Retrospective Spectrum Access Protocol: A Completely Uncoupled Learning Algorithm for Cognitive Networks

Marceau Coupechoux*, Stefano Iellamo*, Lin Chen+
* TELECOM ParisTech (INFRES/RMS) and CNRS LTCI
+ University Paris XI Orsay (LRI)

CEFIPRA Workshop on New Avenues for Network Models
Indian Institute of Science, Bangalore

14 Jan 2014

Introduction

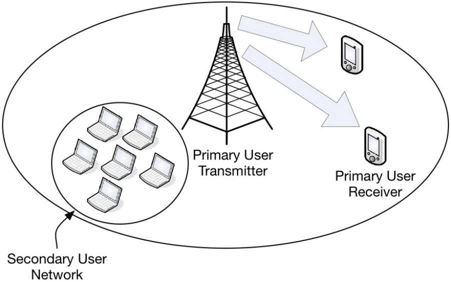
- Opportunistic spectrum access in cognitive radio networks
- SU access freq. channels partially occupied by the licensed PU
- Distributed spectrum access policies based only on past experienced payoffs (*i.e. completely uncoupled dynamics* as opposed to *coupled dynamics* where players can observe the actions of others)
- Convergence analysis based on perturbed Markov chains

Related Work

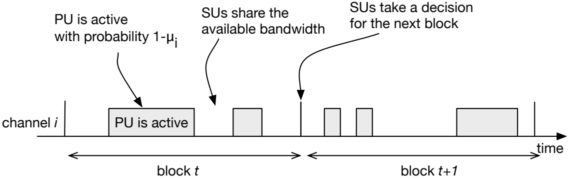
- Distributed spectrum access in CRN:
  - $\#$ SUs $<$ $\#$ Channels: solutions based on *multi-user* Multi-Armed Bandit [Mahajan07, Anandkumar10]
  - Large population of SUs: Distributed Learning Algorithm [Chen12] based on Reinforcement Learning and stochastic approx., Imitation based algorithms [Iellamo13]
- Bounded rationality and learning in presence of noise:
  - Bounded rationality: [Foster90, Kandori93, Kandori95, Dieckmann99, Ellison00]
  - Learning in presence of noise: [Mertikopoulos09]
  - Mistake models: [Friedman01]
  - Trial and Error: [Pradelski12]
  - Similar approaches to our algorithm in other contexts: [Marden09, Zhu13]

System Model I

- A PU is using on the DL a set $\mathcal{C}$ of C freq. channels
- Primary receivers are operated in a synchronous time-slotted fashion
- The secondary network is made of a set $\mathcal{N}$ of N SUs
- We assume perfect sensing



System Model II



- At each time slot, channel i is free with probability μ_i
- Throughput achieved by j along a block is denoted T_j
- Expected throughput when block duration is large: $\mathbb{E}[T_j] = B\mu_{s_j} p_j(n_{s_j})$
- $p_j(\cdot)$ is a function that depends on the MAC protocol, on j and on the number of SUs on the channel chosen by j, n_{s_j}
- We assume $B = 1$, p_j strictly decreasing and $p_j(x) \leq 1/x$ for $x > 0$

Spectrum Access Game Formulation

Definition

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. When a player j chooses strategy $s_j \in \mathcal{C}$, its player-specific utility function $U_j(s_j, \mathbf{s}_{-j})$ is defined as

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

Lemma (Milchtaich96)

For the spectrum access game G, there exists at least one pure Nash equilibrium (PNE).

Motivation

- Find a distributed strategy for SUs to converge to a PNE
- Uniform random imitation of **another** SU leads to the *replicator dynamics* (see Proportional Imitation Rule in [Schlag96, Schlag99])
- Uniform random imitation of **two** SUs leads to the *aggregate monotone dynamics* (see Double Imitation in [Schlag96, Schlag99])
- Imitation on the same channel can be approximated by a double replicator dynamics [Iellamo13]
- We now want to avoid any information exchange between SUs

RSAP I

- Each SU j has a finite memory $\mathcal{H}_j$ containing the history (strategies and payoffs) relative to the H_j past iterations.
- State of the system at t:

$$z(t) \triangleq \{s_j(t-h), U_j(t-h)\}_{j \in \mathcal{N}, h \in \mathcal{H}_j}$$

- Number of iterations passed from the highest remembered payoff:

$$\lambda_j = \min \operatorname*{argmax}_{h \in \mathcal{H}_j} U_j(t-h)$$

- Define *inertia* ρ_j = prob. that j is unable to update its strategy at each t [Alos-Ferrer08] (an endogenous parameter for us)
- Define the *exploration probability* $\epsilon(t) \to 0$

RSAP II

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 $t \geq 1$ **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

RSAP III

Definition (Migration Stable State)

A migration stable state (MSS) ω is a state where no more migration is possible, i.e., $U_j(t) \geq U_j(t-h)\ \forall h \in \mathcal{H}_j\ \forall j \in \mathcal{N}$.

