HAL
archives-ouvertes.fr

Fair Scheduling in Common-Pool Games by Aspiration Learning

Georgios Chasparis, Ari Arapostathis, Jeff Shamma

▶ **To cite this version:**

Georgios Chasparis, Ari Arapostathis, Jeff Shamma. Fair Scheduling in Common-Pool Games by Aspiration Learning. WiOpt'12: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, May 2012, Paderborn, Germany. pp.386-390, 2012. <hal-00764167>

HAL Id: hal-00764167
https://hal.inria.fr/hal-00764167

Submitted on 12 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fair Scheduling in Common-Pool Games by Aspiration Learning

Georgios C. Chasparis Ari Arapostathis Jeff S. Shamma

***Abstract*— We propose a distributed learning algorithm for fair scheduling in common-pool games. Common-pool games are strategic-form games where multiple agents compete over utilizing a limited common resource. A characteristic example is the medium access control problem in wireless communications, where multiple users need to decide how to share a single communication channel so that there are no collisions (situations where two or more users use the medium at the same time slot). We introduce a (payoff-based) learning algorithm, namely aspiration learning, according to which agents learn how to play the game based only on their own prior experience, i.e., their previous actions and received rewards. Decisions are also subject to a small probability of mistakes (or mutations). We show that when all agents apply aspiration learning, then as time increases and the probability of mutations goes to zero, the expected percentage of time that agents utilize the common resource is equally divided among agents, i.e., fairness is established. When the step size of the aspiration learning recursion is also approaching zero, then the expected frequency of collisions approaches zero as time increases.**

Keywords: Aspiration learning; Common-pool games; Resource allocation; Medium-access control

I. Introduction

Lately, there has been considerable research interest in distributed optimization techniques as a means of efficient *coordination* in multiagent systems for *efficient resource allocation*. We are particularly interested in problems in which limited resources need to be shared in a distributed manner among several users, namely *common-pool problems*. For example, in wireless communication networks, multiple users often need to allocate fairly the time slots at which they utilize a shared communication channel (see, e.g., packet radio multiple-access protocols such as the ALOHA protocol [2]). The question that naturally emerges is the following: *Can fair scheduling emerge as the outcome of a distributed learning algorithm?*

In this paper, we approach this question by following a noncooperative game-theoretic formulation, where each agent is acting myopically trying to maximize its *own* utility. There have been several efforts which model common-pool problems in communications in a similar manner, e.g., [3], [4]. Although the notion of *Nash equilibrium* emerges naturally in this framework, Nash equilibria (if exist) might *not* be fair solutions to a common-pool game. This is due to the fact that in these games fair solutions are usually represented as probability distributions in the *joint* strategy space of all agents, also discussed in [4]. For example, when two users need to share the same communication channel, a fair allocation could be as follows: *with probability* $1/2$ *the first user uses the channel, and with probability* $1/2$ *the second user uses the channel.* However, such an allocation cannot be represented by *any* Nash equilibrium, since it is defined in the *joint* strategy space.

Other equilibrium concepts defined on the joint strategy space, such as *correlated equilibria* [5], seem more appropriate. This work is also related to learning dynamics for convergence to correlated equilibria [6], however our goal is more specific, that is to *develop a distributed learning scheme which will "converge" (in a sense to be defined) to fair outcomes in the joint strategy space*. These outcomes may or may not correspond to correlated equilibria, however this is a question we do not answer in this paper.

In this paper, we analyze the asymptotic behavior of a (payoff-based) learning algorithm, namely *aspiration learning*, in common-pool games. Aspiration learning is based on a simple rule of "*win-stay, lose-shift*" [7], according to which a successful action is repeated while an unsuccessful action is dropped. Agents' decisions are also subject to small mistakes (or *mutations*). Prior analysis in aspiration learning [1], [8], [9] has focused only on characterizing the *set* of possible outcomes of aspiration learning, and the asymptotic behavior is usually stated in the form of weak convergence arguments. In this paper, we extend prior work by [1] to common-pool games and we characterize explicitly the expected frequency with which the possible outcomes of the process appear as time increases. In fact, we show that in common-pool games *fairness* is established, i.e., the expected percentage of time that users are utilizing the common resource is divided equally among the users, as the mutation probability approaches zero and time goes to infinity. Furthermore, when the step size of the aspiration learning recursion also approaches zero, the expected number of collisions approaches zero as time increases.

