, SU j evolves between at most two strategies: the current (at t) and the last one (at $t-1$). If they are equal, j doesn't migrate anymore (recall that we study the unperturbed dynamics). Otherwise, as $\rho_j = 0$ and following our assumption, j endlessly switches between two distinct strategies. The number of SUs with a certain strategy at the odd and even iterations is thus constant. Hence, the payoffs experienced by the switching SUs must also be constant every two iterations. This means that they are able to choose between the two strategies and stay with the selected one for ever. This contradicts our assumption and concludes the proof. ■

Remark. Every PNE includes a set of states of the system that are MSSs, i.e., LSs. Let denote Ω^* the union of all these states corresponding to the PNEs.

Proposition 2. *It holds that $R(\omega) = 1 \ \forall \omega \notin \Omega^*$, where Ω^* is the union of LSs at the PNEs and ω is a LS.*

Proof: Suppose that at iteration t the system is in state $z(t) \notin \Omega^*$ and that $z(t)$ is an MSS so that $z(t) \equiv \omega_0$. Thus, in ω_0: $U_j(t-h) \leqslant U_j(t) \ \forall h \in \mathcal{H}_j, \forall j \in \mathcal{N}$ (by definition of MSS). Recall from Theorem 2 that $\mathcal{G}$ has the weak-FIP: there exists at least one player, say j', that can improve its utility by selecting an action $s_{j'}(t+1) \neq s_{j'}(t)$. Thus, if user j' plays strategy $s_{j'}(t+1)$ at iteration $t+1$, then the system state will move from state ω_0 to state $z(t+1)$. It is easy to verify $z(t+1)$ is an MSS different from ω_0 because $U_{j'}(t+1) > U_{j'}(t)$ and $s_j(t+1) = s_j(t) \ \forall j \neq j'$. Hence, if the weak-FIP property holds, under RSAP a single mutation is enough to leave the basin of attraction of any MSS ω not in Ω^*. ■

Proposition 3. *Ω^* can be reached from any LS $L \notin \Omega^*$ by stepwise mutations.*

Proof: Recall the definition of the weak-FIP property: from each strategy profile there exists a sequence of single player improvements that terminates at a PNE after a *finite* number of steps. As shown in the proof of Proposition 2, every intermediate state is a MSS, i.e., a LS. As $\mathcal{G}$ has the weak-FIP (Theorem 2), then the proposition follows straightfully. ■

Lemma 3. *It holds that $CR^*(\Omega^*) = 1$.*

Proof: It follows from Proposition 2, Proposition 3 and the definition of the modified cost. From any state, we reach a LS at zero cost and then, there is a path of LSs towards Ω^*, each with a radius of 1. ■

Lemma 4. *It holds that $R(\Omega^*) > 1$.*

Proof: We show that a single mutation is not sufficient to leave the basin of attraction of Ω^*. As in a LS at a PNE no user has incentive to deviate unilaterally, any single mutation leads to a decrease of payoff for the concerned player, say j. For j, we thus have $U_j(t-1) \geqslant U_j(t)$. Now, if $U_j(t-1) = U_j(t)$, the system reaches a new PNE and is still in Ω^*. Otherwise, $U_j(t-1) < U_j(t)$ and $U_i(t-1) = U_i(t)$, $\forall i \neq j$. According to RSAP, player j will come back to the initial strategy after a finite number of iterations and the system will come back to the initial PNE. ■

Theorem 3 (Convergence of RSAP and convergence rate)**.** *If all SUs adopt the RSAP with exploration probability $\epsilon \to 0$, then the system dynamics converges a.s. to Ω^*, i.e. to a PNE of the game:*

(a) in the general case with endogenous inertia $\rho_j > 0$ for all j;

(b) in the particular case where $H_j = 1$ and $\rho_j = 0$ for all j.

The expected wait until a state in Ω^* is reached, given that the play in the ϵ-perturbed model begins in any state not in Ω^*, is $O(\epsilon^{-1})$ as $\epsilon \to 0$.

Proof: It follows from Lemma 4 and Lemma 3 that $R(\Omega^*) > CR^*(\Omega^*)$. Conclusion comes from Theorem 1. ■

Remark. Our study can be readily extended to other games possessing the weak-FIP. These include dominance solvable games, quasi-acyclic games, power set graphical congestion games and games with the finite improvement property (as defined by Monderer and Shapley [20]).

