Interference Mitigation via Pricing in Time-Varying Cognitive Radio Systems

Alexandre Marcastel, E. Veronica Belmega, Panayotis Mertikopoulos, and Inbar Fijalkow

Abstract Despite the lure of a considerable increase in spectrum usage efficiency, the practical implementation of cognitive radio (CR) systems is being obstructed by the need for efficient and reliable protection mechanisms that can safeguard the quality of service (QoS) requirements of licensed users. This need becomes particularly apparent in dynamic wireless networks where channel conditions may vary unpredictably – thus making the task of guaranteeing the primary users (PUs)' minimum quality of service requirements an even more challenging task. In this paper, we consider a pricing mechanism that penalizes the secondary users (SUs) for the interference they inflict on the network's PUs and then compensates the PUs accordingly. Drawing on tools from online optimization, we propose an exponential learning power allocation policy that is provably capable of adapting quickly and efficiently to the system's variability, relying only on strictly causal channel state information (CSI). If the transmission horizon T is known in advance by the SUs, we prove that the proposed algorithm reaches a "no-regret" state within $O(T^{-1/2})$ iterations; otherwise, if the horizon is not known in advance, the algorithm still reaches a no-regret state within $O(T^{-1/2} \log T)$ iterations. Moreover, our numerical results show that the interference created by the SUs can be mitigated effectively by properly tuning the parameters of the pricing mechanism.

Key words: Cognitive radio, pricing mechanism, time-varying systems, online optimization, exponential learning

Alexandre Marcastel, ETIS/ENSEA – UCP – CNRS, Cergy-Pontoise, France, e-mail: alexandre.marcastel@ensea.fr; E. Veronica Belmega, ETIS/ENSEA – UCP – CNRS, Cergy-Pontoise, France, and Inria, e-mail: belmega@ensea.fr; Panayotis Mertikopoulos, French National Center for Scientific Research (CNRS), LIG F-38000 Grenoble, France and Inria, e-mail: panayotis.mertikopoulos@imag.fr; Inbar Fijalkow, ETIS/ENSEA – UCP – CNRS, Cergy-Pontoise, France, e-mail: inbar.fijalkow@ensea.fr

1 Introduction

Cognitive radio (CR) has been identified as one of the most promising solutions to face the enormous challenges of future and emerging communication networks in terms of capacity, quality of experience, and spectrum efficiency [1]. The proposal to achieve this is to introduce a two-tier hierarchy, based on spectrum licensing: on the one hand, primary users (PUs) have leased part of the spectrum and must be sheltered from harmful interference; on the other hand, the network's secondary users (SUs) are allowed to free-ride on the licensed part of the spectrum, provided that they comply with the PUs' minimum quality of service (QoS) requirements.

This opportunistic spectrum access paradigm gives rise to several major concerns. First, the PUs have no incentive to accept a spectrum lease that leaves them open to free-riding – even under protection against harmful interference. For instance, one of the most widespread ways to guarantee the PUs' contractual QoS guarantees is to impose a so-called *interference temperature* (IT) constraint [2] at the SUs transmission level, i.e. to require that the total interference caused to the licensed user in a given frequency band remain always below a given, fixed tolerance. However, ensuring that the SUs respect a rigid constraint at all times is a highly non-trivial task – e.g. because of channel estimation errors, imperfect SUs coordination (or total lack thereof), malicious SU behavior, etc. Second, the inherent temporal variability of multi-user wireless networks – caused by the users' unpredictable behavior coupled with the random dynamics of the wireless environment – poses a major challenge in protecting the PUs against harmful interference.

To tackle these concerns, we propose a pricing mechanism [3] to *a)* incentivize and reward the network's PUs for allowing SUs to co-exist in the same part of the spectrum; and *b)* act as an effective interference mitigator, keeping the interference created by the SUs at tolerable levels. More precisely, we posit that the system manager imposes a monetary cost for every IT constraint violation caused by the SUs as an increasing function of the violation. These sanctions are then used to reimburse the PUs whose quality of service requirements were violated.