The remainder of the paper is organized as follows. Section II introduces common-pool games and discusses a few examples drawn from medium-access control in wireless networks. Sections III–IV present background material in aspiration learning and finite Markov chains. Section V

This paper is an extension of [1].

G.C. Chasparis is with the Department of Automatic Control, Lund University, 221 00-SE Lund, Sweden (georgios.chasparis@control.lth.se). http://www.control.lth.se/chasparis. This author's work was supported by the Swedish Research Council through the Linnaeus Center LCCC.

A. Arapostathis is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, 1 University Station, Austin, TX 78712 (ari@mail.utexas.edu). http://www.ece.utexas.edu/~ari. This author's work was supported in part by the Office of Naval Research through the Electric Ship Research and Development Consortium.

J.S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (shamma@gatech.edu). http://www.prism.gatech.edu/~jshamma3. This author's work was supported by ONR project #N00014-09-1-0751 and AFOSR project #FA9550-09-1-0538.

978-1-61284-824-2/2012 - Copyright is with IFIP

discusses the asymptotic behavior of aspiration learning in common-pool games and the establishment of fairness. The results are illustrated through simulations in Section VI.

Terminology: We consider the standard setup of finite strategic-form games. There is a finite set of *agents/players*, $\mathcal{I} = \{1, ..., n\}$, $n \geq 2$, and each agent has a finite number of *actions*, denoted by $\mathcal{A}_i$. The set of *action profiles* (or *joint actions*) is the cartesian product $\mathcal{A} \triangleq \mathcal{A}_1 \times ... \times \mathcal{A}_n$; $\alpha_i \in \mathcal{A}_i$ denotes an action of agent i; and $\alpha = (\alpha_1, ..., \alpha_n) \in \mathcal{A}$ denotes an action profile of all agents. We will also denote $|\mathcal{A}|$ the cardinality of the set of $\mathcal{A}$. The *payoff/utility function* of player i is a mapping $u_i : \mathcal{A} \to \mathbb{R}$.

II. Common-Pool Game (CPG)

A. Definition

Common-pool games refer to strategic interactions where two or more agents need to decide unilaterally whether or not to utilize a limited common resource. In such interactions, each agent would rather use the common resource by itself than share it with another agent, which is usually penalizing for both of them.

Definition 2.1 (CPG): A common-pool game (CPG) is a strategic-form game such that, for each agent $i \in \mathcal{I}$, the action space is $\mathcal{A}_i = \{p_0, p_1, ..., p_{m-1}\}$, for some $m \geq 2$ and $0 \leq p_0 < p_1 < ... < p_{m-1}$, and the utility function is:

$$u_i(\alpha) \triangleq \begin{cases} 1 - c_j & \alpha_i = p_j, \alpha_i > \max_{\ell \neq i} \alpha_\ell, \\ -c_j + \tau_j & \alpha_i = p_j, \exists s \in \mathcal{I} \backslash i \text{ s.t. } \alpha_s > \max_{\ell \neq s} \alpha_\ell, \\ -c_j & \alpha_i = p_j, \nexists s \in \mathcal{I} \text{ s.t. } \alpha_s > \max_{\ell \neq s} \alpha_\ell, \end{cases}$$

where

$$0 \leq c_0 < ... < c_{m-1} < 1,$$

$$-c_0 < -c_{m-2} + \tau_{m-2} < ... < -c_0 + \tau_0 < 1 - c_{m-1}$$

and $\tau_j > 0$ for all $j = 0, ..., m-2$.

This definition of a CPG can be considered as a finite-action analog of continuous-action CPG's defined in [10]. Table I presents an example of a two-player and three-action CPG.

	p_0	p_1	p_2
p_0	$-c_0, -c_0$	$-c_0 + \tau_0, 1 - c_1$	$-c_0 + \tau_0, 1 - c_2$
p_1	$1 - c_1, -c_0 + \tau_0$	$-c_1, -c_1$	$-c_1 + \tau_1, 1 - c_2$
p_2	$1 - c_2, -c_0 + \tau_0$	$1 - c_2, -c_1 + \tau_1$	$-c_2, -c_2$

TABLE I

A CPG OF 2 PLAYERS AND 3 ACTIONS.