VI. Performance evaluation

We now conduct simulations to evaluate the performance of the proposed protocol. We simulate a CRN of $N = 50$ SUs and $C = 3$ channels with availability probilities $\mu = [0.3, 0.5, 0.8]$. Each SU is characterized by a generic decreasing noise function $\epsilon(t) \xrightarrow[t\to\infty]{} 0$ and by a user-specific payoff function of the following form: $U_j(.) = w_j f(.)$, where $f(.)$ is a decreasing function common to all the SUs and w_j is a user-specific weight. We set $H_j = 3$ and $\rho_j = 0.3$ for all j.

For performance comparison, we also show the results obtained by simulating the Distributed Learning Algorithm (DLA) recently proposed in [21] and then applied to CRNs in [7]. The idea behind DLA is the following: each player has a prior perception of the payoff performance for each possible strategy and makes a decision relying on this information. The payoff of the chosen alternative is then observed and used to update the perception for that strategy. For a user i, at each iteration, only the perception $Q_j^i(t+1)$ related to the currently played strategy j is updated as follows: $Q_j^i(t+1) = (1-\theta(t))Q_j^i(t) + \theta(t)U_j(n_{s_j})$, where $\theta(t) \in (0,1)$ are smoothing factors (we set $\theta(t) = 1/t$ as in [21] and). Mapping from perceptions to mixed strategies is then given by the Logit rule $\sigma_j^i(t) = \frac{\exp(\gamma Q_j^i(t))}{\Sigma_{i\in\mathcal{C}}\exp(\gamma Q_j^i(t))}$, where γ is the temperature that controls the randomness of channel selections. For the analysis of the fairness of RSAP and DLA

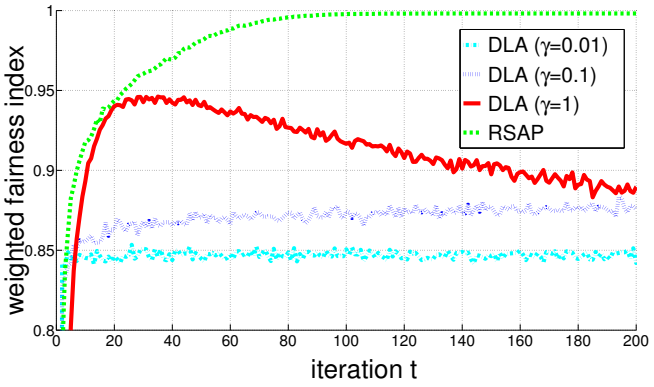

Fig. 4. Weighted fairness index of RSAP and the DLA algorithm proposed in [21]. Each curve represents an average over 1000 independent realizations.

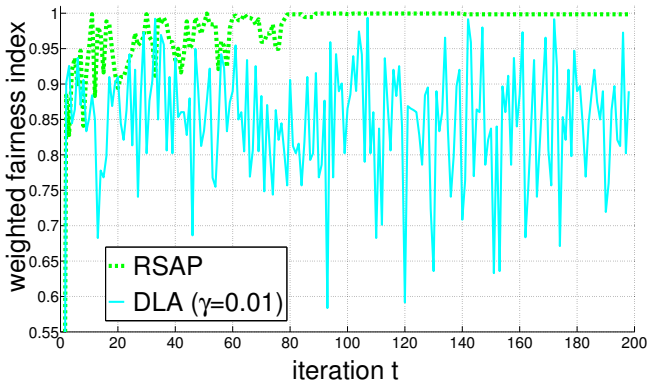

Fig. 5. Weighted fairness index of RSAP and the DLA algorithm proposed in [21]. Each curve represents one single realization of the two algorithms.

we choose a weighted version of the Jain's fairness index $\Upsilon(t) = \frac{(\Sigma_{j\in\mathcal{N}} U_j(t)/w_j)^2}{N\cdot\Sigma_{j\in\mathcal{N}}(U_j(t)/w_j)^2}$, which takes into account the user-specific weights and reaches the maximum of 1 when the resource (the throughput in our case) is equally shared amongst users [22].

Fig. 4 and Fig. 5 show RSAP and DLA convergence trends as a function of the iteration period t. We observe that RSAP convergence is slightly slower with respect to DLA. Nevertheless, while RSAP always converges to a stable NE as $\epsilon(t) \to 0$ (i.e., within 90 iterations), DLA attained equilibrium is not efficient (or, better, not stable as one can infer from the one-realization plot in Fig. 5). This is due to the fact that DLA converges in probability, meaning that only a certain percentage of time at the NE is guaranteed. We notice that although such permanence time increases with the temperature, a too high γ ($\gamma = 1$ in Fig. 4) causes convergence to come to an abrupt stop (at iteration 30 on the example) and fairness starts to decrease. The reason behind this lies in the fact that a high temperature reflects a low randomization, meaning that perceptions are updated not enough often and become obsolete.