Perturbed Markov Chain I

- We have a model of evolution with noise:
  - $Z = \left\{ z \triangleq \{s_j(t-h), U_j(t-h)\}_{j \in \mathcal{N}, h \in \mathcal{H}_j} \right\}$ is the finite state space of the system stochastic process
  - **Unperturbed chain**: $P = (p_{uv})_{(u,v) \in Z^2}$ is the transition matrix of RSAP without exploration (i.e. $\epsilon(t) = 0 \ \forall t$)
  - **Perturbed chains**: $P(\epsilon) = (p_{uv}(\epsilon))_{(u,v) \in Z^2}$ is a family of transitions matrices on Z indexed by $\epsilon \in [0, \bar{\epsilon}]$ associated to RSAP with exploration ϵ
- Properties of $P(\epsilon)$:
  - $P(\epsilon)$ is ergodic for $\epsilon > 0$
  - $P(\epsilon)$ is continuous in ϵ and $P(0) = P$
  - There is a cost function $c : Z^2 \to \mathcal{R}^+ \cup \{\infty\}$ s.t. for any pair of states (u, v), $\lim_{\epsilon \to 0} \frac{p_{uv}(\epsilon)}{\epsilon^{c_{uv}}}$ exists and is strictly positive for $c_{uv} < \infty$ and $p_{uv}(\epsilon) = 0$ if $c_{uv} = \infty$

Perturbed Markov Chain II

- Remarks:
  - ϵ can be interpreted as a small probability that SUs do not follow the rule of the dynamics. When a SU explores, we say that there is a **mutation**
  - The cost c_{uv} is the rate at which $p_{uv}(\epsilon)$ tends to zero as ϵ vanishes
  - c_{uv} can also be seen as the **number of mutations** needed to go from state u to state v
  - $c_{uv} = 0$ when $p_{uv} \neq 0$ in the unperturbed Markov chain
  - $c_{uv} = \infty$ when the transition $u \rightarrow v$ is impossible in the perturbed Markov chain
  - The unperturbed Markov chain is not necessarily ergodic.
    It has one or more **limit sets**, i.e., **recurrent classes**

Perturbed Markov Chain III

Lemma (Young93)

There exists a limit distribution $\mu^* = \lim_{\epsilon \to 0} \mu(\epsilon)$

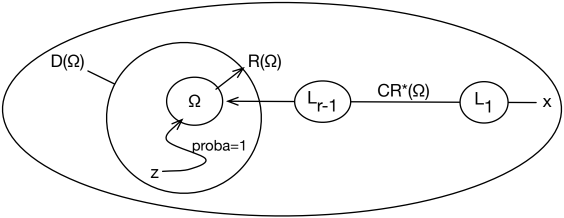
Definition

A state $i \in Z$ is said to be **long-run stochastically stable** iff $\mu_i^* > 0$.

Lemma (Ellison00)

The set of stochastically stable states is included in the recurrent classes (limit sets) of the unperturbed Markov chain (Z, P).

Ellison Radius Coradius Theorem I



- Ω: a union of limit sets of (Z, P)
- $D(\Omega)$: **basin of attraction**, the set of states from which the unperturbed chain converges to Ω w.p.1
- $R(\Omega)$: **radius**, the min cost of any path from Ω out of $D(\Omega)$
- $CR(\Omega)$: **coradius**, maximum cost to Ω
- $CR^*(\Omega)$: **modified coradius**, obtained by substracting from the cost, the radius of intermediate limit sets

Ellison Radius Coradius Theorem II

Theorem (Ellison00, Theorem 2 and Sandholm10, Chap. 12)

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:

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

Proof idea

Uses the Markov chain tree theorem and the fact that it is more difficult to escape from Ω than to return to Ω.

RSAP Convergence Analysis I

Proposition

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

Proof idea

Every MSS is obviously a LS. (a) There is a positive probability that no SU change its strategy for $\max_j H_j$ iterations. After such an event, the system is in a MSS. (b) If the system is in a LS, every SU must switch between at most two strategies. As the system is deterministic, the system alternates between two states. So the LS has a unique state because every SU can choose between two payoffs.

RSAP Convergence Analysis II

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

Lemma

It holds that $R(\omega) = 1 \ \forall \omega \notin \Omega^*$, *where* ω *is a LS.*

Proof idea

For a congestion game $\mathcal{G}$ with player specific decreasing payoff functions, the weak-FIP property holds [Milchtaich96]. Using weak-FIP, we show that a single mutation is enough to leave the basin of attraction of any MSS not in Ω^* and to reach a new MSS.

RSAP Convergence Analysis III

Lemma

$CR^*(\Omega^*) = 1$

Proof idea

From any state, there is a path of null cost to reach a MSS and then (from weak-FIP property) a path, which is a sequence of MSSs. Each MSS has a radius of 1.