We will characterize by "*success*" any action profile in which one player's action is strictly greater than any other player's action. Accordingly, any other situation will correspond to a "*failure*." More specifically, in any CPG, we define the set of "*successful*" action profiles as

$$\bar{\mathcal{A}} \triangleq \{\alpha \in \mathcal{A} : \exists i \in \mathcal{I} \text{ s.t. } \alpha_i > \max_{\ell \neq i} \alpha_\ell\}.$$

For example, this set of joint actions will correspond to the off-diagonal action profiles in Table I. It is also evident that the set $\bar{\mathcal{A}}$ *payoff-dominates* the set $\mathcal{A} \backslash \bar{\mathcal{A}}$.

According to Definition 2.1, a player is rewarded when it "succeeds," i.e., when its action is higher than everyone else's action. On the other hand, all players are penalized when there is a "failure." However, the penalty for the players who "fail" is smaller when there is a "success."

B. Example: Medium Access Control in Wireless Networks

In packet radio multiple-access protocols (cf., [11, Chapter 5]), there are multiple users which compete for access to a single communication channel. Each user needs to decide whether or not to occupy the channel in a given time slot based only on *local* information. If more than one user is occupying the channel at a given time slot, then a collision occurs and the user needs to resubmit the data. The base station can send signals to the users indicating a successful or unsuccessful transmission.

An example of such multiple-access protocols is the ALOHA protocol [2], where users transmit a packet according to a probabilistic pattern. Another example is the carrier-sense multiple-access protocol (CSMA) and its variations (cf., [11, Chapter 5]), according to which the users first detect whether the channel is occupied and then transmit. For example, in the *p-persistent CSMA*, if the medium is idle, users transmit with probability p. In CSMA/CA instead, if the medium is occupied, each user waits a random time (also known as *backoff factor*) before resubmission of the packet.

Due to the distributed nature of the problem, the multiple access problem can be formulated as a strategic-form game. In fact, there has been a large amount of research efforts discussing such possibilities including [3], [4], [12]. Some of the proposed strategic-form games can be formulated as a CPG, as the following discussion will reveal.

In particular, in [3], a random access game is considered, where users decide on whether to submit a packet or wait. The users receive the hishest payoff when they succeed on transmitting a packet, while they receive a penalty each time there is a collision, i.e., when there are two or more users transmitting at the same time slot. It is shown that the game with two users exhibits three Nash equilibria two of which are pure. A similar framework is considered in [4], where the action space of each user consists of multiple power levels of transmission. If a user transmits with a power level that is strictly larger than the power level of any other user, then it is able to transmit successfully, otherwise a collision occurs and the transmission is not possible. This game can be formulated in a straightforward manner as a CPG, where the action profile of each user is defined as the power level of transmission and the utilities are defined according to Definition 2.1. It is noted, as expected, that for such a random access game, the Nash equilibria do not achieve a socially optimum solution. As also shown in [4], socially optimum solutions can be achieved through *correlated strategies* (i.e., probability distributions in the joint action space). However, implementing correlated strategies requires the presence of a referree.

It is worth mentioning that there has been several other approaches for addressing the same problem. For example, in [12], users have discretion over the size of the waiting time in a CSMA/CA framework which is assumed to evolve in a finite set. Again, it is noted that Nash equilibria will not be fair since users have the incentive to set a waiting time which is as small as possible. As a result, at a Nash equilibrium either only one user is occupying the channel, or more than one user is occupying the channel leading to a collision. Nash equilibria, as in [4], do not correspond to socially optimal solutions, thus reference [12] addresses this problems through a Nash bargaining framework from the theory of cooperative games.

The above formulations of packet radio multiple-access protocols reveal the necessity for distributed schemes that will guarantee *fair* use of the common medium among several users. Similarly to the model of [4], we wish to address this question through a noncooperative game formulation, where actions correspond to the power level of the transmission and utilities are given by Definition 2.1. We will also demonstrate how fair solutions can be established when users learn over time through a *distributed* learning rule, namely aspiration learning.

III. Background on Aspiration Learning

In this section, we present a (payoff-based) learning algorithm developed and analyzed in [1].