Pricing mechanisms of this type have already been considered as efficient means of managing the interference in *static* multi-user networks [4–6]; the major difference here lies in the temporal variability of the wireless networks which introduces a vastly different (temporal) dimension in the analysis of said mechanisms. To account for these difficulties, we take an approach based on online optimization which provides a suitable framework for studying dynamically varying systems [7]. Building upon these tools, we propose an adaptive exponential learning policy [8], which relies only on strictly causal channel state information.

Our first theoretical result is that if a SU knows his transmission horizon T, he can match the performance of the best fixed *a posteriori* power allocation policy within $O(T^{-1/2})$. In other words, even though the proposed power allocation policy only requires *strictly causal* knowledge of past information, it matches asymptotically the performance of the best fixed policy that can be achieved with *non-causal* knowledge of the system's evolution. This result remains true even if the transmission horizon T is not known and the algorithm is used with a variable step-size

parameter; in that case however, the algorithm's *regret* (the gap between our algorithm and the best fixed policy) grows slightly to $O(T^{-1/2} \log T)$. These results are then validated by our numerical simulations which show that the network's SUs reach a no-regret state within a few tens of iterations in realistic wireless conditions.

2 System Model and Problem Formulation

We consider a cognitive radio (CR) network composed of M licensed *primary users* (PUs) and K unlicensed *secondary users* (SUs), transmitting simultaneously to a common access point (AP) over a shared frequency band of width W. Every PU $m \in \{1, \dots, M\}$ is assumed to lease a block of S_m orthogonal channels and transmits only over the leased part of the spectrum; by contrast, the network's SUs are assumed to free-ride over all available subcarriers. As a result, the Shannon rate of k-th user is given by the standard expression

$$R_k(\mathbf{p}; t) = \sum_{s=1}^{S} \log\left(1 + \frac{p_{ks}g_{ks}}{\sigma_{ks}^2 + \sum_{j \neq k} p_{js}g_{js} + p_s^{PU}g_s^{PU}}\right), \tag{1}$$

where $g_{ks} = |h_{ks}|^2$ is the (time-varying) channel gain between the k-th SU and the AP, $\sigma_{ks}^2 = \mathbb{E}[w_{ks}^\dagger w_{ks}]$ is the variance of the noise, p_{ks} is the transmit power of the k-th SU over the s-th subcarrier, and $\mathbf{p} = (p_{ks})_{k,s}$ is the power profile of all SUs.

In a power-constrained setting, the total power $P_k = \sum_{s \in S} p_{ks}$ of the k-th SU is de facto limited by the maximum transmit power $\bar{P}_k$ of the user's wireless device. As a result, the feasible set of the k-th SU is defined as:

$$\mathcal{P}_k = \{\mathbf{p}_k \in \mathbb{R}^S : p_{ks} \geq 0 \text{ and } \sum_{s=1}^{S} p_{ks} \leq \bar{P}_k\}. \tag{2}$$

In a CR context, the network operator must also shelter the PUs' contractual QoS guarantees from harmful interference by the SUs. This requirement often takes the form of a maximal interference threshold per sub-carrier [2], i.e.

$$\sum_{k=1}^{K} p_{ks}g_{ks} \leq I_s \quad \forall s \tag{3}$$

where I_s denotes the maximal interference tolerated by the PU who has leased sub-carrier s. This requirement depends on the powers of all SUs in an aggregate way; however, given that the SUs do not coordinate with one another (and also to induce fairness among the SUs), we assume here that the network operator also imposes a *user-specific* maximal interference requirement of the form

$$p_{ks}g_{ks} \leq P_s, \quad \forall s, \forall k, \tag{4}$$

thus providing an additional safety net to the PUs' transmissions.

Of course, both (3) and (4) represent a time-varying requirement from the SUs' viewpoint because the channel gains g_{ks} may vary unpredictably over time. In particular, given the lack of coordination between SUs and the fact that they do not have perfect channel state information before the transmission, it is impossible to ensure that these constraints will be met for all time. In turn, this raises a major concern for CR paradigm as the PUs have no incentive to pay for spectrum access rights that can be compromised at any given time.

