Online Power Allocation for Opportunistic Radio Access in Dynamic OFDM Networks
Alexandre Marcastel, E Veronica Belmega, Panayotis Mertikopoulos, Inbar Fijalkow

To cite this version:

HAL Id: hal-01387044
https://hal.archives-ouvertes.fr/hal-01387044
Submitted on 25 Oct 2016

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.
Online Power Allocation for Opportunistic Radio Access in Dynamic OFDM Networks

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

* ETIS/ENSEA – UCP – CNRS, Cergy-Pontoise, France
† Inria
‡ French National Center for Scientific Research (CNRS), LIG F-38000 Grenoble, France

Abstract—User mobility has become a key attribute in the design of optimal resource allocation policies for future wireless networks. This has become increasingly apparent in cognitive radio (CR) systems where the licensed, primary users (PUs) of the network must be protected from harmful interference by the network’s opportunistic, secondary users (SUs): here, unpredictability due to mobility requires the implementation of safety net mechanisms that are provably capable of adapting to changes in the users’ wireless environment. In this context, we propose a distributed learning algorithm that allows SUs to adjust their power allocation profile (over the available frequency carriers) “on the fly”, relying only on strictly causal channel state information. To account for the interference caused to the network’s PUs, we incorporate a penalty function in the rate-driven objectives of the algorithms that are provably capable of adapting to changes in the users’ wireless environment [2].

One of the most promising solution candidates for these challenges is that of cognitive radio (CR), a paradigm which allows opportunistic spectrum access by otherwise unlicensed users so as to maximize its utilization [3, 4]. In more detail, cognitive radio systems introduce a hierarchy based on spectrum licensing: on the one hand, the network’s primary users (PUs) have leased part of the spectrum and must be protected from harmful interference by opportunistic spectrum access; on the other hand, the network’s secondary users (SUs) try to free-ride on unallocated parts of the spectrum without compromising the PUs contractual quality of service guarantees. In particular, SUs are allowed to transmit over a shared set of channels, provided that the interference they induce to the network’s PUs is kept below a pre-negotiated threshold.

In this paper, we focus on opportunistic radio access in orthogonal frequency division multiplexing (OFDM) systems where the wireless environment and the users’ loads and demands change dynamically over time, in an unpredictable way. Specifically, we study the problem of SUs throughput maximization while keeping the interference caused to the network’s PUs under a fixed tolerance set by the PUs. This problem has attracted considerable interest in the literature [5–8], but the vast majority of works on this topic have focused on the case where the users’ channels remain static – or, at least, stationary – throughout the transmission horizon.

In more realistic network scenarios however, the users’ mobility, their unpredictable behavior (going online and offline in an ad hoc manner), and the complex multi-path fading attributes of the wireless network cause this stringent stationarity assumption to fail. As a result, static solution concepts (such as social optima or Nash equilibria) [5–7] are no longer relevant when the users’ wireless environment changes arbitrarily over time. Instead, a suitable solution framework is provided by online optimization and regret minimization methods [9, 10] which allow users to adapt to changes in the wireless environment, quickly and efficiently.

To achieve this, we propose a dynamic power allocation policy based on exponential learning [11, 12] that relies only on strictly causal channel state information [7, 13]. To illustrate the performance of the proposed algorithm, we focus on a simple network composed of one PU and several SUs which transmit
simultaneously to a common access point (AP) over several orthogonal frequency bands. In this context, we are able to show that the SUs’ average regret vanishes as $O(T^{-1} \log S)$ where $S$ is the number of subcarriers and $T$ is the transmission horizon. As a result, the algorithm is able to reach a no-regret state quickly, even for large numbers of subcarriers.

The closest work to ours is [7] where the authors use a similar learning technique assuming stationary channels and show that SUs converge to a Nash equilibrium; however, this result is no longer relevant in our case because the users’ environment evolves dynamically over time. This dynamic aspect is present in [13] where the authors use online learning to maximize the SUs’ throughput; nonetheless, because interference constraints are absent in [13], the algorithm proposed therein leads users to transmit at full power, thus causing significant interference to the PUs. Instead, the techniques developed herein enable the network’s SUs to maximize their throughput while staying below the PUs’ interference tolerance level, despite the system’s unpredictability.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a wireless cognitive radio network with one PU and $K$ SUs, each transmitting to a common receiver via a shared channel over $S$ non-interfering subcarriers. In this multiple access channel (MAC) context, the received signal at the AP in the subcarrier $s$ is given by the familiar baseband model:

$$r_s = \sum_{k=1}^{K} x_{ks} h_{ks} + h_{s} x_{s}^U + w_s,$$

where $x_{ks}$ is the transmitted signal of the $k$-th SU over subcarrier $s$, $h_{ks}$ is the associated transfer coefficient between the $k$-th SU and the AP, $x_{s}^U$ and $h_{s}^U$ are the corresponding quantities for the PU, and $w_s$ is the ambient noise in the channel.