For some constants $\zeta > 0$, $\epsilon > 0$, $\lambda \geq 0$, $c > 0$, $0 < h < 1$, and $\underline{\rho}, \overline{\rho} \in \mathbb{R}$, such that

$$-\infty < \underline{\rho} < \min_{\alpha \in \mathcal{A}, i \in \mathcal{I}} u_i(\alpha) \leq \max_{\alpha \in \mathcal{A}, i \in \mathcal{I}} u_i(\alpha) < \overline{\rho} < \infty,$$

the aspiration learning iteration initialized at $(\alpha(0), \rho(0))$ is described by Table II. According to this algorithm, each agent i keeps track of an aspiration level, ρ_i, which measures player i's desirable return and is defined as a discounted running average of its payoffs throughout the history of play. Given the current aspiration level, $\rho_i(t)$, agent i selects a new action $\alpha_i(t+1)$. If the previous action, α_i, provided utility higher than $\rho_i(t)$, then agent i is "satisfied" and selects the same action, $\alpha_i(t+1) = \alpha_i$. Otherwise, the new action is selected randomly over all available actions, where the probability of selecting again α_i depends on the level of discontent measured by the difference $u_i(\alpha) - \rho_i(t) < 0$.

Define the state-space as $\mathcal{X} \triangleq \mathcal{A} \times [\underline{\rho}, \overline{\rho}]^n$, i.e., pairs of i) joint actions α and ii) vectors of aspiration levels, ρ_i, $i \in \mathcal{I}$.

We assume the standard notation of continuous-space Markov chains [13]. Let also:

- $\mathcal{B}(\mathcal{X})$: the Borel σ-algebra on $\mathcal{X}$.
- $\mathcal{P}(\mathcal{X})$: the set of probability measures on $\mathcal{B}(\mathcal{X})$ endowed with the topology of weak convergence.
- $T : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ is a *transition probability function* if i) $T(x, \cdot)$ is a probability measure for all $x \in \mathcal{X}$ and ii) $T(\cdot, B)$ is measurable on $\mathcal{X}$ for all $B \in \mathcal{B}(\mathcal{X})$.
- For $\mu \in \mathcal{P}(\mathcal{X})$ and transition probability function T, $\mu T \in \mathcal{P}(\mathcal{X})$ is the probability measure defined by $\mu T(B) \triangleq \int_{\mathcal{X}} \mu(dx) T(x, B)$.

At any instance $t = 0, 1, ...,$

1) Agent i plays $\alpha_i(t) = \alpha_i$ and measures utility $u_i(\alpha)$.
2) Agent i updates its aspiration level according to

$$\rho_i(t+1) = \mathtt{sat}[\rho_i(t) + \epsilon[u_i(\alpha) - \rho_i(t)] + r_i(t)]$$

where

$$r_i(t) \triangleq \begin{cases} 0, & \text{w.p. } 1-\lambda \\ \mathtt{rand}[-\zeta, \zeta], & \text{w.p. } \lambda \end{cases},$$

and

$$\mathtt{sat}[\rho] \triangleq \begin{cases} \overline{\rho}, & \rho > \overline{\rho} \\ \rho, & \rho \in [\underline{\rho}, \overline{\rho}] \\ \underline{\rho}, & \rho < \underline{\rho} \end{cases}.$$

3) Agent i updates its action:

$$\alpha_i(t+1) = \begin{cases} \alpha_i & \text{w.p. } \phi(u_i(\alpha) - \rho_i) \\ \mathtt{rand}(\mathcal{A}_i \backslash \alpha_i) & \text{w.p. } 1 - \phi(u_i(\alpha) - \rho_i) \end{cases}$$

where

$$\phi(z) \triangleq \begin{cases} 1 & z \geq 0 \\ \max(h, 1+cz) & z < 0 \end{cases}.$$

4) Agent i updates the time and repeats.

TABLE II
Aspiration Learning

- $\mu \in \mathcal{P}(\mathcal{X})$ is an *invariant measure* for the transition probability function T if $\mu = \mu T$.
- For transition probability functions T_1 and T_2, the transition probability function $T_1 T_2$ is defined by $T_1 T_2(x, B) \triangleq \int_{\mathcal{X}} T_1(x, dy) T_2(y, B)$.
- δ_x: the Dirac measure defined by $x \in \mathcal{X}$.