To overcome this, instead of treating (3) and (4) as physical constraints at the SU level, we posit that the network operator charges a monetary cost to the SUs for any violation of the PUs' requirements, as a function of the severity of the violation; this cost is then reimbursed (at least partially) to the PUs whose QoS requirement was violated. More concretely, this pricing mechanism can be expressed by a *cost function* of the form

$$C_k(\mathbf{p};t) = \sum_{s=1}^{S} C\left(\sum_{k=1}^{K} p_{ks}g_{ks} - I_s\right) + \sum_{s=1}^{S} C\left(p_{ks}g_{ks} - P_s\right), \tag{5}$$

where $C(\cdot)$ is a non-decreasing, Lipschitz continuous and convex *pricing function*. For instance, a standard example of such a pricing function is the piecewise linear penalty

$$C(x) = \begin{cases} \lambda x & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where λ is the price per dBm of violation.

Putting all this together, the SUs' utility can be expressed as:

$$U_k(\mathbf{p}_k;t) = R_k(\mathbf{p};t) - C_k(\mathbf{p};t), \tag{7}$$

i.e. as the trade-off between the SU's achieved throughput and the cost paid to achieve it. Thus, given the system's evolution over time, we obtain the online optimization problem:

$$\begin{aligned} \text{maximize} \quad & U_k(\mathbf{p}_k;t) \\ \text{subject to} \quad & \mathbf{p}_k \in \mathcal{P}_k \end{aligned} \tag{P}$$

Given that the objective of each SU depends *explicitly* on time (via the channel gains $g_{ks}(t)$), our goal will be to determine a dynamic power allocation policy $\mathbf{p}_k(t)$ that remains as close as possible to the (evolving) solution of (P). However, due to the temporal variability of the channel gains g, the power $\mathbf{p}_k^*(t)$ that solves (P) at every given time t cannot be calculated ahead of time with strictly causal information.

On account of that, we will focus on the fixed power allocation policy that is optimal on average and in hindsight, i.e. the solution of the time-averaged problem:

$$\mathbf{p}_k^* \in \arg\max_{\mathbf{p}_k \in \mathcal{P}_k} \sum_{t=0}^{T} U_k(\mathbf{p}_k;t), \tag{8}$$

where $\mathcal{P}_k$ is the feasible set of the user k defined in (2). As before, the mean optimal solution $\mathbf{p}_k^*$ can only be calculated offline (i.e. it requires knowing the evolution of the system over the entire transmit horizon), and is only used as a benchmark for a dynamic power allocation policy $p_k(t)$ that relies only on strictly causal information. To make this comparison precise, we define a user's (cumulative) *regret* [7, 9] as:

$$\mathrm{Reg}_k(T) = \sum_{t=1}^{T} U_k(\mathbf{p}^*; t) - U_k(\mathbf{p}_k; t) \tag{9}$$

In words, the user's regret over the transmission horizon T measures the cumulative performance gap between the dynamic power strategy $\mathbf{p}_k(t)$ and the average optimum profile $\mathbf{p}_k^*$. In particular, if $\mathrm{Reg}_k(T)$ grows linearly with T, the user is not able to track changes in the system sufficiently fast. Accordingly, we will say that a power control policy $\mathbf{p}_k(t)$ leads to *no regret* if

$$\limsup_{T \to \infty} \mathrm{Reg}_k(T)/T \le 0 \quad \text{for all } k, \tag{10}$$

irrespectively of how the system evolves over time. If this is the case, it means that there is no fixed power profile yielding a higher utility in the long run; put differently, (10) provides an asymptotic guarantee that ensures that $\mathbf{p}(t)$ is at least as good as the mean optimal solution. We will further explore this property in Section 4.

3 Exponential Learning

To devise an online policy $\mathbf{p}_k(t)$ that leads to no-regret, our starting point will be as follows: First, each user's policy tracks the direction of gradient (or subgradient) ascent of their utility, without taking into account the problem' constraints as defined in (2). Subsequently, this "aggregated gradient" is mapped back to the feasible region via a suitably chosen exponential map, and the process repeats.