For decoding purposes, we assume single user decoding (SUD) at the receiver, meaning that interference from non-designated transmitters is treated as additive (Gaussian) noise. In this case, the Shannon rate of user $k$ at time $t$ will be:

$$R_k(p_k; t) = \sum_{s=1}^{S} \log \left( 1 + \frac{g_{ks}(t)p_{ks}}{\sigma^2_s + \sum_{j \neq k} g_{js}(t)p_{js} + g_{s}^U p_{s}^U} \right),$$

where $p_{ks} = \left( p_{ks} \right)_{s=1}^{S}$ is the transmit power profile of user $k$, $g_{ks}(t) = \| h_{ks}(t) \|^2$, $s = 1, \ldots, S$ are the associated channel gains at time $t$, and $\sigma^2_s = \mathbb{E}[w_s w_s]$ is the variance of the noise.

In the context of power-limited users, the users’ total transmit power $P_k = \sum_{s=1}^{S} p_{ks}$ will be bounded from above by the maximum transmit power $P_{\text{max}}$ of their wireless devices. Thus, the feasible power region of each user is

$$\mathcal{P}_k = \{ p_k \in \mathbb{R}^S : p_{ks} \geq 0 \text{ and } \sum_{s=1}^{S} p_{ks} \leq P_{\text{max}} \}.$$

In the absence of other considerations, the unilateral objective of each user would be the maximization of its rate subject to the total power constraint (3) above. However, in a cognitive radio context, the network operator must also safeguard the contractual quality of service (QoS) guarantees that the PU has already paid for typically in the form of a maximum interference tolerance per subcarrier. On that account, the network operator also imposes to each SU the requirement

$$g_{ks}(t)p_{ks} \leq I_{\text{max}}.$$  

In contrast to the maximum power constraint of (3), the requirement (4) varies with time (because the SUs’ channels themselves vary with time), so it cannot be enforced a priori: since there is no way to predict one’s channel in advance, it is not possible to devise a policy that always respects this requirement either. Hence, instead of treating (4) as a (dynamic) physical constraint, we incorporate it in the SUs’ (dynamic) utility function defined as follows:

$$U_k(p_k; t) = R_k(p; t) - \sum_{s=1}^{S} C(g_{ks}(t)p_{ks}/I_{\text{max}} - 1),$$

where $C(x)$ is a Lipschitz continuous, convex penalty function which is non-decreasing in $x$.

In the above, the convexity assumption for $C(x)$ essentially acts as an interference control mechanism. Specifically, it implies that the same increase in the incurred interference leads to a higher violation penalty when the network operates in a high-interference state (as opposed to a mild-interference one). As a result, using a convex penalty scheme drives the network’s SUs to transmit at lower powers relative to the PU’s QoS requirements. As such our archetypal example will be the piecewise linear cost function:

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

where $\lambda$ is a sensitivity parameter that represents the incurred penalty when a SU violates the interference tolerance requirement (4).

In view of all this, we obtain the online problem:

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

Given that each SU’s objective depends explicitly on time (via its dependence on the channel gains $g_{ks}(t)$), our aim will be to determine a dynamic power control policy $p_k(t)$ that is as close as possible to maximizing the above objective over time. However, given that it is not possible to predict the channel gains $g_{ks}(t)$ ahead of time, it is not possible to predict the optimal transmit power profile $p^*_k(t)$ which solves (P) in a real-time, online manner. Instead, we will focus on power allocation policies that can be implemented with strictly causal knowledge and which are asymptotically optimal in hindsight, in a sense made precise below.

To make all this precise, fix some horizon $T$ over which the problem (P) is run, and let $p^*_k$ denote the optimum (fixed) power profile over the horizon, i.e. the solution of the time-averaged problem:

$$p^*_k \in \arg \max_{p_k \in \mathcal{P}_k} \int_0^T U_k(p_k; t) \, dt,$$

where $\mathcal{P}_k$ is the feasible set of user $k$ defined by the constraints (3). Of course, this solution can only be computed in hindsight
yielding a higher utility in the long run; put di

irrespectively of how the system evolves over time.

provides an asymptotic guarantee that ensures that the policy

that a power control policy leads to

no regret

in the system su

ffi

further explore this property in Section IV.

scheme:

will be mapped back onto the problem’s feasible region via an

the problem’s constraints; subsequently, the resulting trajectory

step, we will track the direction of steepest ascent of each

no regret in (P), our main idea will be as follows: as a first

More precisely, we will consider the exponential learning

III. EXPONENTIAL LEARNING

To devise an online power allocation policy that leads to

no regret in (P), our main idea will be as follows: as a first

step, we will track the direction of steepest ascent of each

user’s utility (via its gradient) without taking into account

the problem’s constraints; subsequently, the resulting trajectory

will be mapped back onto the problem’s feasible region via an

“exponential projection” step.