Lemma

$R(\Omega^*) > 1$

Proof idea

Comes from the definition of the PNEs and of RSAP.

RSAP Convergence Analysis IV

Theorem (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. 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 $\epsilon \to 0$.

Remark. Our study can be readily extended to other games possessing the weak-FIP and hence the FBRP, weak-FBRP and the FIP [Monderer96] since FIP $\Rightarrow$ FBRP $\Rightarrow$ weak-FBRP $\Rightarrow$ weak-FIP.

Simulation Settings

- We compare our algorithm to Trial and Error (T&E, Pradelski's optimized learning parameters in [Pradelski&Young 2012]) and to the Distributed Learning Algorithm (DLA) [Chen&Huang 2012].
- We consider two networks:
  - **Network 1:** We consider $N = 50$ SUs, $C = 3$ channels characterized by the availability probabilities $\mu = [0.3, 0.5, 0.8]$ and user specific payoffs: $U_j(.) = w_j f(.)$, 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.
  - **Network 2:** We set $N = 10$, $C = 2$ and $\mu = [0.2, 0.8]$. We set $H_j = 1$ and $\rho_j = 0$ for all j.

Fairness index in Network 1 I

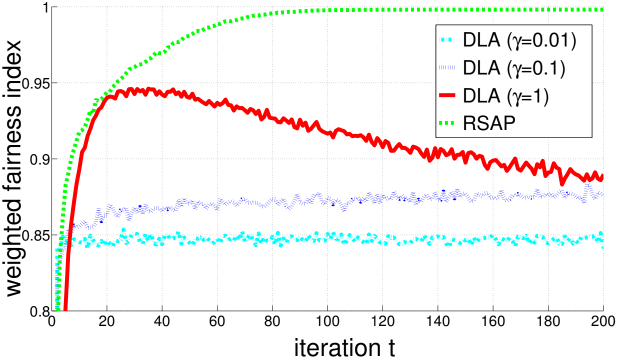


Figure : Weighted fairness index of RSAP and the DLA algorithm proposed in [Chen&Huang 2012]. Each curve represents an average over 1000 independent realizations.

Fairness index in Network 1 II

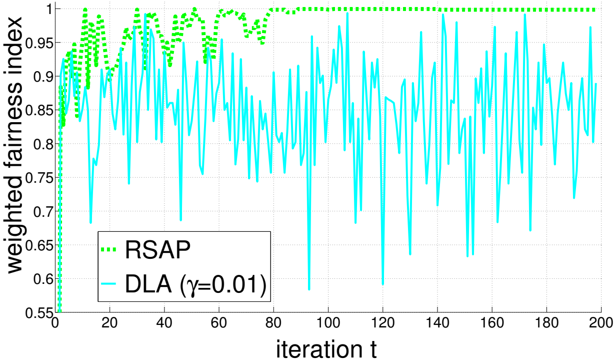


Figure : Weighted fairness index of RSAP and the DLA algorithm proposed in [Chen&Huang 2012]. Each curve represents a single realization of the two algorithms.

RSAP vs T&E

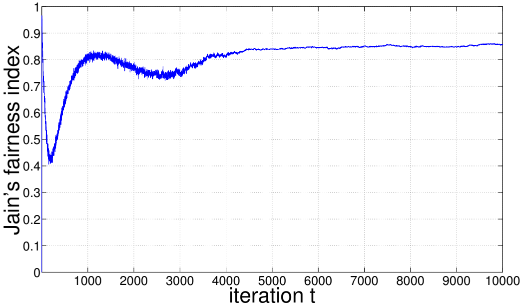


Figure : Trial and Error fairness index on Network 2 (average of 1000 trajectories).

RSAP vs T&E

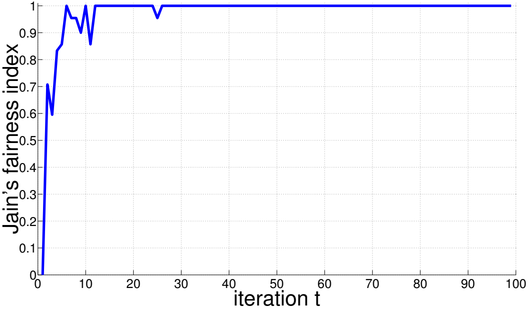


Figure : RSAP fairness index on Network 2.

Conclusion

- We discussed the distributed resource allocation problem in CRNs
- We have proposed a fully distributed scheme without any information exchange between SUs and based on self-imitation
- We have proved convergence using Ellison radius-coradius theorem
- We have compared RSAP to T&E [Pradelski&Young 2012] and to DLA [Chen&Huang 2012]

Further Work

- Congestion games on graphs
- More realistic models of the channel between the SU transmitter and the SU receiver
- Learning in presence of noise (SUs get only an estimate of the mean throughput at each iteration)
- Joint sensing and access problem