To be more precise, this procedure can be described by the recursion

$$\begin{aligned}
\mathbf{y}_k(t+1) &= \mathbf{y}_k(t) + \delta(t)\mathbf{v}_k(t), \\
p_{ks}(t+1) &= \bar{P}\frac{\exp\left(\mathbf{y}_{ks}(t+1)\right)}{1 + \sum_{s'=1}^{S} \exp\left(\mathbf{y}_{ks'}(t+1)\right)},
\end{aligned} \tag{DXL}$$

where $\mathbf{v}_k(t) = \partial_{k_k} U_k(\mathbf{p}; t)$ denotes the gradient of the k-th user's utility function and $\delta(t)$ is a non-decreasing step-size parameter (for an algorithmic implementation, see Algorithm 1 above).

Our goal in what follows will be to examine the no-regret properties of the online power allocation policy (DXL). To do so, let V denote an upper bound for $\mathbf{v}_k$, i.e.

$$\|\mathbf{v}_k\|^2 \le V_k^2. \tag{11}$$

Algorithm 1 Discrete-time exponential learning.

Parameter: step-size $\delta(t) > 0$.

Initialization: $\mathbf{y}_k \leftarrow 0; t \leftarrow 0$.

Repeat

$\quad \mid \quad t \leftarrow t + 1;$

$\quad \mid \quad$ allocate powers: $p_{ks} \leftarrow \bar{P} \frac{\exp(y_{ks})}{1 + \sum_{s'=1}^{S} \exp(y_{ks'})};$

$\quad \mid \quad$ get gradient data $\mathbf{v}_k \leftarrow \partial_{\mathbf{p}_k} U_k(\mathbf{p}_k; t);$

$\quad \mid \quad$ update scores: $\mathbf{y}_k \leftarrow \mathbf{y}_k + \delta(t)\,\mathbf{v}_k;$

$\quad \lfloor \quad$ **until** termination criterion is reached.

With all this at hand, our first result concerns the case where the transmission horizon is known in advance (for instance, as in a timed call), and (DXL) is employed with a constant, optimized step-size δ^* (the proof is omitted due to space limitations):

Theorem 1. *Assume that* (DXL) *is run for a given time horizon T with the optimized step-size $\delta_k^* = V_k^{-1}\sqrt{\log(1 + S)/T}$. Then, it enjoys the regret bound*

$$\mathrm{Reg}_k(T) \le 2V_k\bar{P}\sqrt{T\log(1 + S)}. \tag{12}$$

Consequently, the users' average regret $\mathrm{Reg}_k(T)/T$ vanishes as $O(T^{-1/2})$, i.e. (DXL) *leads to no regret.*

The above result is contingent on the SUs knowning the transmission horizon T in advance. If this is not the case, it is more advantageous to consider a strictly decreasing step-size so as to reduce the algorithm's jitter in fluctuations of unknown length. We illustrate this in Theorem 2 below (again, we omit the proof due to space limitations):

Theorem 2. *Assume that* (DXL) *is run for an unknown time horizon T with the variable step size $\delta(t) = at^{-1/2}$ for some $a > 0$. Then, it enjoys the regret bound:*

$$\mathrm{Reg}_k(T) \lesssim \bar{P}\left(\frac{\log(1 + S)}{a} + aV_k^2\right)T^{1/2} + a\bar{P}V_k^2 T^{1/2}\log T. \tag{13}$$

Consequently, the users' average regret $\mathrm{Reg}_k(T)/T$ vanishes as $O(T^{-1/2}\log T)$, i.e. (DXL) *leads to no regret.*

This implies that, in both cases (known vs. unknown horizon), the rate of regret minimization depends on the system parameters. We further remark that the users' average regret vanishes faster if the transmission horizon T is known in advance, but the level of this disparity ($\log T$) is fairly moderate. This disparity can be overcome completely by means of a more complicated step-size policy known as a "doubling trick" [7] but, for simplicity, we do not present this approach here.
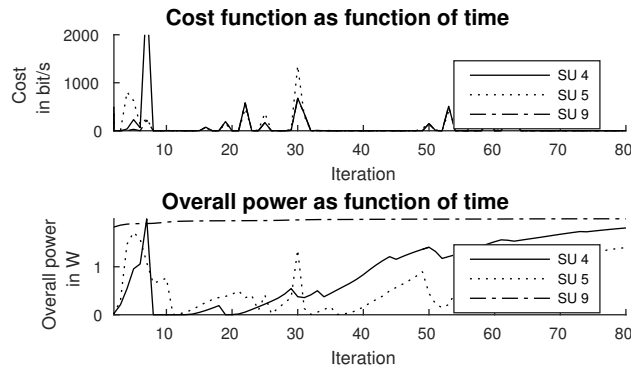
Cost function as function of time