More precisely, we will consider the exponential learning scheme:

\[ \dot{p}_{ks} = \frac{\exp(v_{ks})}{1 + \sum_{r=1}^{S} \exp(y_{kr})} \cdot \frac{\partial \ln \Pi_{P}}{\partial p_{k}} \]

where \( v_{k} \) is the subgradient of the \( k \)-th user’s utility function (for a pseudocode implementa-

tion, see Algorithm 1 below). As noted above, the raison d’être

of the exponentiation step in (XL) is to project the auxiliary

variable \( y_{k} \) back to \( P_{k} \); it is easy to see that \( \sum p_{ks} \leq P_{\text{max}} \), so

the power allocation policy induced by (XL) is a feasible one.

With this in mind, our main theoretical result for (XL) is as

follows:

**Theorem 1.** The power allocation policy (XL) enjoys the regret bound

\[ \text{Reg}_{k}(T) \leq P_{\text{max}} \log(1 + S). \]

Consequently, the users’ average regret \( \text{Reg}_{k}(T)/T \) vanishes

asymptotically as \( O(1/t) \), i.e. (XL) leads to no regret.

**Proof:** See Appendix A.

Importantly, the regret bound provided in Theorem 1 is

universal and only depends on the “size” of the users’ feasible

regions (the number of spectral degrees of freedom \( S \) and the

users’ maximum transmit power \( P_{\text{max}} \)). In particular, the

guarantee (10) does not depend on the system’s average channel

quality, the number of users, or other attributes of the system. We

thus conclude that (XL) is particularly flexible and can be used

“as is” in a fairly wide range of decentralized CR systems: as

long as the number of OFDM subcarriers shared by the focal

users remains (roughly) constant, the users will converge to a

no-regret state in the same rate.

IV. NUMERICAL RESULTS

To validate our theoretical results we performed extensive

numerical simulations of which we exhibit a representative

sample below.

Our focus is an uplink cellular network with a fixed AP.

Specifically, we consider a wireless system operating over a

10 MHz frequency band centered around the carrier frequency

\( f_{c} = 2 \) GHz. The cell is a square of side-length equal to

2 km with the AP at its center. The network’s PU is randomly

positioned inside the cell and we consider \( K = 9 \) SUs, also

placed randomly in the cell, following a Poisson point process.

The maximum interference is fixed at \( I_{\text{max}} = -83 \) dBm

and we assume that the SUs’ wireless devices have a maximum

transmit power of \( P_{\text{max}} = 30 \) dBm. Furthermore, each SU is

assumed to be mobile with a speed chosen arbitrarily between

10 and 130 km/h. Finally, the channels between the wireless

users and the AP are generated according to the realistic COST-

HATA model for a suburban macro-cellular network [14] with

fast- and shadow-fading attributes as in [15].

In Fig. 1, we plot the evolution of the SUs’ channel gains

and their respective Shannon rates as a function of time. To

reduce graphical clutter, we only illustrate this data for three

representative SUs at various distances from the AP. Specifi-

cally, the distance from the AP of each of the three focal

users is \( d_{2} = 131.8 \) m for SU 2, \( d_{3} = 172.7 \) m for SU 3, and

\( d_{7} = 779.6 \) m for SU 7; respectively, the SUs’ speeds

are \( v_{2} = 50 \) km/h, \( v_{3} = 90 \) km/h, and \( v_{7} = 10 \) km/h. A first

observation is that the SUs’ rate is directly correlated to their

channel gains; moreover, there are rapid variations in the SUs’

throughput that are directly correlated with the variations – and

responses – in the SU’ power allocation policies.

In Fig. 2, we plot the evolution of the users’ transmit powers

(in dBm) and the cost for inducing harmful interference to the

system’s PU. In this case, if the SUs’ channel gains are low, the

induced interference is also low, so users can transmit at max-
imum power – i.e. at $P_{\text{max}} = 30$ dBm (SU 7). On the contrary, when the channel gains become high, the induced interference also increases. As a result, the SUs transmitting at high powers are penalized via the penalty function (6) and decrease their transmit powers as a result thereof. Hence, the penalty function plays a key role in CR interference management.

In Fig. 3, 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 non-positive values at a rate which depends on the user’s individual channels and on the penalty parameter $\lambda$ – cf. Eq. (6). As a result, the online power allocation policy we propose matches the best fixed transmit profile in hindsight within a few tens of iterations, despite the channels’ significant variability over time.