Aspiration learning defines an $\mathcal{X}$-valued Markov chain. Let $P_\lambda(x, \cdot)$ denote the corresponding transition probability function. We will refer to this process with $\lambda > 0$ as the *perturbed process*. Let also $P(x, \cdot)$ denote the transition probability function of the *unperturbed* process.

The analysis of the asymptotic behavior of aspiration learning can be related to the *pure strategy states*:

Definition 3.1 (Pure strategy state): A pure strategy state is a state $x = (\alpha, \rho) \in \mathcal{X}$ such that for all $i \in \mathcal{I}$, $u_i(\alpha) = \rho_i$.

We denote the set of pure strategy states by $\mathcal{S}$. The set $\mathcal{S}$ is isomorphic to $\mathcal{A}$ and can be identified as such.

Let also $Q(x, \cdot)$ denote the transition probability function induced by the aspiration learning algorithm where exactly one player trembles. For some $s \in \mathcal{S}$ define the sets $N_\varepsilon \triangleq [\alpha_s, [\rho_s - \varepsilon, \rho_s + \varepsilon])$, $\varepsilon > 0$, where (α_s, ρ_s) denote the action and aspiration level of s. For any two pure strategy states, $s, s' \in \mathcal{S}$, define also

$$\hat{P}_{ss'} \triangleq \lim_{t \to \infty} Q P^t(s, N_\varepsilon(s'))$$

for some $\varepsilon > 0$ sufficiently small. As we showed in Proposition 3.2 in [1], $\hat{P}_{ss'}$ is independent of the selection of ε. It can be interpreted as the probability that under the dynamics $QPP...$ with initial condition s the process has been "captured" by s', i.e., action $\alpha_{s'}$ is being played

repeatedly after some time t.

Define also the stochastic matrix $\hat{P} \triangleq [\hat{P}_{ss'}]$. By Proposition 3.5 in [1], $\hat{P}$ is irreducible and aperiodic. The following proposition relates the asymptotic behavior of aspiration learning with the unique invariant distribution of $\hat{P}$ as $\lambda \to 0$.

Proposition 3.1 (Theorem 3.1 in [1]): There exists a unique probability vector $\pi = (\pi_1, ..., \pi_{|\mathcal{S}|})$ such that for any collection of invariant probability measures $\{\mu_\lambda \in \mathcal{P}(\mathcal{X}) : \mu_\lambda P_\lambda = \mu_\lambda, \lambda > 0\}$, we have

$$\lim_{\lambda \downarrow 0} \mu_\lambda(\cdot) = \hat{\mu}(\cdot) \triangleq \sum_{s \in \mathcal{S}} \pi_s \delta_s(\cdot),$$

where convergence is in the weak sense. Furthermore, π is the unique invariant distribution of $\hat{P}$.*

Using Proposition 3.1 and Birkhoff's individual ergodic theorem, e.g., [13, Theorem 2.3.4], the expected percentage of time that the process spends in any set $B \in \mathcal{B}(\mathcal{X})$ such that $\partial B \cap \mathcal{S} \neq \varnothing$ is $\hat{\mu}(B)$ as λ approaches zero and time increases (see [1, Theorem 3.1]). Thus, $\hat{\mu}$ and, therefore, π characterizes the asymptotic behavior of aspiration learning.

IV. Background on Finite Markov Chains

In order to compute the invariant distribution of a finite-state, irreducible and aperiodic Markov chain, we are going to consider a characterization introduced by [14]. In particular, for finite Markov chains an invariant measure can be expressed as the ratio of sums of products consisting of transition probabilities. These products can be described conveniently by means of graphs on the set of states of the chain. In particular, let $\mathcal{S}$ be a finite set of states, whose elements will be denoted by s_k, s_ℓ, etc., and let a subset $\mathcal{W}$ of $\mathcal{S}$.