Overall power as function of time

Fig. 1 Evolution of the cost function and overall power for SUs 4, 5 and 9 as function of time. When the channel gains are low (e.g SU 9) the interferences created by the SUs are also small and their cost function is null, i.e they meet acpPU' requirement. In contrast, when the interference created by a SU exceeds the requirements (because of strong channel gains for example) the SU is immediately penalized by imposing a suitable cost which results in a decrease of the transmit power at the next iterations (see SUs 4 and 5).

4 Numerical results

To validate our theoretical results we have performed extensive numerical simulations of which we exhibit a representative sample below.

We focus on an uplink cellular network with a fixed AP. Specifically, we consider a wireless system operating over a 10 MHz band centered around the carrier frequency $f_c = 2$ GHz. The total bandwidth is divided in 64 sub-carriers. We consider 1 PU and 9 SUs randomly positioned inside a square cell of side 2 km, following a Poisson Point Process. The maximum interference temperature in each sub-carrier in (3) is fixed at $I_s = -90$ dBm for all s and the constraint per sub-carrier (4) is limited at $P_s = -110$ dBm for all s and the variance of the noise is set at: $\sigma^2 = -120$ dBm. We also assume that the SUs' have a maximum transmit power of $\bar{P} = 30$ dBm. The channels between the wireless users and the AP are generated according to the realistic COST-HATA model for a suburban macro-cellular network [10] with fast and shadow-fading attributes as in [11]. Each SU is assumed to be mobile with a speed chosen arbitrarily between 10 and 130 km/h.

In Fig. 1, we plot the cost function defined in (5) and the overall power consumption as function of the time by SU. To reduce graphical clutter, we only illustrate this data for three representative SUs at various distances from the AP. Specifically, the initial distance from the AP of each of the three focal users is $d_4 = 600.1$ m for SU 4, $d_5 = 943.8$ m for SU 5, and $d_9 = 979.4$ m for SU 9; respectively, the SUs' speeds are $v_4 = 50$ km/h, $v_5 = 10$ km/h, and $v_9 = 90$km/h. If interfering channel gains are low, as the 9th SU, the users can transmit at maximum power (i.e. at $\bar{P} = 30$ dBm) without creating harmful interference to the PUs. At the opposite, when the channel gains become high, the induced interference also increases. As a
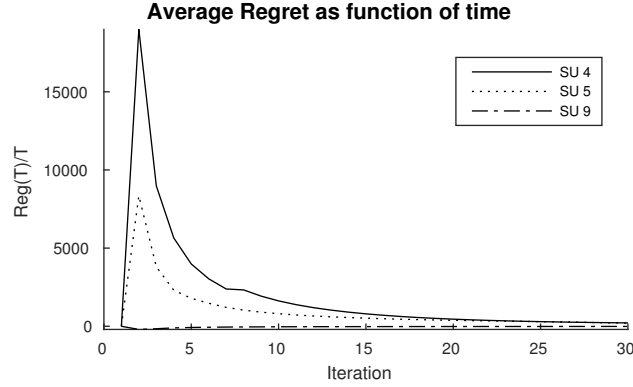
Average Regret as function of time