VII. Conclusion

In this paper, we have developed and analyzed a distributed spectrum access protocol called retrospective spectrum access protocol. We have demonstrated that the developed protocol can be implemented in a distributed fashion based on only local observations and the network is guaranteed to converge stochastically to an equilibrium state within a bounded delay. For future research, we plan to extend our analysis in the paper to the more generic multi-hop network paradigm and to the challenging case of noise corrupted payoff evaluations.

References

[1] Dean Foster and Peyton Young. Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38(2):219–232, October 1990.

[2] Michihiro Kandori, George J. Mailath, and Rafael Rob. Learning, Mutation, and Long Run Equilibria in Games. *Econometrica*, 61(1):29–56, January 1993. ArticleType: research-article / Full publication date: Jan., 1993 / Copyright 1993 The Econometric Society.

[3] H. P. Young. The Evolution of Conventions. *Econometrica*, 61(1):57–84, Jan. 1993.

[4] Michihiro Kandori and Rafael Rob. Evolution of Equilibria in the Long Run: A General Theory and Applications. *Journal of Economic Theory*, 65(2):383–414, April 1995.

[5] A. Mahajan and D. Teneketzis. *Foundations and Applications of Sensor Management*, chapter Multi-armed Bandit Problems, pages 121–151. Springer-Verlag, 2007.

[6] A. Anandkumar, N. Michael, and A. Tang. Opportunistic Spectrum Access with Multiple Users: Learning under Competition. In *Proc. IEEE International Conference on Computer Communication (INFOCOM)*, San Diego, CA, Apr. 2010.

[7] Xu Chen and Jianwei Huang. Spatial spectrum access game: nash equilibria and distributed learning. In *Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '12, pages 205–214, New York, NY, USA, 2012. ACM.

[8] Stefano Iellamo, Lin Chen, and Marceau Coupechoux. Proportional and double imitation rules for spectrum access in cognitive radio networks. *Computer Networks*, 2013.

[9] G. Ellison. Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Review of Economic Studies*, 67, 2000.

[10] Tone Dieckmann. The evolution of conventions with mobile players. *Journal of Economic Behavior & Organization*, 38(1):93–111, January 1999.

[11] J.W. Friedman and C. Mezzetti. Learning in Games by Random Sampling. *Journal of Economic Theory*, 98(1):55–84, May 2001.

[12] Jason R. Marden, H. Peyton Young, Gürdal Arslan, and Jeff S. Shamma. Payoff-Based Dynamics for Multiplayer Weakly Acyclic Games. *SIAM Journal on Control and Optimization*, 48(1):373–396, January 2009.

[13] H. Peyton Young. Learning by trial and error. *Games and Economic Behavior*, 65(2):626–643, March 2009.

[14] J.R. Marden, H.P. Young, and L.Y. Pao. Achieving pareto optimality through distributed learning. In *2012 IEEE 51st Annual Conference on Decision and Control (CDC)*, pages 7419–7424, 2012.

[15] Bary S.R. Pradelski and H. Peyton Young. Learning efficient Nash equilibria in distributed systems. *Games and Economic Behavior*, 75(2):882–897, July 2012.

[16] Dean P. Foster, University of Pennsylvania Wharton School, H. Peyton Young, Johns Hopkins University Oxford, and University of. Regret testing: learning to play Nash equilibrium without knowing you have an opponent. September 2006.

[17] Minghui Zhu and Sonia Martínez. Distributed Coverage Games for Energy-Aware Mobile Sensor Networks. *SIAM Journal on Control and Optimization*, 51(1):1–27, January 2013.

[18] I. Milchtaich. Congestion Games with Player-Specific Payoff Functions. *Games and Economic Behavior*, 13(1):111–124, Mar. 1996.

[19] Carlos Alós-Ferrer. Learning, bounded memory, and inertia. *Economics Letters*, 101(2):134–136, 2008.

[20] Dov Monderer and Lloyd S. Shapley. Potential Games. *Games and Economic Behavior*, 14(1):124–143, May 1996.

[21] Roberto Cominetti, Emerson Melo, and Sylvain Sorin. A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior*, 70(1):71–83, September 2010.

[22] R. Jain, D. Chiu, and W. Hawe. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. Research Report TR-301, DEC, 1984.