Definition 4.1: ($\mathcal{W}$-graph) A graph consisting of arrows $s_k \to s_\ell$ ($s_k \in \mathcal{S}\backslash\mathcal{W}, s_\ell \in \mathcal{S}, s_\ell \neq s_k$) is called a $\mathcal{W}$-graph if it satisfies the following conditions:

1) *every point $k \in \mathcal{S}\backslash\mathcal{W}$ is the initial point of exactly one arrow;*
2) *there are no closed cycles in the graph; or, equivalently, for any point $s_k \in \mathcal{S}\backslash\mathcal{W}$ there exists a sequence of arrows leading from it to some point $s_\ell \in \mathcal{W}$.*

We denote by $\mathcal{G}\{\mathcal{W}\}$ the set of $\mathcal{W}$-graphs; we shall use the letter g to denote graphs. If $\hat{P}_{s_k s_\ell}$ are nonnegative numbers, where $s_k, s_\ell \in \mathcal{S}$, define also the transition probability along path g as

$$\varpi(g) \triangleq \prod_{(s_k \to s_\ell) \in g} \hat{P}_{s_k s_\ell}.$$

The following Lemma holds:

Lemma 4.1 (Lemma 6.3.1 in [14]): Let us consider a Markov chain with a finite set of states $\mathcal{S}$ and transition probabilities $\{\hat{P}_{s_k s_\ell}\}$ and assume that every state can be reached from any other state in a finite number of steps. Then, the stationary distribution of the chain is $\pi = [\pi_s]$, where

$$\pi_s = \frac{R_s}{\sum_{s_i \in \mathcal{S}} R_{s_i}}, s \in \mathcal{S}$$

where $R_s \triangleq \sum_{g \in \mathcal{G}\{s\}} \varpi(g)$.

V. Fair Scheduling in CPG's

In this section, using Proposition 3.1 and Lemma 4.1 we establish fairness in CPG's.

First, define the subset of pure-strategy states that correspond to "successful" states for agent i:

$$\bar{\mathcal{S}}_i \triangleq \{s \in \mathcal{S} : \alpha_i > \alpha_j, \forall j \in \mathcal{I}\backslash i\},$$

i.e., $\bar{\mathcal{S}}_i$ corresponds to the set of pure-strategy states in which the action of agent i is strictly larger than the action of any other agent $j \neq i$. Let also $\bar{\mathcal{S}} \triangleq \bigcup_{i \in \mathcal{I}} \bar{\mathcal{S}}_i$, which is isomorphic to the set of successful action profiles, $\bar{\mathcal{A}}$.

For any two states $s, s' \in \mathcal{S}$, we define the following equivalence relation, denoted by $\sim$.

Definition 5.1 (State equivalence): In any CPG and for any two pure-strategy states $s, s' \in \mathcal{S}$ such that $s \neq s'$, let α and α' denote the corresponding action profiles. We write $s \sim s'$ if there exist $i, j \in \mathcal{I}$, $i \neq j$, such that

1) $\alpha'_i = \alpha_j$,
2) $\alpha_i = \alpha'_j$, *and*
3) $\alpha'_k = \alpha_k$ *for all* $k \neq i, j$.

Note that the equivalence relation $\sim$ defines an isomorphism among the states of any two sets $\bar{\mathcal{S}}_i$ and $\bar{\mathcal{S}}_j$ for any $i \neq j$. An immediate implication of the above equivalence property is the following:

Claim 5.1: In any CPG, let $s, s' \in \mathcal{S}$ be such that $s \sim s'$. Let also α and α' be the corresponding action profiles of s and s', respectively. Then, there exist $i, j \in \mathcal{I}$, $i \neq j$, such that:

1) $u_j(\alpha') = u_i(\alpha)$,
2) $u_i(\alpha') = u_j(\alpha)$, *and*
3) $u_k(\alpha') = u_k(\alpha)$ *for all* $k \neq i, j$.

Proof: This is a direct consequence of the symmetry in the payoff function of the CPG and the definition of the state equivalence property (Definition 5.1). ■

Lemma 5.1 (Fairness): For any CPG, $\pi_{\bar{\mathcal{S}}_1} = ... = \pi_{\bar{\mathcal{S}}_n}$.