Fig. 2 Evolution of the SUs average regret as function of time. We see that the SUs' online power allocation policy quickly leads to zero average regret; specifically, (DXL) matches the optimal fixed transmit profile in hindsight within a few tens of iterations.
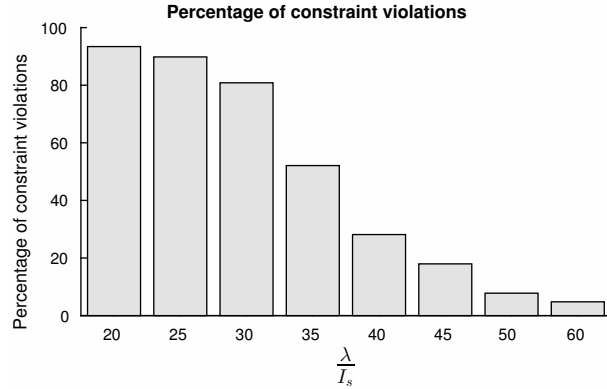
Percentage of constraint violations

Fig. 3 Fraction of time at which the PU's interference constraint are violated. The higher is λ the higher is the penalty applied to the SUs in case of QoS requirement violations. This results in less interferences violations from the SUs. Thus, the operator can control the interference by tuning λ and can efficiently protect the PU.

result, the SUs transmitting at high powers are penalized via the cost function and decrease their transmit powers as a result.

In Fig. 2, we plot the evolution of the opportunistic users' average regret as a function of time. We see that each SU's regret quickly drops to zero at a rate which depends on the user's individual channels, on the choice of the step parameter α and on the penalty parameter λ – cf. Eq. (6). As a result, the online power allocation policy we propose matches the best fixed transmit profile within a few number of iterations, despite the channels' significant variability over time for the same SUs.

Finally, in Fig. 3, we plot the fraction of time at which the PU's interference constraints are violated. To be precise, we plot the fraction of iterations at which at least one SU creates interference aboves the maximum tolerated levels. As ex-

pected, higher λ values leads to fewer constraint violations. Therefore, the exponential learning policy (DXL) with the cost function defined in (5) and (6) allows the operator to efficiently use the total bandwidth by allowing SUs to transmit while protecting the PUs and that despite the unpredictability of the system's variation over time.

5 Conclusions and perspectives

We have investigated a dynamic multi-user CR system in which multiple opportunistic users are allowed to co-exist with the PUs. In order to control the interference created by the SUs, the system owner implements a pricing mechanism which also serves a second purpose, i.e. as a reward incentive for the PUs to accept an open spectrum license. In this context, we propose an exponential learning algorithm that allows the SUs to adapt their power allocation policies to the dynamic changes in the environment in an optimal way regarding the tradeoff between their achievable rate and the cost for the harmful interference they inflict. Our simulations show that by tuning the parameters of the cost function, the system owner can efficiently control the interference created by the SUs - and, hence, protect the PUs's transmissions - despite the dynamic and arbitrary variations of the system.

References

[1] Qualcomm, "The 1000x data challenge," *Technical Report*, 2014.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," vol. 23, no. 2, pp. 201–220, 2005.

[3] J. Huang and L. Gao, "Wireless network pricing," *Synthesis Lectures on Communication Networks*, vol. 6, no. 2, pp. 1–176, 2013.

[4] T. Alpcan, T. Başar, R. Srikant, and E. Altman, "Cdma uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, no. 6, pp. 659–670, 2002.

[5] C. Saraydar, N. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *Communications, IEEE Transactions on*, vol. 50, no. 2, pp. 291–303, 2002.

[6] S. D'Oro, P. Mertikopoulos, A. L. Moustakas, and S. Palazzo, "Interference-based pricing for opportunistic multicarrier cognitive radio systems," vol. 14, no. 12, pp. 6536–6549, 2015.

[7] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[8] P. Mertikopoulos, A. L. Moustakas *et al.*, "The emergence of rational behavior in the presence of stochastic perturbations," *The Annals of Applied Probability*, vol. 20, no. 4, pp. 1359–1388, 2010.

[9] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[10] G. F. Pedersen, *COST 231-Digital mobile radio towards future generation systems*. EU, 1999.

[11] G. Calcev, D. Chizhik, B. Göransson, S. Howard, H. Huang, A. Kogiantis, A. F. Molisch, A. L. Moustakas, D. Reed, and H. Xu, "A wideband spatial channel model for system-wide simulations," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 2, pp. 389–403, 2007.