Finally, in Fig. 4, we plot the fraction of times at which the PU’s tolerated interference levels are violated – specifically, the fraction of iterations at which at least one SU causes interference above $I_{\text{max}}$ to the PU. As expected, higher values of $\lambda$ lead to fewer constraint violations. Hence, by combining the exponential learning policy (XL) with the penalty scheme (6), the network operator is able to allow opportunistic spectrum usage while effectively – and efficiently controlling the induced interference and protecting the PU’s transmission – and, all this, despite the unpredictable variability of the network’s channels over time.

V. Conclusions and perspectives

In this paper, we proposed a distributed power allocation algorithm for the uplink of a time-varying cognitive radio network based on online optimization and exponential learning. Our algorithm allows the network’s opportunistic users to achieve an optimal average performance in terms of throughput while ensuring at the same time that the induced interference is kept on average below a maximum, tolerated level. The proposed algorithm is simple, distributed, it relies on strictly causal channel state information, and its gap to the $a$ $posteriori$ fixed optimal policy decays as $O(T^{-1} \log S)$ in the number of subcarriers ($S$) and the transmission horizon ($T$). All these properties make for a promising power allocation policy in flexible and
dynamic future wireless communications. Moreover, from a CR viewpoint, we show that the system owner can effectively allow opportunistic access to the spectrum while controlling the primary user’s transmissions by tuning a scalar parameter which controls the trade-off between the opportunistic users’ rates and their created interference in an adaptive, dynamic way.

**APPENDIX**

**Proof of Theorem 1:** The first step in proving the regret bound (10) is to use the concavity of the users’ utility function to write

\[
\text{Reg}_k(T) = \int_0^T U_k(p^*_k(t); t) - U_k(p_k(t); t) \, dt
\]

\[
\leq \int_0^T \langle v_k(p_k(t); t) \rangle \left| p^*_k(t) - p_k(t) \right| \, dt
\]

\[
= \langle y_k(T) | p^*_k \rangle - \int_0^T \langle y_k(t) | p_k(t) \rangle \, dt, \tag{11}
\]

where we have used the fact that \( y_k = v_k \) and that \( v_k = \partial_{p_k} U_k \) by construction – recall the definition of the policy (XL) in Sec. III. Note now that the exponentiation in (XL) can be written as

\[
\frac{\exp(y_{ks})}{1 + \sum_{s'=1}^S \exp(y_{ks'})} = \frac{\partial}{\partial y_{ks}} \log \left(1 + \sum_{s'=1}^S \exp(y_{ks'})\right). \tag{12}
\]

Thus, letting \( f(y_k) = P_{\max} \log \left(1 + \sum_{s'=1}^S \exp(y_{ks'})\right) \), the bound (11) becomes:

\[
\text{Reg}_k(T) \leq \langle y_k(T) | p^*_k \rangle - \int_0^T \langle y_k(t) | \nabla_{y_k} f(y_k(t)) \rangle \, dt
\]

\[
= \langle y_k(T) | p^*_k \rangle - f(y_k(T)) + f(0), \tag{13}
\]

where we used the fact that (XL) is initialized with \( y(0) = 0 \).

By Fenchel’s inequality [16], we then get

\[
f(y) + f^*(p) \geq \langle y | p \rangle, \quad \text{for all } p, y \in \mathbb{R}^S, \tag{14}
\]

where \( f^*(p) \) denotes the convex conjugate of \( f \), viz.

\[
f^*(p) = \sup_{y \in \mathbb{R}^S} \langle y | p \rangle - f(y). \tag{15}
\]

Thus, substituting in (13), we obtain

\[
\text{Reg}_k(T) \leq f^*(p_k^*_T) + f(0) = f^*(p_k^*_T) + P_{\max} \log(1 + S), \tag{16}
\]

so we are left to show that \( f^*(p) \leq 0 \) for all \( p \in \mathcal{P} \equiv \{ p \in \mathbb{R}^S : p_s \leq 0 \text{ and } \sum_s p_s \leq P_{\max}\} \). To that end, let \( x_s = p_s/P_{\max} \); then, it suffices to show that

\[
\sum_{s=1}^S x_s y_s \leq \log \left(1 + \sum_{s=1}^S \exp(y_s)\right), \tag{17}
\]

for all \( y \in \mathbb{R}^S \) and for all \( x \in \Delta \equiv \{x \in \mathbb{R}^S : x_s \leq 0 \text{ and } \sum_s x_s \leq 1\} \). However, since the log-sum-exp function is convex in \( y \), Jensen’s inequality readily yields:

\[
\exp \left(\sum_{s=1}^S x_s y_s\right) \leq \sum_{s=1}^S x_s \exp(y_s) \leq 1 + \sum_{s=1}^S x_s \exp(y_s), \tag{18}
\]

and (17) follows by taking logarithms on both sides. We conclude that \( f^*(p_k^*_T) \leq 0 \) for all \( p_k^*_T \in \mathcal{P} \), and our claim follows.

**References**