Proof: (sketch) Let $i \in \mathcal{I}$ and $s \in \bar{\mathcal{S}}_i$. Consider also the set of $\mathcal{G}\{s\}$-graphs according to Definition 4.1. We can define a sequence of pure strategy states in $\mathcal{S}$, $\{s_1, ..., s_L\}$ such that $s_L \equiv s$ and

$$g = \bigcup_{\ell=1}^{L-1} (s_\ell \to s_{\ell+1}) \in \mathcal{G}\{s\},$$

for some $L \in \mathbb{N}$. Consider any other agent $j \neq i$ and a state $s' \in \bar{\mathcal{S}}_j$ such that $s' \sim s$. For any path $g \in \mathcal{G}\{s\}$ of length L, there exists a *unique* path $g' \in \mathcal{G}\{s'\}$ consisting of a sequence of states of length L, $\{s'_1, ..., s'_L\}$, such that $s'_\ell \sim s_\ell$, for all $\ell \in \{1, ..., L\}$. Lastly, due to the symmetry of any CPG and Claim 5.1, we have that

$$\hat{P}_{s_\ell s_{\ell+1}} = \hat{P}_{s'_\ell s'_{\ell+1}}$$

for any $\ell \in \{1, ..., L-1\}$. Hence, we conclude that $\varpi(g') = \varpi(g)$. In other words, there exists an isomorphism between the graphs in the sets $\mathcal{G}\{s\}$ and $\mathcal{G}\{s'\}$, such that any two isomorphic graphs have the same transition probability. Accordingly, we have that $\pi_s = \pi'_s$ for any two states

s, s' such that $s \sim s'$. Since any two sets $\bar{\mathcal{S}}_i$ and $\bar{\mathcal{S}}_j$ are isomorphic with respect to the equivalence relation $\sim$, we further conclude that $\pi_{\bar{\mathcal{S}}_1} = ... = \pi_{\bar{\mathcal{S}}_n}$. ■

Lemma 5.2: For any CPG and for sufficiently small $\zeta > 0$, $\pi_{s_i} \to 0$ as $\epsilon \to 0$, for all $s_i \notin \bar{\mathcal{S}}$.

Proof: (sketch) The proof follows from a generalization of Theorem 4.1 in [1] derived for *strict coordination games* to the case of CPG's. ■

Theorem 5.1 (Fairness for small ϵ): For any CPG and for sufficiently small $\zeta > 0$,

$$\pi_{\bar{\mathcal{S}}_i} \to {}^1\!/_n \text{ as } \epsilon \to 0,$$

for all $i \in \mathcal{I}$.

Proof: First, recognize that the sets $\{\bar{\mathcal{S}}_i\}$ are mutually disjoint, and $\bigcup_{i=1}^n \bar{\mathcal{S}}_i = \bar{\mathcal{S}}$. Then, by Lemma 5.2, we have that $\pi_{\bar{\mathcal{S}}} = \sum_{i=1}^n \pi_{\bar{\mathcal{S}}_i} \to 1$ as $\epsilon \to 0$. Thus, according to Lemma 5.1 the conclusion follows. ■

In other words, we have shown that the invariant distribution π puts equal weight on either agent "*succeeding*." Furthermore, it puts zero weight on states outside $\bar{\mathcal{S}}$ (i.e., "*failures*") as $\epsilon \to 0$.

VI. Simulations

We may combine Proposition 3.1 with Lemma 5.1 and Theorem 5.1 to provide a characterization of the asymptotic behavior of aspiration learning in CPG's as λ and ϵ approach zero. In fact, according to Proposition 3.1 and Lemma 5.1, the expected percentage of time that the aspiration learning spends in any one of the pure strategy sets $\bar{\mathcal{S}}_i$ should be equal to each other as $\lambda \to 0$ and $t \to \infty$ (i.e., *fairness* is established). Furthermore, according to Theorem 5.1, the expected percentage of "*failures*" (i.e., states outside $\bar{\mathcal{S}}$) approaches zero as $\epsilon \to 0$ and $t \to \infty$.

We consider the following setup for aspiration learning:

$$\lambda = 0.001, \epsilon = 0.001, h = 0.01, c = 0.05, \zeta = 0.05.$$

Also, we consider a CPG of 2 players and 4 actions, where $c_0 = 0, c_1 = 0.1, c_2 = 0.2$, $c_3 = 0.3$ and $\tau_0 = \tau_1 = \tau_2 = \tau_3 = 0.8$. Under this setup, Figure 1 demonstrates the response of aspiration learning. We observe, as Theorem 5.1 predicts, that the frequency with which either agent utilizes the common resource approaches ${}^1\!/_2$ as time increases. Furthermore, the frequency of collisions (i.e., the actions in which neither agent utilizes the resource successfully) approaches zero as time increases.

VII. Conclusions

We proposed a distributed learning scheme for establishing fair scheduling in CPG's. Agents make decisions based only on their own prior experience of the game, i.e., actions played and received rewards. We showed analytically and demonstrated through simulations that aspiration learning provides a fair solution to CPG's. In fact, as time increases, the expected frequency with which agents utilize the common resource is equally divided among agents, while the expected frequency of collisions approaches zero.

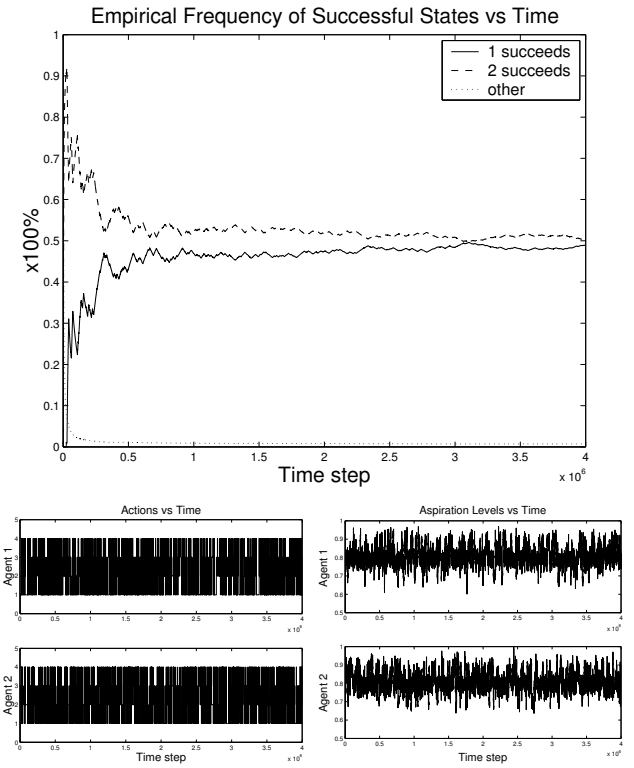

Fig. 1. A typical response of aspiration learning in a CPG with 2 players and 4 actions.

References

[1] G. Chasparis, J. Shamma, and A. Arapostathis, "Aspiration learning in coordination games," in *IEEE Conference on Decision and Control*, Atlanta, GA, 2010.

[2] N. Abramson, "The Aloha system - another alternative for computer communications," in *Proc. 1970 Fall Joint Computer Conference*, A. Press, Ed., 1970, pp. 281–285.

[3] H. Inaltekin and S. Wicker, "A one-shot random access game for wireless networks," in *International Conference on Wireless Networks, Communications and Mobile Computing*, 2005.

[4] H. Tembine, E. Altman, R. ElAzouri, and Y. Hayel, "Correlated evolutionary stable strategies in random medium access control," in *International Conference on Game Theory for Networks*, 2009, pp. 212–221.

[5] R. J. Aumann, "Correlated equilibrium as an expression of bayesian rationality," *Econometrica*, vol. 55, pp. 1–18, 1987.

[6] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, pp. 1127–1150, 2000.

[7] M. Posch, A. Pichler, and K. Sigmund, "The efficiency of adapting aspiration levels," *Biological Sciences*, vol. 266, no. 1427, pp. 1427–1435, July 1999.

[8] R. Karandikar, D. Mookherjee, and D. Ray, "Evolving aspirations and cooperation," *Journal of Economic Theory*, vol. 80, pp. 292–331, 1998.

[9] I. K. Cho and A. Matsui, "Learning aspiration in repeated games," *Journal of Economic Theory*, vol. 124, pp. 171–201, 2005.

[10] H. Meinhardt, "Common pool games are convex games," *Journal of Public Economic Theory*, vol. 1, no. 2, pp. 247–270, 1999.

[11] Z. Han and K. R. Liu, *Resource Allocation for Wireless Networks*. Cambridge University Press, 2008.

[12] M. Félegyházi, M. Cagali, and J. Hubaux, "Efficient MAC in cognitive radio systems: A game-theoretic approach," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1984–1995, 2009.

[13] O. Hernandez-Lerma and J. B. Lasserre, *Markov Chains and Invariant Probabilities*. Birkhauser Verlag, 2003.

[14] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*. New York, NY: Springer-Verlag, 1984.